The Impact of PISA on Students' Learning: a Chinese Perspective

Huiming Ding

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Education

April, 2020

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

During the journey of my PhD, I have received support and help from various people. I would like to express my gratitude to them.

First, I want to thank my supervisor, Dr Matt Homer, for providing meticulous, insightful and constructive feedback throughout my PhD which stimulates me to critical thinking of my research. My appreciation also goes to your constant support and encouragement not only in supervising my PhD but also in developing me as a researcher. I would like extends my thanks to my co-supervisor, Innocent Tasara, who also gave patient supervision and continuous support for my research. Your comments provided many insights into my research, especially in the qualitative strand. The discussions with either of you in the supervision meetings were always fruitful and inspiring.

Thanks also go to Jim Ryder and Gill Main, my thesis transfer panel members, for your suggestions on improving my research.

I would also express my thanks to the School of education for providing me with the scholarship. Thanks also to the PGR community within the School for its support and various interesting activities available for PGRs, which added to my PhD journey many beautiful memories about doing research and beyond.

Thanks to Lei Wang and Ning Han, my former line managers in China's National Education Examinations Authority, for your understanding of my decision of applying for a PhD, and your support during the time when I was preparing for the application. Thanks also to you for giving me the approval of using Fangshan PISA data, and to my former colleagues who helped with processing my application of using the data and sending me the data files.

I am also grateful for Fangshan Education commission and my contacts there and in schools for their support for my research in terms of approaching and recruiting interviewees. I also would like to thank all the interviewees involved in my research for their active participation, and positively sharing their experience, perceptions, and thoughts.

Thanks to Francesco Avvisati and Miyako Ikeda in the OECD for answering my enquiries about PISA approaches to processing data for adjusting performance trends and multilevel modelling of complex models.

Thanks to Ana Ilse Benavides Lahnstein and Liza Isyqi Binti Ramli who helped with reviewing my interview guides and providing insightful suggestions on conducting interviews before I started fieldwork. Thanks also to the mathematics teacher who was involved in my pilot for sharing the thoughts about the interview questions for mathematics teachers and providing comments.

I would also like to say many thanks to my officemates, Ana Ilse Benavides Lahnstein, Syakima Ibrahim, and Mohamed Suleiman for your help during my PhD. Thanks to you all for kindly and actively familiarising me with the facilities in the office and the research centre, and generously sharing your experience of doing a PhD, which made me feel very welcomed in joining you and adapt quickly to the new environment. Thanks also go to all my officemates who contributed to building and maintaining the friendly, respectful, and supportive working space. Especially thanks to Sulaiman Al Jamoudi for your kindness and support.

Thanks to the friends who listened to the progress of my research. A great thanks to my friend Yan Gu who is always there with support and care.

Finally, but most importantly, I must express my deepest gratitude to my beloved parents and sister. Thanks to you for understanding me, and for your constant and unconditional love, trust, and support.

Abstract

This thesis investigates PISA's impact on students' learning in a local context, Fangshan District of Beijing, in China. PISA's growing influence on educational policymaking in domestic education systems has been widely discussed, but concerns about policy borrowing and PISA's governing power on education have been raised. However, these discussions rarely look further into local contexts to investigate PISA's impact on students' learning.

Through mixed methods research, this thesis presents an investigation into the impact PISA has had on students' learning and how this impact occurs. A conceptual framework employing theories about washback effect and ecological systems theory was developed to underpin the research. Sixteen local educational policymakers and practitioners were interviewed to identify how PISA is used in the local context, and to gather their views on the perceived impact of PISA on students' learning. Fangshan PISA data were used to triangulate their perceptions, and also to expand the understanding of their use and translation of PISA data in policymaking.

Thematic analysis of interview data reveals that PISA is perceived as a new perspective, which is different from domestic assessment in some aspects, for benchmarking the quality of local education. PISA is also used as an impetus motivating local initiatives to improve educational quality, in which some PISA concepts are integrated. The mechanism of PISA's impact on students' learning was conceptualised via making the factors of different contextual layers which negotiate PISA's impact explicit. Interviewees perceived that through school enactment of the local PISA-motivated initiatives, and reforms of national curriculum and assessment, to some extent, students' learning has been gradually fostered. These perceptions are largely supported by the trend analysis and multilevel modelling of PISA data. Research findings also indicate the challenges that the local PISA data users face for appropriately interpreting and translating PISA data to inform educational policymaking.

Table of contents

Acknowledgements .... **iii**
Abstract .... **v**
Table of contents .... **vi**
List of Tables .... **xii**
List of Figures .... **xiii**
Abbreviations .... **xiv**
Chapter 1 Introduction **1**
1.1 Background 1
1.2 Aims 5
1.3 The structure of the thesis 6
Chapter 2 Education context in mainland China **8**
2.1 An overview of education system 8
2.2 Curriculum reform 13
2.3 Assessment system reform 15
Chapter 3 Literature Review **21**
3.1 An overview of PISA 21
3.2 The impact of PISA on educational policies 24
3.2.1 Policy responses to PISA findings vary across education systems 25
3.2.2 Policy responses to PISA findings change over PISA cycles 27
3.2.3 The phenomenon of policy borrowing 28
3.2.4 The governing power of PISA on global education 29
3.3 The use and impact of PISA in mainland China 31
3.3.1 The use and impact of PISA at China national level 31
3.3.2 The use and impact of PISA at China regional level 33
3.3.3 PISA facilitates China's engagement in global communication of education policy 35
3.4 PISA results about factors accounting for students' performance .. 36
3.4.1 Input factors 37
3.4.1.1 Socioeconomic status 37
3.4.1.2 Gender 39
3.4.1.3 Age 40

3.4.1.4 Grade ... 41
3.4.2 Processes factors ... 42
3.4.2.1 Opportunities to learn ... 42
3.4.2.2 Teaching practices and quality ... 44
3.4.3 Non-cognitive outcomes factors ... 46
3.4.3.1 Motivation to learn mathematics ... 47
3.4.3.2 Self-beliefs in learning mathematics ... 48
3.5 PISA limitations ... 51
3.5.1 Measurement model ... 51
3.5.2 Sampling ... 52
3.5.3 Translation ... 53
3.5.4 Content coverage ... 54
3.5.5 Causality inference ... 55
3.6 Developing a conceptual framework ... 56
3.6.1 Washback effect ... 57
3.6.2 Ecological systems theory ... 61
3.7 Summary ... 63
Chapter 4 Methodology ... 66
4.1 Research aims and research questions ... 66
4.2 Research paradigm ... 68
4.3 Research design ... 68
4.4 Research methods ... 70
4.5 Data collection ... 72
4.5.1 Qualitative data collection ... 72
4.5.1.1 The context of the fieldwork site ... 72
4.5.1.2 Piloting ... 73
4.5.1.3 Sampling of interview participants ... 73
4.5.1.4 The process of interviews ... 77
4.5.2 Quantitative data obtainment ... 78
4.6 Data analysis procedures ... 79
4.6.1 Thematic analysis of the interview data ... 80
4.6.2 Quantitative data analyses on PISA data ... 82
4.6.2.1 Trend analysis ... 82
4.6.2.2 Multilevel modelling analyses ... 84

4.6.2.2.1 Explanatory variables 84
4.6.2.2.2 Missing values and multiple imputation 89
4.6.2.2.3 Modelling strategy 93
4.7 Ethical considerations 94
4.7.1 Informed consent 95
4.7.2 Confidentiality and anonymisation 95
4.7.3 Data management and protection 96
4.8 Trustworthiness 96
Chapter 5 Findings 99
5.1 The use of PISA in the local context 99
5.1.1 PISA is perceived as a new perspective for reviewing the quality of local education 99
5.1.1.1 Evaluating the quality of local education 100
5.1.1.2 Enlightening educational policymakers on teaching 104
5.1.2 PISA has catalysed various initiatives for improving the quality of local education 107
5.1.2.1 Increasing students’ exposure to reading 107
5.1.2.2 Integrating PISA concepts into teaching to improve mathematics and science learning 112
5.1.2.2.1 Involving teachers to improve students’ mathematics and science learning 113
5.1.2.2.2 Improving mathematics learning through improving instructional design 115
5.1.2.2.3 Translating PISA science framework into teaching practices 119
5.1.3 Implementation of PISA-motivated initiatives varies across schools 123
5.1.4 The influence of national education contextual factors on PISA-motivated initiatives 128
5.1.4.1 National curriculum’s role in PISA-motivated initiatives 128
5.1.4.2 High-stakes examinations’ role in PISA-motivated initiatives 130
5.1.4.3 Parents’ role in these effects 132
5.1.5 The local reflection on embracing PISA 132
5.2 The perceived impact of the use of PISA on mathematics learning 135
5.2.1 Perceived impact of the use of PISA on the process of mathematics teaching and learning 135

5.2.1.1 Teachers’ refreshed understanding of teaching has brought changes to teaching practices 136

5.2.1.1.1 Teaching has been transformed to highlight student orientation and cognitive activation......... 137

5.2.1.2 Improved teaching practices influence students’ learning content 139

5.2.1.2.1 Highlighting applied mathematics........................ 139

5.2.1.2.2 Highlighting mathematics-related reading 139

5.2.1.3 The impact on teaching is influenced by teachers’ involvement in PISA and PISA-motivated mathematics initiatives 140

5.2.2 Perceived impact of the use of PISA on students’ mathematics learning outcomes.. 142

5.2.2.1 Perceived impact on students’ mathematics non-cognitive learning outcomes 143

5.2.2.1.1 Perceived impact on students’ self-beliefs in learning mathematics 143

5.2.2.1.2 Perceived impact on students’ interest in learning mathematics.. 143

5.2.2.2 Perceived impact on students’ mathematics cognitive learning outcomes.. 144

5.2.2.2.1 Perceived impact of reading initiatives on students’ mathematics cognitive learning outcomes... 145

5.2.2.2.2 Perceived impact of mathematics initiatives on students’ mathematics cognitive learning outcomes... 147

5.2.3 National education context moderates the effect of PISA and PISA-motivated initiatives on mathematics learning 150

5.3 Students’ performance trends and factors associated with performance... 153

5.3.1 Students’ performance trends in PISA............................... 153

5.3.2 Multilevel analysis of Fangshan PISA data......................... 156

5.4 Summary 162

Chapter 6 Discussion.. 164

6.1 The use of PISA in the local context 164

6.1.1 Using PISA as a new perspective to review local educational quality and teaching practices 165

6.1.2 Using PISA as an impetus to improve local educational quality 167

6.2 The mechanism of PISA's impact on students' learning 173
6.2.1 Factors in local context 174
6.2.2 Factors in school context 176
6.2.3 Factors in interconnection between school context and home context 178
6.2.4 Factors in national education context 179
6.3 PISA's impact on student's mathematics learning outcomes....... 181
6.3.1 Synthesising qualitative findings about PISA's impact on mathematics learning with quantitative findings 181
6.3.2 The local mathematics performance has been improved ... 186
6.3.3 The factors contributing to the improvement of mathematics performance 187
6.3.3.1 Increasing students' exposure to reading 188
6.3.3.2 Increasing students' exposure to applied mathematics problems 189
6.3.3.3 Increasing use of student-oriented teaching approach and cognitive activation... 190
6.3.3.4 Enhancing self-beliefs and interest 191
6.4 Contributions of this research 193
6.4.1 Contributions to knowledge 193
6.4.2 Contributions to methodology... 194
6.4.3 Contributions to theories... 195
6.5 Limitations of the research... 196
6.5.1 Limitations in generalisation and transferability of findings. 196
6.5.2 Limitations in making causal inference 197
6.6 Summary 197
Chapter 7 Conclusion 200
7.1 Research implications... 200
7.1.1 Implications for researchers 200
7.1.2 Implications for assessment developers............................. 201
7.1.3 Implications for education policymakers 202
7.2 Future research 203
7.3 Methodological reflection 204
7.3.1 Timing for conducting interviews 204
7.3.2 Background of interviewees 205
7.3.3 Design of assessment instrument 205

7.3.4 Display of the process of synthesising findings 206
7.4 Personal reflexive account 206
List of References 208
Appendices 228
A.1 Interview guides 228
A.2 Phone scripts of inviting potential participants 231
A.3 Letter seeking permission from school gatekeepers 233
A.4 Background questionnaire for mathematics teachers (English version) 234
A.5 Information sheet for the local educational official (English version) 235
A.6 Consent form for the local educational official (English version) .. 238
A.7 Application for using Fangshan PISA data (English version) 239
A.8 Confidentiality statement on using PISA data (Chinese version) . 240
A.9 Confidentiality statement on using PISA data (English version) ... 242
A.10 Transcription notation 243
A.11 Memos of qualitative analysis 244
A.12 Screenshot of categorising codes 245
A.13 Ethical approval letter 246
A.14 Journal of quantitative analyses 248
A.15 The process of conceptualising PISA’s impact on students’ learning…. 249

List of Tables

Table 1.1 Mainland China's involvement in PISA 3

Table 2.1 The structure of courses in secondary education 14

Table 3.1 Hughes' washback model 58

Table 4.1 Schools and students in Fangshan (including Yanshan Subdistrict) 72

Table 4.2 The participants of interviews 75

Table 4.3 Fangshan's sample size in PISA 79

Table 4.4 Demographic composition of Fangshan's student sample in PISA 83

Table 4.5 Descriptive statistics of explanatory variables 87

Table 4.6 Patterns of missingness........ 90

Table 4.7 Bivariate correlation matrix of variables........ 92

Table 4.8 Terms appropriate to the four aspects of trustworthiness .. 97

Table 5.1 Summary of the local PISA-motivated initiatives........ 122

Table 5.2 Students' mean performance trends in PISA 154

Table 5.3 Students' mean performance trends in PISA, after adjusting for demographic composition 154

Table 5.4 Analysis of the null model........ 156

Table 5.5 Analysis of multilevel models with explanatory variables . 158

Table 6.1 Pillar integration process 183

List of Figures

Figure 2.1 Offical school ages by level of education (UNESCO, 2019).. 8

Figure 2.2 Enrolment rates by level of education 10

Figure 3.1 The growing participation in PISA 24

Figure 4.1 Mixed methods design 70

Figure 4.2 Imputation results for ANXMAT 93

Figure 5.1 Trends in the percentages of mathematics low-achievers and high-achievers 155

Figure 5.2 Adjusted trends in the percentages of mathematics low-achievers and high-achievers 155

Figure 6.1 The mechanism of PISA's impact on students' learning .. 173

Abbreviations

ARS	Acquiescence response style
BFLPE	Big-Fish-Little-Pond effect
BIB	Balanced Incomplete Block
ERS	Extreme response style
ESCS	PISA index of economic, social and cultural status
FIMS	The First International Mathematics Study
GPA	Grade Point Average
ICC	The interclass correlation coefficient
IEA	The International Association for the Evaluation of Education Achievement
ILSA	International large-scale assessment
IQ	Intelligence quotient
IRT	Item Response Theory
MAR	Missing at random
MCMC	Monte Carlo Markov Chain
MEXT	Ministry of Education, Culture, Sports, Science, and Technology
MLM	Multilevel modelling
MOE	Ministry of Education
MPR	Midpoint responding
NAEP	National Assessment of Educational Progress
NAEQ	National Assessment for Education Quality
NEEA	National Education Examinations Authority
NES	National Education Standards
OECD	The Organisation for Economic Co-operation and Development
OTL	Opportunity to learn
PE	Physical education
PIACC	The Programme for the International Assessment of Adult Competencies
PIP	Pillar Integration Process
PIRLS	The Progress in International Reading Literacy Study
PISA	The Programme for International Student Assessment
PV	Plausible value
RQ	Research question
SD	Standard deviation
SES	Socioeconomic status
SHPISA	Shanghai PISA Secretariat and Research Centre
TIMSS	The Trends in International Mathematics and Science Study
UBID	Unit-based instructional design
UNESCO	United Nations Educational, Scientific, and Cultural Organization
VIF	Variance inflation factors

Chapter 1 Introduction

This thesis presents the research on the impact of the Programme for International Student Assessment (PISA) on students' learning, and particularly investigates that impact in a local context in China. In this chapter, I will present the research background, aims, and the structure of the thesis.

1.1 Background

Early in the 1950s, with the recognition that international standards on identifying strengths and weaknesses of educational quality across nations were not available, a cross-national study was firstly proposed in Hamburg at the Institute for Education of the United Nations Educational, Scientific, and Cultural Organization (UNESCO) (Husén, 1979). Soon in 1960, the International Association for the Evaluation of Education Achievement (IEA), which was originally composed of a group of educational psychologists, psychometricians and sociologists, launched the Pilot Twelve-Country Study in 12 countries to explore the feasibility of administrating an assessment across countries of different school systems (Husén, 1979). As modern techniques and theories of educational assessment advance, international large-scale assessments (ILSAs) have been not only feasible, but also had growing impact on domestic educational policymaking in various aspects, such as pedagogy, teacher training, and instruction (Heyneman and Lee, 2014). The methodological innovations such as Item Response Theory (IRT) and Balanced Incomplete Block (BIB) item sampling design enable ILSAs to focus on more and complex questions of policy interests in education (Kirsch et al., 2013). Programmes such as the Trends in International Mathematics and Science Study (TIMSS), the Progress in International Reading Literacy Study (PIRLS), PISA, and the Programme for the International Assessment of Adult Competencies (PIACC) are typical examples of the most influential ILSAs in terms of its global reach and impact.

As to PISA, it is a triennial study launched by the Organisation for Economic Co-operation and Development (OECD) in 1997, with PISA 2000 as its first cycle. PISA assesses how well 15-year-old students are prepared and equipped to meet the challenges in their adult life by measuring their performance in reading, mathematics and science literacy (OECD, 1999). PISA also collects information such as student background, learning experience, and school system through questionnaires[1] distributed to students and school leaders. With the data of questionnaires and the cognitive test, PISA produces policy-oriented indices about student performance to promote participating education systems' understanding of their systems (OECD, 1999). In the most recent cycle, PISA 2018, about 0.6 million 15-year-old students in 79 participating education systems attended this assessment (OECD, 2019a).

Mainland China has been involved in PISA for several cycles. Early in PISA 2006, under the administration of National Education Examinations Authority (NEEA), four small areas in mainland China carried out PISA as a trial on a voluntary basis. Following that, some regions participated in PISA 2009 China Trial and PISA 2012 China Trial. Different from these areas, Shanghai formally took part in PISA 2009 and PISA 2012, which were administered by Shanghai PISA team (i.e. Shanghai PISA Secretariat and Research Centre [SHPISA]). Starting from PISA 2015, mainland China expanded its formal participation by involving four provinces/municipalities. In 2015, 9841 students from 268 schools in Beijing, Shanghai, Jiangsu, and Guangdong, aggregated as one sample representing the whole of these four regions, attended and completed this assessment (OECD, 2016a). In 2018, the sample consisted of 12058 students from 361 schools selected in Beijing, Shanghai, Jiangsu, and Zhejiang (OECD, 2019a). There is a local area, Fangshan District of Beijing, which continuously participated in PISA 2009 China Trial and PISA 2012 China Trial with its locally representative sample, and as a part of Beijing, took part in PISA 2015 with all its eligible schools. Table 1.1 (Yang, 2008; 2015; Wang et al., 2017; OECD, 2019a) below summarises the regions and sub-regions that have been involved in PISA, either through China Trials or formal participation.

[1] The optional questionnaires for teachers and parents are employed to gather more comprehensive information and also to cross-validate the questionnaire information collected from students and school leaders.

Table 1.1 Mainland China's involvement in PISA

PISA 2006	PISA 2009		PISA2012		PISA2015	PISA2018
China Trial	China Trial	Formal participation	China Trial	Formal participation	Formal participation	Formal participation
Haidian District of Beijing	Fangshan District of Beijing	Shanghai	Fangshan District of Beijing	Shanghai	Beijing-Shanghai-Jiangsu-Guangdong	Beijing-Shanghai-Jiangsu-Zhejiang
Chaoyang District of Beijing	Tianjin		Tianjin			
Tianjin	Hebei		Hebei			
Weifang City of Shandong	Jilin		Jilin			
	Jiangsu		Jiangsu			
	Zhejiang		Zhejiang			
	Hubei		Hainan			
	Hainan		Sichuan			
	Sichuan		Yunnan			
	Yunnan		Ningxia			
	Ningxia					

Since the first cycle, PISA has widely attracted the attention of educational policymakers (Sellar and Lingard, 2014). According to the survey conducted amongst PISA Governing Board representatives from the education systems that participated in PISA 2009, PISA has been playing an important role in policymaking in most of these systems (Breakspear, 2012). Policymakers use PISA findings to justify policy proposals (e.g. Australia), to drive educational evaluation and set monitoring indicators (e.g. Japan), and also to set or revise curriculum standards (e.g. Korea) (Breakspear, 2012). Besides the national activities in light of PISA, PISA has expanded its policy effect at regional level with the sub-national participation in the countries such as Canada and USA (Engel and Frizzell, 2015). Along with the policy impact on domestic education policymaking, there have been extensive debates focusing on borrowing policies from top-performers of PISA (Johansson, 2016). As we know, PISA results reports provide the international league table, which is simply generated from students' average performance of each participating education system. Since "every country has its own unique features and context" (OECD, 2016b), it is suggested to pay attention to the findings of other factors to comprehensively understand students' performance and education system (Wu, 2009; Ho, 2016). However, it seems that deep analyses of PISA data for informing educational policymaking are rarely seen (Pons, 2012). PISA's power of governing global education is also argued (Meyer and Benavot, 2013). Researchers claim that this assessment tool is governing policymaking in domestic education systems by numbers (Grek, 2009).

Despite that there have been a range of discussions on PISA's impact, most of them centre on the use and consequences of PISA in educational policymaking (e.g. Ertl, 2006), while whether the impact as claimed by policymakers has practical influence on students' learning in schools remains unclear. This is also the case in China (e.g. Wang and Tong, 2015). According to my work experience as a member of PISA national team in China, I realised national policymakers' concern of students' performance in PISA, and local PISA data consumers' interest in and attempts at using PISA for seeking solutions to address local educational issues. Therefore, I think, as to the use of PISA data, besides publishing PISA results to describe the strengths or weaknesses in performance, there should be more that we can do to advance appropriate ways to fully make use of PISA data for improving practices. To better and more appropriately use PISA data for informing educational policies and practices, it is important to enter into local context and look into how PISA data are used and translated in educational policies and their practical influence on students' learning.

1.2 Aims

My research aims to empirically investigate PISA's impact on students' learning in schools in a local Chinese context. The local context is Fangshan District of Beijing, which continuously participated in three cycles of PISA from PISA 2009 China Trial to PISA 2015 with its locally representative samples. My research attempts to first understand the overall picture of the impact of PISA on students' learning and how this impact occurs; and then further investigate in detail the impact on students' learning in a specific learning domain, mathematics. The research seeks answers to the following three research questions (RQs):

RQ 1. How is PISA used by Fangshan educational policymakers and practitioners in the local context?

- 1a. How do Fangshan educational policymakers perceive PISA?
- 1b. What kind of initiatives have been motivated by PISA due to these perceptions?
- 1c. How are these initiatives implemented at school level?
- 1d. What is the influence of national education contextual factors on the formation and implementation of PISA-motivated initiatives?

RQ 2. What are the perceptions of Fangshan educational policymakers and practitioners of the impact of the use of PISA in educational policies and practices on mathematics learning?

- 2a. What are their perceptions of the impact of the use of PISA in educational policies and practices on the process of mathematics teaching and learning in schools?
- 2b. What are their perceptions of the impact of the use of PISA in educational policies and practices on mathematics learning outcomes?
- 2c. What is the perceived role of national education contextual factors in the impact on mathematics learning?

RQ 3. To what extent have the PISA-motivated initiatives actually influenced students' learning outcomes in mathematics, as evidenced by PISA data?

RQ 1, as listed above, is addressed through seeking answers to four sub-questions. It explores the overall picture of how PISA is used in the local context, and how its data, such as PISA results, databases, assessment frameworks, and released items, are interpreted and translated in the local PISA-motivated initiatives. This exploration is an important prerequisite for investigating and understanding the perceptions of the local educational policymakers and practitioners about the impact of PISA and PISA-motivated initiatives on students' learning, which is the focus of RQ 2. RQ2 is addressed through answering three sub-questions. RQ 3 seeks evidence from quantitative data of PISA to triangulate the perceptions identified from qualitative data in the answer to RQ 2, and also to expand the understanding about the perceptions. In RQ 2 and RQ 3, mathematics is particularly focused upon. I will describe in detail all the three research questions later in Section 4.1.

1.3 The structure of the thesis

The main thesis consists of seven chapters. The present Chapter 1 introduces the background, the aims of my research, and outlines the structure of my thesis.

Chapter 2 provides the description of the education context in mainland China, including an overview of the education system, curriculum reform, and assessment system reform. It shows readers the picture of the evolution of national policymakers' conceptions of education, the development of the focus and goals of education, and the growing connection between national education policymaking and global education.

Chapter 3 gives an overview of PISA, and introduces the impact of PISA on global educational policies and in mainland China. It then discusses PISA results in relation to the factors associated with students' performance and PISA's limitations that are important to be noticed in interpreting PISA data for informing policymaking. After that, this chapter presents the conceptual framework of my research.

Chapter 4 focuses on the description of methodology of my research. It details the research aims and research questions, justifies the research paradigm, design, and methods. Following these contents, it gives details about data collection and the procedures of data analyses. Finally, ethical considerations and research trustworthiness are discussed.

Chapter 5 first presents the findings which are identified from interview data with the thematic analysis approach. These qualitative findings, displayed in two sections, demonstrate the use of PISA in Fangshan local educational policies and practices, and the perception of the local educational policymakers and practitioners about the impact of that use of PISA on students' mathematics learning. This chapter then presents the quantitative findings based on the analyses of Fangshan PISA data, demonstrating the local students' mathematics performance trends in PISA, and the factors accounting for students' mathematics performance.

Chapter 6 discusses the three key findings drawn on the results presented in Chapter 5, and how these findings are related to literature. In these discussions, argumentation on the use of PISA in the local context is made. The mechanism of how PISA's impact on students' learning occurs in the local context is then conceptualised. Based on the synthesis of both interview data and Fangshan PISA data, to what extent PISA has had impact on students' mathematics learning is discussed. Following that, discussion on the contributions and limitations of the research is made.

Chapter 7 concludes with implications of this research for researchers, assessment developers, and educational policymakers, and also presents the possibilities for future research which is related to or could extend my research. In addition, this chapter also offers my methodological reflection and personal reflexive account which might be useful for researchers doing similar research in the future.

Chapter 2 Education context in mainland China

In this chapter, I will first present the overview of the education system in mainland China, introducing the structure of the education system and its development. Then I will present the curriculum reform and the assessment system reform.

2.1 An overview of education system

In mainland China, since the “opening-up policy” and economic reform was initiated in 1978, education has been greatly developing. As documented in China’s Education Law, in its education system there are four education levels, denoting pre-primary, primary, secondary (lower and upper), and tertiary education. The official school ages of each education level are shown below in Figure 2.1.

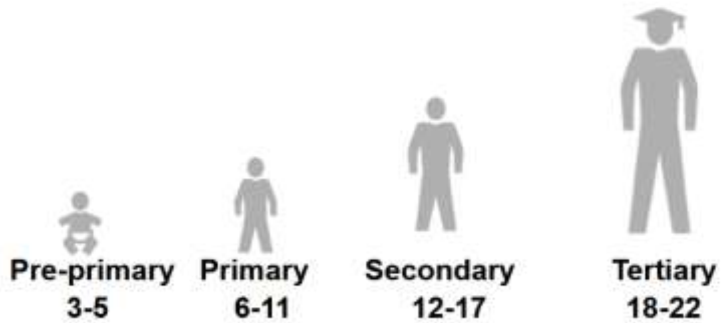
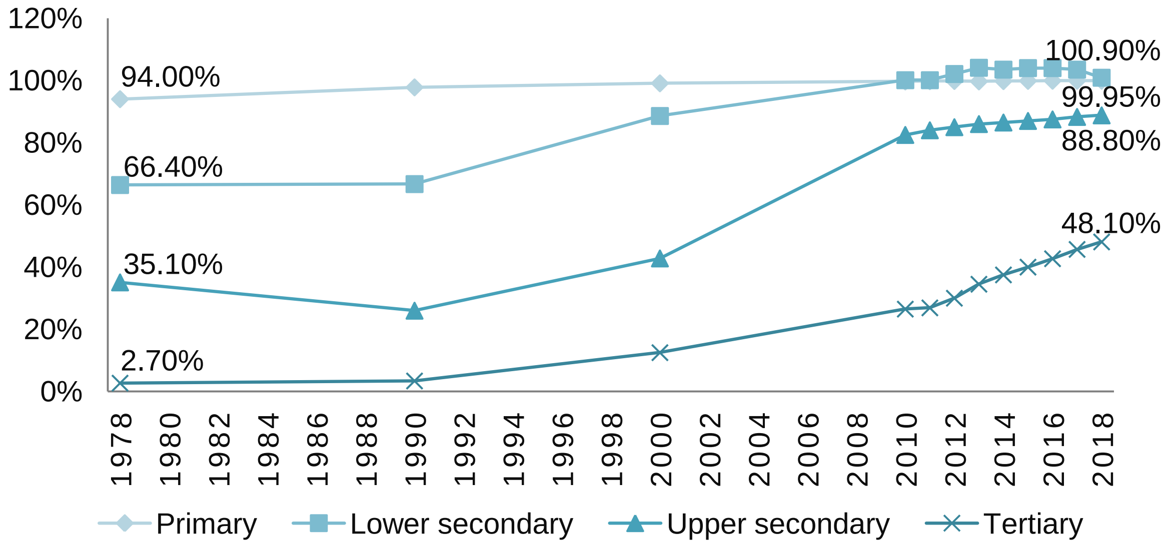


Figure 2.1 Offical school ages by level of education (UNESCO, 2019)

As indicated in the above Figure 2.1, pre-primary education lasts for three years, primary and secondary education each contains six years of schooling. Obtaining an undergraduate degree typically requires four years of study, while in some specific disciplines (e.g. medicine), five years of study may be required. In addition to the tertiary degree programmes, there are higher vocational colleges available for students which usually require three years of study. In terms of secondary education in which 15-year-olds are usually enrolled, it is comprised of three school years of lower-secondary education (from 7th grade to 9th grade) and three school years of upper-secondary education (from 10th grade to 12th grade). The total nine-year education in primary and lower-secondary is compulsory, as specified in the Compulsory Education Law which was enacted in 1986. As specified in the Education Law, the management of education in China is taken charge by the State Council and all levels of governments in regions based on the principle of hierarchical management and administrative division. Secondary and pre-secondary education are managed by local governments under the leadership of the State Council, while tertiary education is managed by the State Council and provincial governments.

Admission of students in post-compulsory education is undertaken in the national education examination system. After the completion of compulsory education, usually at 15 years old or so, students who would like to continue to receive upper-secondary education usually need to take the Senior High School Entrance Examination (known as *Zhongkao*), and choose between the regular (i.e. academic) track and the vocational track. *Zhongkao* is curriculum-based, and also serves as the graduation examination of lower-secondary education (Ministry of Education [MOE], 2002). It is administrated by provincial or city-level education departments (MOE, 2008). In upper-secondary academic education, to be qualified for graduation, students need to pass the curriculum-based Academic Achievement Test for Regular Senior High School Students. Students who would like to go to university or college are generally required to sit the National College Entrance Examination (known as *Gaokao*). *Gaokao* basically includes mandatory examinations in Chinese, mathematics, and foreign language, and examinations in elective subjects selected by students from physics, chemistry, biology, history, geography, and politics. For different provinces, the mode providing available choices to students to choose elective subjects may vary. Since over years *Zhongkao*/*Gaokao* has served as the basis for senior high school/higher education enrolment, they are considered as crucial high-stakes examinations for students, and are valued highly by them, their parents, educational policymakers, and practitioners.

The current education system in China is shaped by continuous reforms initiated since 1978, which were motivated by the need of advancing socio-economic, scientific, and technical development (Shi and Zhang, 2008). With 40 years of development, China has made great improvement in the universalisation of education, as shown in Figure 2.2.



Note: for primary education, the data are net enrolment rates; for other education levels, the data are gross enrolment rates. For the years without data points, data of enrolment rates are not available.

Data source: http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/

Figure 2.2 Enrolment rates by level of education

In 1977, *Gaokao*, which was aborted for one decade due to the Cultural Revolution, was resumed. This event is considered as the beginning of restoring the education system (Fan and Sun, 2018; Gu et al., 2019). During the period of restoring, universalising primary education in 1980's was proposed as the focus of education reform (Shi and Zhang, 2008). In response to the initiation of economic and technological system reform, *Decision on the reform of education system* (*Zhonggong zhongyang guanyu jiaoyu tizhi gaige de jueding*) was issued in 1985 to increase school autonomy and motivate the initiative of local administration of education (Shi and Zhang, 2008; Fan and Sun, 2018). This document specifies that basic education is managed under the principle of "local responsibility and hierarchical management". Along with the decentralisation of education management, the Compulsory Education law which was then enacted in 1986 further underpins the development of basic education.

In 1992, China started to establish the socialist market economy system, which drove the focus of education reform to transit to building the education system adaptable to the economy system. In 1993, the *Guideline for education reform and development of China* (*Zhongguo jiaoyu gaige he fazhan gangyao*) was therefore issued to improve the quality of citizens, and transform the heavy burden of huge population to the advantage of human resources (State Council, 1993). Regarding the hierarchical management of secondary and pre-secondary education, this document further specifies that they are managed by local governments under the guidance of the central government, and proposes the principal responsibility system (State Council, 1993). The conception of “Quality Education” was then for the first time documented in 1994 in the central government’s policy document *Opinions on further strengthening and improving moral education in schools* (*Zhonggong zhongyang guanyu jinyibu jiaqiang he gaijin xuexiao deyu gongzuo de ruogan yijian*), in which legal education, education of social morality and work ethics, physical and aesthetic education, and students’ mental quality are proposed to be improved (State Council, 1994).

The connotation of quality education is further clarified in the *Decision on deepening education reform and comprehensively promoting quality education* (*Zhonggong zhongyang guowuyuan guanyu shenhua jiaoyu gaige, quanmian tuijin suzhi jiaoyu de jueding*) which was issued in 1999, and starts to be proposed for all the four levels of education afterwards (State Council, 1999). This policy document writes that quality education aims to pave the way for students’ all-round and healthy development (State Council, 1999). It emphasises that not just intellectual education but also moral, physical, and aesthetic education shall be crucial components in educational activities (State Council, 1999). In terms of intellectual education, this document suggests to adopt the heuristic approach and discussion activities in teaching, and pay attention to developing students’ abilities of analysing and solving problems, communication, and so on (State Council, 1999). It also starts the enrolment expansion in upper-secondary and tertiary education to relieve the stress of students seeking post-compulsory education (State Council, 1999). To facilitate the implementation of quality education, this policy suggests to reform the examination and assessment system in which assessing students’ capabilities and comprehensive quality shall be underlined; and suggests to reform the curriculum system and adopt the three-level (i.e. national, local, school) curriculum model to make curriculum content more comprehensive and practical (State Council, 1999). To guarantee the implementation of quality education, provision of teacher education targeting all teachers and improving their overall quality is proposed in this document (State Council, 1999). This policy was then further developed and specified later in the same year in the *Regulations on the further education for teachers at primary and secondary schools*

(*Zhongxiaoxue jiaoshi jixu jiaoyu guiding*) that teachers' further education is administrated by city and county educational sectors under the supervision of provincial educational departments, and is mainly funded by governments (MOE, 1999).

Along with the suggestion of "the scientific outlook on development" which was raised by the former General Secretory Hu Jintao in 2003 and then added into the Communist Party of China's constitution as one of the guiding principles for social and economic development, educational reform moves its focus to the equity of education (Fan and Sun, 2018). In 2005, the *Opinions on further promoting balanced development of compulsory education* (*Jiaoyubu guanyu jinyibu tuijin yiwu jiaoyu junheng fazhan de ruoganyijian*) was issued to prevent the widening educational gaps between urban and rural areas, between regions, and between schools (MOE, 2005).

Further, in 2010, the *Outline of China's national plan for medium and long-term education reform and development (2010-2020)* (*Guojia zhongchangqi jiaoyugage he fazhan guihua gangyao (2010-2020)*) was issued "to enhance citizens' overall quality", and "boost educational development in a scientific way" (State Council, 2010). This plan specifies five principles for comprehensively implementing quality education. That is, "giving priority to development, taking the cultivation of people as the starting point, carrying out reform and innovation, promoting equity, and improving quality" (State Council, 2010). It suggests to update the concepts about talent or professional cultivation (State Council, 2010). Besides all-round development, diversity of talents or professionals is suggested, and personalised development is also encouraged to develop every student's potential, considering individual students' differences in personality (State Council, 2010). This plan also proposes to reform the enrolment and examination system, suggesting the role of this reform to be the breakthrough to change the practice that a student's achievement in one single round of examinations (i.e. *Zhongkao* or *Gaokao*) determines his or her destiny (State Council, 2010). Hence, it suggests to gradually transform this system to be with diversified evaluation and enrolment criteria (State Council, 2010). The Academic Achievement Test for Junior High School Students and the evaluation of students' comprehensive capabilities are suggested to be improved and used for the senior high school enrolment (State Council, 2010). Advancing the tertiary-level examinations and enrolment system reform in terms of its contents and forms is suggested, and assessing students' comprehensive quality and capabilities is underlined (State Council, 2010). It is suggested to explore the ways to hold several rounds of examinations per year for some certain subjects (e.g. English) (State Council, 2010). Gaokao scores which still serve as the basis for enrolment, in addition to the Academic Achievement Test for Regular Senior High School Students and the evaluation of overall quality, are suggested to be used in university/college enrolment (State Council, 2010). To consolidate the scientific basis for educational policymaking, this plan proposes to establish and improve the national educational quality monitoring and evaluation system, and release evaluation reports regularly (State Council, 2010). It also particularly suggests further promoting the openness of education through the ways such as international communication and cooperation (State Council, 2010). Advanced educational concepts and experience in the world are encouraged to learn from and inform domestic education reform, so as to enhance the international position, influence, and competiveness of China's education (State Council, 2010).

2.2 Curriculum reform

In response to the policy of promoting quality education, the *Guidelines for basic education curriculum reform (pilot)* (*Jichu jiaoyu kecheng gaige gangyao (shixing)*) which covers education levels from pre-primary to upper-secondary was issued in 2001 to develop quality-oriented basic education curriculum system (MOE, 2001a). Specifically, this curriculum reform aims to transform the curriculum focus from imparting knowledge to students to shaping students' positive and active role in learning; to set integrated courses with the consideration of the interconnection between subjects and across education levels, diversity of local contexts, and individual differences of students; and to develop students' abilities of collecting and processing information, acquiring new knowledge, analysing and solving problems, communication and team-working (MOE, 2001a).

Since this curriculum reform, a three-level (i.e. national, local, school) model of curriculum management has been adopted to adapt various needs of different regions, schools, and students (MOE, 2001a). At the national level, MOE develops national curriculum, specifies the categories of courses and class hours, and makes policies in relation to curriculum management. Based on these and regional contexts, provincial education departments may develop the local plan for implementing national curriculum and local curriculum. When implementing national curriculum and local curriculum, schools can develop or select school-based courses according to their own school conditions, students' interests and needs, under the guidance and supervision of education administration sectors of all levels.

Accordingly, the *Curriculum standards for compulsory education (experimental)* and the *Curriculum standards for regular upper-secondary education (experimental)* were released respectively in 2001 and 2003 to replace the syllabus which was used before for decades. Compared with the syllabus, the national curriculum focuses on stipulating the basic requirements for the quality of students, rather than the details of the process of teaching (MOE, 2001b). Curriculum content is organised based on three dimensions: knowledge and skills, process and methods, attitudes and values. The curriculum contains three categories of courses, namely, subjective courses, integrative courses (e.g. integrated science), and integrated practical activities (e.g. information technology, enquiry-based learning, community service, labour and technical practice). Besides mandatory courses, schools are encouraged to set elective courses. For secondary education, the structure of courses is as shown below in Table 2.1.

Table 2.1 The structure of courses in secondary education

Educational level	Mandatory/ Elective	Courses
Lower secondary	Mandatory	Ideology and morality, Chinese, mathematics, foreign language, science (or physics, chemistry, biology), history and society (or history, geography), physical education (PE) and health, art (or music, fine arts)
	Mandatory	Integrated practical activities
	Elective	Elective courses
Upper secondary	Mandatory	Chinese, mathematics, foreign language, politics, history, geography, physics, chemistry, biology, art (or music, fine arts), PE and health, technology
	Mandatory	Integrated practical activities
	Elective	Elective courses

Based on 10 years of experience of curriculum implementation and the investigation of the perceptions of principals, teachers, students and parents, the *Curriculum standards for compulsory education (experimental)* was then further improved in the *Curriculum standards for compulsory education (version 2011)*, which also incorporates the ideas of curriculum reforms in 42 countries (Yang, 2012). Under the guidance of the “outlook of scientific development” and “socialist core values”, the version 2011 of curriculum emphasises on developing students’ innovation spirit and the ability of applying knowledge to solving real-life problems (Yang, 2012). It also highlights the education of fine cultural traditions of Chinese nation.

Later, the *Curriculum standards for regular upper-secondary education (version 2017)* was released. This revised curriculum further clarifies the goal of regular upper-secondary education that it is not only for preparing students for entering into higher education, but also for preparing them for social life and career development, and laying the foundation for their life-long development. It aims to improve students’ comprehensive quality with the focus on developing their core competencies. For this end, another category of courses, that is, alternative mandatory courses, is added into the curriculum structure to further meet the needs of students’ all-round and personalised development. Regarding the core competencies, for mathematics curriculum, for example, they are specified as six core competencies, denoting mathematical abstraction, logical reasoning, mathematical modelling, intuitive imagination, mathematical operations, and data analyses. Similar to the *Curriculum standards for compulsory education (version 2011)*, this revised curriculum also underlines strengthening the education of China's fine traditional culture (MOE, 2018). It is important to note that, in this revision, curriculum experts as well as examinations and assessment experts were involved to allow for strengthening the link between the new curriculum and the ongoing *Gaokao* reform (Liu, 2018).

2.3 Assessment system reform

It is widely acknowledged that the examination and enrolment system in China has made considerable contribution to the improvement of the quality in education, and also has been playing an important role in improving social mobility (Liu, 2000; Dai, 2007). However, it has been criticised for using scores in high-stakes examinations as the single measure evaluating students and even schools and teachers, which has brought issues, for example, examination-oriented education which impedes students’ all-round development (Dai, 2007). Over the years, students’ achievement in

Zhongkao or *Gaokao* is generally the single measure determining whether they could go to a senior high school or a university. Furthermore, in some areas, students' progression rates are used as an important indicator in the evaluation of schools and teachers (Wu, 2008). Although quality education has been promoted along with curriculum reforms for years, it still has not been well implemented in educational practices, due to the resistance of various stakeholders (Wu, 2008).

To promote quality education, the *Circular on actively promoting reforms in the evaluation and examination system in primary and secondary education* (*Jiaoyubu guanyu jiji tuijin zhongxiaoxue pingjia yu kaoshi zhidu gaige de tongzhi*) was issued in 2002, in which the aim of assessment in primary and lower-secondary education is specified as developing students' comprehensive quality and teachers' teaching quality (MOE, 2002). It stipulates that the examination items of *Zhongkao* must be developed based on national curriculum (MOE, 2002). With consideration both of students' overall quality and individual differences, it suggests to explore the way of conducting multiple evaluation in the enrolment of senior high schools (MOE, 2002).

In 2008, the *Opinions on further promoting and improving Zhongkao Reform* (*Jiaoyubu guanyu shenru tuijin he jinyibu wanshan zhongkao gaige de yijian*) was issued to motivate nationwide implementation of quality education, and improve the development of education equity (MOE, 2008). To address the issue that, in some areas and some schools, scores in *Zhongkao* are still used as the single measure in enrolling students, and the progression rate is used as the single measure evaluating school effectiveness, this policy proposes to gradually replace scores with grades for rating students' performance in *Zhongkao* (MOE, 2008). Regarding assessment content, it emphasises students' mastery of basic knowledge and ability of analysing and solving problems in contexts (MOE, 2008). As stated in this document, the subjects included in *Zhongkao* are determined by provincial or city-level education departments (MOE, 2008).

Since 2016, *Zhongkao* has been combined with the Academic Achievement Test for Junior High School Students, starting to serve not only for senior high school enrolment but also as the graduation examination, to lighten students' heavy academic burden in terms of preparing examinations. As stated in the *Opinions on further promoting reforms in the examination and enrolment system in upper-secondary education* (*Jiaoyubu guanyu jinyibu *tuijin gaozhong jieduan xuexiao kaoshi zhaosheng zhidu de gaige yijian*), the examination and enrolment system involving performance in the Academic Achievement Test for Junior High School Students in addition to comprehensive quality evaluation is expected to be established in 2020 (MOE, 2016). To further address the unintended consequences of assessment, the MOE continues to make efforts through the item development in the academic test with its recently-released policy *Opinions on strengthening the item development of the Academic Achievement Test for Junior High School Students* (*Jiaoyubu guanyu jiaqiang chuzhong xueye shuiping kaoshi mingti gongzuo de yijian*) (MOE, 2019). This policy stipulates the abandonment of the Specifications of the Academic Achievement Test for Junior High School Students, and requires that all subjects except integrated practical activities, which are specified in the curriculum for compulsory education, shall be included in the academic achievement test (MOE, 2019). It suggests the use of assessment for guiding teaching in terms of motivating teachers to adopt teaching approaches employing contexts and motivating students' deep thoughts and high engagement (MOE, 2019). Regarding developing items, it proposes to improve the design of contexts with full consideration of the real life of students and the differences between urban and rural areas, and employ contexts typical and relevant for students (MOE, 2019).

Gaokao, as a selective examination, mainly serves for enrolling students into universities, rather than comprehensively evaluating students' achievement in meeting curriculum requirements. Therefore, although *Gaokao* examination papers are developed with reference to the curriculum, it is admitted that only a part of curriculum content is covered by *Gaokao* (Yang, 1989). To compensate for this, the Academic Achievement Test for Regular Senior High School Students was launched in 1991 by the MOE, serving as the graduation examination assessing the extent to which students meet the requirements of the curriculum and examining whether they are qualified for graduation (Han, 2001). Despite that multiple evaluation has been suggested for years for university enrolment, *Gaokao* is still widely and generally used as the single measure in enrolment practices. In some areas, the rate of progression to higher education is even used for evaluating school effectiveness and teaching quality, and linked to the track record of leaders in local education sectors (Wu, 2008). Because of its crucial role and high stakes for students, parents, educational policymakers and practitioners, *Gaokao* unintentionally drives examination-oriented teaching in senior high schools such that teachers only teach those specified in *Gaokao specifications* and *Gaokao oulines* (Xia, 2004). Its negative influence on teaching and learning has been argued (Zheng, 2002).

Gaokao reform has experienced three stages, respectively focus on knowledge, abilities, and comprehensive assessment (Yu, 2017). In the first few years after *Gaokao* was resumed in 1977, with the influence of the lack of expertise in educational measurement and examination development, and the knowledge-focused teaching method at the time, *Gaokao* mainly assessed candidates' mastery of basic knowledge (Liu, 2014). This assessment focus is then criticised that it fosters rote learning (Yu, 2017). To address this issue, *Gaokao* moved to focus on assessing abilities. The MOE of China made reference to examinations in other countries, and started *Gaokao* reform of promoting standardised examination in 1985 (Yu, 2017). *Gaokao specifications* and *Gaokao outlines* for each subject were therefore developed afterwards (Yu, 2017), in which the constructs, components, and levels of abilities, and the approaches for assessment are specified and described (Han, 2001). Examination items are developed based on the table of specifications which include two dimensions, namely, content and ability, to assure appropriate coverage of knowledge and skills (Liu and Zhang, 1986). In response to the *Curriculum standards for regular upper-secondary education (experimental)*, *Gaokao* has gradually integrated the newly-added curriculum content and ability dimensions into assessment (Ren and Chen, 2013).

In recent years, *Gaokao* reform has stepped into the stage of multiple evaluation. In 2014, the State Council promulgated the *Implementation opinions on deepening reform of examination and enrolment system* (*Guowuyuan guanyu shenhua kaoshi zhaosheng zhidu gaige de shishi yijian*) and started the new round of *Gaokao* reform to foster integrity and all-round development of students, and to guide teaching. This reform which targets both the content and the form of *Gaokao* is considered as the most comprehensive and systematic one ever since 1977 (Jiang, 2017). Actively making use of the guiding role of *Gaokao*, it aims to motivate the implementation of the overall education reform for "facilitating students' all-round and personalised development, enhancing students' comprehensive quality and schooling quality, improving talent selection and the equity of education" (Jiang, 2017, p.3). In terms of the content reform, this document stipulates that *Gaokao* shall be based on national curriculum and demands of talent selection for universities, and it focuses on assessing students' ability of thinking individually and the ability of analysing and solving problems with the knowledge they have acquired (State Council, 2014). It specifies that students' total scores shall consist of those in the *Gaokao* examinations in Chinese, mathematics, and foreign language, and those in the three subjects students choose for the Academic Achievement Test for Regular Senior High School Students. It allows students to sit the foreign language examination for two times and use the higher score obtained. In terms of the enrolment system, it pushes forward the establishment of a multiple assessment system involving not just *Gaokao* but also the Academic Achievement Test for Regular Senior High School Students and student files of comprehensive quality. This reform started piloting in Shanghai and Zhejiang immediately in 2014, followed by further piloting in Beijing, Tianjin, Shandong, and Hainan in 2017. To provide the theoretical framework for conducting the new *Gaokao*, the NEEA has established China's *Gaokao* assessment system with the incorporation of the ideas in the new curriculum, domestic and international research outcomes on educational development and assessment (MOE, 2020). It specifies "essential knowledge, crucial abilities, competencies of disciplines, and core values" as the four categories of assessment content (MOE, 2020). In terms of assessing students' competencies of disciplines, contexts in addition to knowledge and skills are built into the dimensions of examination items (Li et al., 2019).

To inform national assessment system reform, mainland China started to conduct a PISA Trial in 2006 in four small areas (Wang, 2009) (see Table 1.1 in Section 1.1). This PISA Trial was administrated by the NEEA, which also administrates *Gaokao* at national level and is the core department undertaking national assessment system reform. After PISA 2006 China Trial, NEEA continued to conduct PISA 2009 China Trial in Fangshan District of Beijing and ten provincial regions, and then carried out PISA 2012 China Trial in almost the same regions (Wang et al., 2017) (see Table 1.1 in Section 1.1). As stated by Lei Wang (2007; 2009) who was responsible for national administration of PISA China Trials and was national project manager of PISA 2015, the aims of conducting PISA China Trials are to learn from the advanced assessment concepts, theories and techniques employed in PISA for building a large-scale educational assessment system suitable to national conditions; and to facilitate the reforms in the content and form of national assessment. Because of these aims, PISA China Trials were administrated in line with most of the technical standards of formal participation except for some aspects. To be specific, PISA China Trials targeted 15-year-old students in academic track, while those in vocational track were excluded; results reports were generated by the PISA team in NEEA rather than the OECD; and data were not submitted to the OECD (Wang and Jing, 2013). Besides PISA China Trials, Shanghai formally participated in PISA 2009 and 2012. In PISA 2015 and PISA 2018, China expanded its formal participation by involving four provincial regions (see Table 1.1 in Section 1.1). Assessment system reform still gains insights from PISA data through China's formal participation in PISA (See Section 3.3.1).

In summary, this chapter depicts the overall picture of national education context in mainland China by introducing the education system and reforms of curriculum and assessment. The education system has been gradually evolving to focus on all-round as well as personalised development of students. During this process, the decentralisation of management in basic education has increased, which situates the understanding that local educational sectors, to some extent, have the authority of taking locally-adaptive policy initiatives for improving local education. Meanwhile, the system tends to show greater openness of education in terms of embracing advanced educational concepts in the world and participating in international education activities. In this context, the reforms in national curriculum and national assessment system have incorporated some concepts from other countries and international studies such as PISA, in addition to keeping and developing domestic-featured contents.

Chapter 3 Literature Review

This chapter contains seven sections. Section 3.1 introduces PISA in terms of the background of its initiation, its policy-driven design, its data that can be used for informing policymaking, and its expansion of global reach. Section 3.2 reviews the literature on discussing the impact of PISA on educational policies to explore the features and issues in the impact that have been widely documented. Section 3.3 particularly reviews the literature on the use and impact of PISA in mainland China. Section 3.4 and 3.5 respectively discuss PISA results about factors accounting for students' performance and PISA limitations, both of which are important for consideration in making policy initiatives in response to PISA, and in evaluating PISA's practical impact. Based on the literature review and my research focus, in Section 3.6, a conceptual framework is developed for framing the methodological design of my research, informing data analyses, conceptualising and interpreting my research findings. Section 3.7 provides a summary for the chapter.

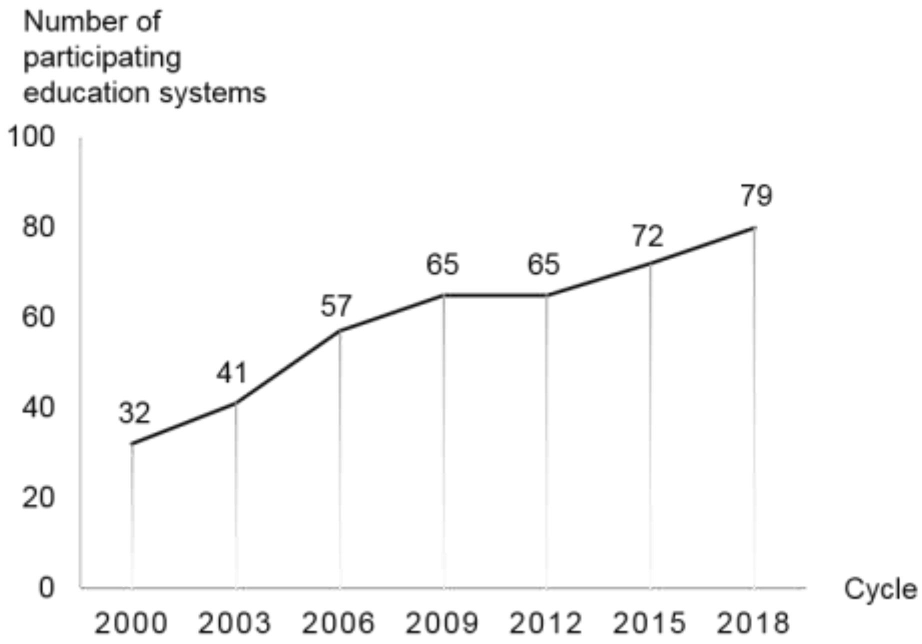
3.1 An overview of PISA

As briefly introduced in Section 1.1, PISA is a triennial study launched by the OECD where representatives of member countries work together to discuss and address challenges in globalisation in various policy areas such as economics and education (OECD, 2016c). As Brown et al. (1997, p.8) write, "the competitive advantage of nations is frequently redefined in terms of the quality of national education and training systems judged according to international standards". In response to the OECD member countries' demand of detecting their potential global competitiveness (Sellar and Lingard, 2014), in 1997, the OECD launched PISA. PISA assesses 15 year-old students' performance in reading, mathematics, and science literacy for evaluating "the degree of preparedness of young people for adult life and, to some extent, the effectiveness of education systems" (OECD, 1999, p.11). PISA is an age-based programme. To be specific, it targets students aged from 15 years and 3 months to 16 years and 2 months, who are enrolled in 7^{th} grade and higher, considering that most students in their last year of compulsory schooling are 15 years old in OECD countries (Adam and Wu, 2002).

In each cycle, PISA has one majority assessment domain which shares almost 2/3 of the total testing time (OECD, 2016d), and the other two are minority domains. Reading, mathematics, and science alternatively act as the majority domain along the cycles since PISA 2000. Although the three assessed domains correspond to school subjects, PISA does not intend to examine to what extent students master the curriculum, instead, it assesses students' acquisition of knowledge and skills in these domains which will be necessary for adult life, and their ability of applying what they have acquired to the situations they may encounter in future life (OECD, 1999). With this purpose and focus, PISA introduces "contexts" into assessment materials to evaluate how well students could apply their knowledge and skills in personal, social, and global contexts (OECD, 2016d). Since the initiation of PISA, interests of policymakers had already been considered during the development of its framework. It is "a collaborative effort, bringing together scientific expertise from the participating countries and steered jointly by their governments on the basis of shared, policy-driven interests" (OECD, 1999, p.3). As claimed by the OECD, PISA produces policy-oriented indicators about student performance, which are internationally comparable, to promote participating countries' and regions' understanding of the relative strengths and weaknesses of their own education systems, and inform their educational policymaking (OECD, 1999; 2013a). These indicators are framed based on a school effectiveness model which includes three categories of factors predicting cognitive performance: input (e.g. students' demographic characteristics, socioeconomic status), processes (e.g. opportunity to learn, teaching practices, teaching quality), and outcomes (e.g. non-cognitive outcomes such as motivation and self-beliefs) (Purves, 1987; Scheerens, 1990; Scheerens and Bosker, 1997; OECD, 2013a). In each cycle, most of the indicators about the processes category and the non-cognitive outcomes category focus on the majority domain in that cycle. In addition to the cognitive test, questionnaires for students and school leaders are typically used in PISA to collect information about the effectiveness of school systems so as to construct the policy-oriented indicators allowing for interpreting and explaining students' performance (OECD, 2016e).

In each cycle of PISA, a range of data, including cognitive test and questionnaire databases, released items, assessment framework, and reports, that can be used for informing educational policymaking, are generated. PISA items in the cognitive test are developed by PISA international contractors and participants under PISA assessment framework, most of which are recycled across the rounds of PISA. Therefore, PISA item bank is not available for public. However, some of the items have been released and are not going to be reused by PISA, from which public could have a general idea of what PISA looks like. Questionnaires are released by the OECD and available for public use after the administration of each cycle. The PISA assessment framework is available for the public as well. In each cycle of PISA, performance of each participating education system is transformed onto PISA international scale (OECD, 2014a). For each assessment domain, the international scale, which was established in the cycle when the domain was assessed as the focus for the first time, has the mean=500 and standard deviation (SD)=100 for the average of OECD countries (OECD, 2014a). Questionnaire indicators are scaled with mean=0 and SD=1 for OECD average (OECD, 2014a). Hence, PISA allows international comparison of education systems against its international scale. It also enables participating education systems to track their performance trends, with the approach of employing common items linked the assessments across cycles, and equating performance of participating education systems over cycles onto the same scale. In each cycle, PISA results reports generated by the OECD typically release international rankings of students' average performance of participating education systems in each assessment domain, describe average performance of different student subgroups (e.g. gender) and performance trends across cycles, and display the findings about the factors related to students' performance.

Education systems that participate in PISA typically employ two-stage sampling. In pre-2015 cycles, for each country, a minimum sample of 150 schools and 35 students in each school was required (OECD, 2014a). The student sample is randomly drawn from the selected schools (OECD, 2014a), rather than being obtained by sampling classes. Since PISA 2015, 42 students are needed to be selected from each sample school for the participating education systems which choose computer-based assessment mode (OECD, 2016a). Although PISA was originally launched within OECD countries, it has attracted participation globally (OECD, 2006). The number of its participants has been continuously growing since its first cycle in 2000, as shown in Figure 3.1 below. Its extension programmes, PISA for Development and PISA-based Test for Schools respectively further extend its reach to wider developing countries and directly to schools (Lewis et al., 2016; Addey, 2017; Lewis, 2017).



Data source: PISA results reports released by the OECD

Figure 3.1 The growing participation in PISA

PISA's influence is not only reflected in the respect that it has attracted worldwide participation, it has also stirred global educational policy discourse. In the next section, I will review the literature on its impact on domestic education policies.

3.2 The impact of PISA on educational policies

According to the survey on representatives of PISA Governing Board from 65 education systems that participated in PISA 2009, to various extents, PISA has had impact on national education policies across these systems (Breakspear, 2012). As introduced in last section, PISA results usually include international rankings of participating education systems, their students' cognitive performance in reading, mathematics and science, performance trends over cycles, and also explore factors associated with students' performance. Amongst these aspects of content, international rankings, students' cognitive performance and performance trends have played a more critical role in influencing domestic education policies in education systems that participated in PISA (Breakspear, 2012). Some countries, for instance, Australia, even officially declared the goal of reaching the top echelon in PISA international rankings (Gorur and Wu, 2015); while in Italy, objectives of reducing the percentages of low reading performers and low mathematics performers in PISA have been officially proposed (Damiani, 2016). In the following, I will review the policy responses to PISA findings that vary and change in domestic education systems, policy borrowing that emerges in some of these responses, and debates on the governing power of PISA.

3.2.1 Policy responses to PISA findings vary across education systems

In terms of policy responses to international rankings, Steiner-Khamsi (2003, p.2) suggests three extreme types:

> Scandalization (highlighting the weaknesses of one's own educational system as a result of comparison), glorification (highlighting the strengths of one's own educational system as a result of comparison), and indifference.

Policy responses in Germany are a typical example of scandalization, which have been extensively discussed (e.g. Ertl, 2006; Grek, 2009; Neumann et al., 2010; Kelly and Kotthoff, 2017; Niemann et al., 2017). In PISA 2000, the gap between its below-average performance and its previously perceived world-leading education system brought Germany national “shock” (Ertl, 2006; Martens and Niemann, 2010; Ringarp, 2016). Besides the unexpected performance, PISA 2000 results also indicate that Germany was one of the countries where students’ performance was closely associated with socio-economic and migration factors (Ertl, 2006). Because of the “shock”, a number of measures taken by the Federal Ministry of Education and Research have been triggered (Ertl, 2006). The most significant initiative is the introduction of National Education Standards (NES) to secondary education (Ertl, 2006), in which the literacy as defined in PISA framework is incorporated (Neumann et al., 2010). NES serves as the basis of developing an assessment system for the end of secondary education for continuously monitoring the quality of education (Klieme et al., 2003, cited in Ertl, 2006). Ertl (2006) argues that these reform actions only address certain aspects of education, while little attention was paid to the issues in other aspects such as early selection of students.

In contrast to the hot debate aroused by PISA in Germany and Germany's policy responses to its PISA results, the responses to PISA results in USA are historically indifferent until in PISA 2009 in which Shanghai-China ranked on the top (Lingard and Lewis, 2017). According to Martens and Niemann (2010), USA's indifference in its PISA results could also be explained by the gap between the perception of performance and the results in empirical assessments. Different from Germany, the public in the USA has already had awareness of its relatively poor educational performance through domestic assessments prior to PISA (Dobbins and Martens, 2010).

In Denmark which had performance from PISA 2000 to PISA 2006 placed on the OECD average or even below the average, its unsatisfactory scores and rankings have catalysed a number of policy initiatives targeting Danish compulsory school system (Egelund, 2008; Dolin and Krogh, 2010). Amongst these initiatives, the most significant one is the introduction of national tests (Dolin and Krogh, 2010). Despite the acknowledgement that PISA results have facilitated "a necessary and useful development" in Danish education in terms of arousing the attention to the value of evaluation in elementary schools, Dolin and Krogh (2010) caution that there would also be a risk of test-driven instruction which would bring serious consequences to the entire educational system.

Japan, a country that traditionally performs well in PISA is also concerned about its decline in PISA international rankings (Nakayasu, 2016). It is known that Japan prompted education reforms by redirecting the *yutori* (low pressure/no cramming) curricula and introducing national achievement testing in response to its declined performance in PISA 2003 compared with PISA 2000 (Breakspear 2012; Chung, 2016). In the revised curricula, PISA-type competencies were decided to be incorporated to improve students' academic performance, and PISA-type items are being employed in the national testing (Breakspear 2012). Nevertheless, articles also question this point of view that these reforms were directly catalysed by PISA 2003 results. The unpopular *yutori* curricula which were announced in 1998 and started implementation in 2002 had experienced criticism since its announcement (Takayama, 2008). Takayama (2008) points out that before the release of PISA 2003 results, Japan's Ministry of Education, Culture, Sports, Science, and Technology (MEXT) had already taken consideration of both rechannelling the *yutori* and introducing the national achievement testing. He argues that under the pressure from the public concern and the Prime

Minister’s Office, MEXT treated the moderately declined performance in PISA 2003 as a crisis and used it as an external resource to legitimise the reforms (Takayama, 2008). As commented by Takayama (2008), without deeply interpreting its PISA data, these reforms are more like “political necessity”.

3.2.2 Policy responses to PISA findings change over PISA cycles

National policy responses to PISA are not immutable over PISA cycles. According to the changes of national rankings in PISA or the changes in domestic contexts, the reactions of policymakers at national level can evolve. For example, in the UK, although its data in the first two rounds of PISA could not be used for comparison due to UK samples’ low response rate in the assessment (OECD, 2014b), it is considered that UK did well compared with other participating education systems, and no concrete initiatives were steered at national level as responses to PISA results (Grek, 2009). However, this has been changing since the release of PISA 2009 results in 2010 which has led to a growing recognition of high-performing countries' education policies (Knodel and Walkenhorst, 2010; Knodel et al., 2013). In PISA 2009, the slipped rankings of UK motivated policymakers to take actions on the reforms of schools. As the then Education Secretary Michael Gove wrote, PISA 2009 is considered “as a spur to action” and “provides clear pointers to how we can reform our schools system to make it one of the best in the world” by learning from the education systems of PISA high-performing countries (Gove, 2010). As claimed by Gove (2011), based on the perceived characteristics of education systems successful in PISA, the schools white paper *The importance of teaching* was launched to raise the quality of teaching, enhance school autonomy, and improve accountability mechanisms. However, Baird et al. (2011) argue that the relationship between some content of these policies in the white paper with PISA findings is not obvious.

For France, Pons (2016) classifies the evolution of French policy debate on PISA between 2001 and 2014 into three periods, and argues that this evolution is brought by the changes in domestic policy context. In the first period from 2001 to 2004, PISA results were collectively ignored due to domestic policy configuration, and there was only narrow debate centring on the statistical limits of PISA methodology (Pons, 2016). In the second period from 2005 to 2010, the policy context changes under the implementation of Lisbon strategy (Lisbon European Council, 2000) which encourages member countries to identify objectives and benchmarks for their education systems, the debate focus therefore moved progressively to political discussion on educational reforms and expected education system (Pons, 2016). In the third period, political speeches increased, in which PISA was frequently used by policymakers to legitimise their ideas or initiatives (Pons, 2016).

3.2.3 The phenomenon of policy borrowing

Amongst the policy actions or reforms catalysed by PISA or proposed under the context that unsatisfactory PISA results were received, sometimes policy borrowing or learning from those in PISA high-performing education systems is involved (e.g. Gove, 2011; Forestier et al., 2016). As denoted by Phillips and Ochs (2003, p. 451), “borrowing” covers “whole range of issues relating to how the foreign example is used by policymakers at all stages of the processes of initiating and implementing educational change.” PISA high-performers, such as Finland and some countries and regions in Asia, therefore are framed as “reference societies” for the education reforms (Sellar and Lingard, 2013a).

For example, the initiatives in Germany made references not only to PISA itself but also to other countries such as Finland (Etrl, 2006; Ringarp, 2016). In the German context, the education reforms focusing on national educational standards was justified by the Federal Ministry of Education that countries successful in PISA widely have had national educational standards, and the Ministry specifically investigated the development and application of educational standards in these countries (Ertl, 2006). When borrowing policies from elsewhere, domestic policymakers tend to use foreign examples selectively, conditioned by domestic factors such as their prior perceptions of reference countries and economic relations with these countries (Waldow et al., 2014; Waldow, 2017). For Germany, Finnish education, as opposed to Asian education, is considered as a positive model and more acceptable (Waldow et al., 2014; Waldow, 2017). Finland’s success in PISA is interpreted based on the historically-rooted positive image that Finnish education is child-centred and encourages intrinsically-motivated learning (Waldow, 2017). Policy debates in some other contexts (e.g. Australia, UK, USA) tend also to “look East”, with the perceptions that Asian economies are significant economic competitors (Sellar and Lingard, 2013b; Waldow et al., 2014). For example, as reflected in the book edited by Tucker (2011), *Surpassing Shanghai: An agenda for American education built on the world's leading systems*, USA has been drawing lessons from

Asian high-performing systems such as Shanghai-China, Japan, and Singapore since PISA 2009.

On the other side of policy borrowing is policy lending. According to the survey on the policy impact of PISA conducted by the OECD (Breakspear, 2012), education systems with high performance are frequently cited that they were made reference to by other systems in policy reforms. For Finland which has had gained worldwide admiration for its success in PISA, with this reputation, Finnish government even has taken initiatives of making its education a nation brand to promote education exportation (Schatz et al., 2017). “PISA tourism” (Chung, 2010) emerges in Finland that many educational policymakers and practitioners of other education systems go to Finland for learning something.

It is argued that internationally comparable data, as provided by PISA, have become an important policy technology, which makes policy borrowing and lending more prevalent than they used to be (Steiner-Khamsi and Waldow, 2012). Along with the globalised phenomenon of policy borrowing that countries and regions search for “best practice” of education elsewhere (Kamens, 2013), debates and critiques of policy borrowing have been raised (e.g. Oates, 2011; Auld and Morris, 2016). It is admitted that crudely transplanting others’ policies without the consideration of contextual differences is problematic (Auld and Morris, 2016). Yet, rather than genuinely import policies from other context and apply them to one’s own context, there have been cases that high-performing education systems and PISA are used by policy-makers to legitimise their own pre-shaped ideas with regard to educational reforms (Pons, 2012; Chung, 2016). In this case, the claimed policy impact is considered as “political rhetoric” (Baird et al. 2016), and it may not have an explicit link with those claimed references (Morris, 2015).

3.2.4 The governing power of PISA on global education

Along with the widely claimed impact of PISA on domestic education policymaking in a range of education systems, PISA’s power of governing global education is argued (Grek, 2009; Meyer and Benavot, 2013; Sjøberg, 2015; Fulge et al., 2016; Teltemann and Klieme, 2016). PISA seems to bring pressure to participating education systems (Lockheed and Wagemaker, 2013) and governs their educational policymaking by numbers (Grek, 2009), due to the international league tables of the performance of participating education systems published in PISA results reports. As reviewed above, some countries even made policies to achieve top rankings on PISA international league tables (Gorur and Wu, 2015). To this end, researchers consider that PISA has brought the consequences of prioritising achievement results in schooling (Engel and Rutkowski, 2014). Furthermore, since various education systems tend to follow the educational concepts suggested in PISA and the policy experience of the systems which performed well in PISA, it is argued that PISA has been contributing to policy convergence across education systems (Bieber and Martens, 2011). Hence, some researchers even raise the caution that PISA has also brought the consequence of narrowing the understanding of the definition of education (Riley and Torrance, 2003; Labaree, 2014).

Despite the concerns about PISA's power of governing domestic education, there is also voice suggesting that such power may not be as strong as assumed, and can be diluted in domestic contexts. For example, Michel (2017) argues that policy convergence brought by PISA is rather limited, considering historical, sociological, and cultural diversity of domestic contexts among education systems. Carvalho and Costa (2015, p.640) further suggest that, during the process of translating PISA concepts in domestic education systems, "reinterpretation, de-contextualisation, and re-contextualisation" are involved, in which "national, local, regional and international agencies intertwine". Similarly, Tikly (2015) considers that international assessments' impact on students' learning is mediated by the structures and mechanisms at other system levels.

Moreover, when the policy initiatives are put into practice through school implementation, contextual factors at school level are also involved in the mechanism of PISA's impact. Researchers therefore suggest that, this process is "enactment" rather than implementation of policies, in which policies are further recontexualised in schools (Braun et al., 2010; Braun et al., 2011). For example, more policy actors, such as school leaders and teachers, can play a role in influencing the interpretation and translation of the policy initiatives (Braun et al., 2010; Braun et al., 2011). Teachers are even considered as the final policy brokers during school enactment; policies would not affect what and how students learn at schools unless teachers change their teaching practices (Cohen and Ball, 1990; Spillane, 1999). To what extent teaching practices are changed in response to policy initiatives is subject to available resources, teaching and learning challenges in specific school settings (Braun et al., 2011), teachers' perceived power of policies (Schwille et al., 1982), and teachers' will as well as capacity for enacting policies (McLaughlin, 1990). Therefore, although PISA's policy impact has been discussed and documented in a large body of literature, it is possible that this impact only remains in policy speeches or texts, and has brought few actual changes to educational practices in schools (Niemann et al., 2017). According to literature as reviewed above, despite that there have been a range of discussions and debates on PISA's impact in domestic education systems, and the unintended consequences PISA may have brought, few of them looked further into local contexts and empirically investigated PISA's practical impact on students' learning in schools.

3.3 The use and impact of PISA in mainland China

Moving to the context in mainland China, PISA has also been becoming influential in domestic policymaking and educational practice, as claimed by educational policymakers. In the following, I will review the use and impact of PISA at national level as well as regional level in Chinese context.

3.3.1 The use and impact of PISA at China national level

Perhaps, because mainland China only has participated in PISA with student sample representing some regions rather than the whole country, it seems that China's performance in PISA has not explicitly motivated policy actions at national level. Rather, China tends to use more of the framework and techniques of PISA for informing the assessment reform in terms of improving the contents and forms of assessment (Wang, 2007; 2009).

During the participation in PISA, PISA national team in NEEA actively explore the feasibility of translating PISA concepts and techniques (especially those about item development, scoring, management and administration of large-scale assessments) into national education examinations such as *Gaokao* (e.g. Wang and Tong, 2015; 2016). With the efforts made by colleagues in other departments in NEEA, PISA national team's research outcomes have been starting to be applied in the practice of assessment. For example, inspired by PISA, real-life contexts across a range of subjects have been employed as the basis for developing materials in *Gaokao* (Wang and Tong, 2015). It is considered that, in the recent *Gaokao* reform, the inclusion of the dimension "contexts" is in line with the assessment concept in PISA (Li et al., 2019). As to reading, PISA features practical reading by considerably employing non-continuous texts. It is considered that PISA 2009, in which reading was the majority domain, has heated up domestic attention to developing and assessing students' ability of reading non-continuous texts (Ye, 2017). In 2017, the Chinese examination in *Gaokao* involved the use of multiple texts of practical reading (Zhao, 2017).

At national level, PISA's impact is not only limited to influencing the examinations developed by the NEEA. PISA framework has been an important reference in the reform of national educational assessment. Over years, as I introduced in Section 2.3, quality in primary and secondary education was usually evaluated only based on students' academic achievement and schools' progression rates. To reverse this practice which hinders students' all-round development, in 2013, MOE issued the *Opinions on promoting reforms of comprehensive evaluation of the quality in primary and secondary education* (*Jiaoyubu guanyu tuijin zhongxiaoxue jiaoyu zhiliang zonghe pingjia gaige de yijian*) to reform the evaluation system for primary and secondary education (MOE, 2013a). As clearly stated by MOE, making reference to "some concepts of PISA" in addition to Shanghai's experience of educational assessment reform, this policy suggests that assessment shall focus on student development as the measurement of educational quality, and collect data for identifying potential factors influencing educational quality (MOE, 2013b).

National Assessment of Education Quality (NAEQ) which started pilots from 2007 for monitoring basic education was formally launched by MOE in 2015. PISA, TIMSS, and USA's National Assessment of Educational Progress (NAEP) are important references for its development (MOE, 2015). NAEQ is even referred to as "Chinese PISA" (Chai and Xiang, 2015), although it is national curriculum-based (Song, 2015), which is different from PISA. Tao Xin, the deputy director of National Assessment Centre for Education Quality, the authority taking charge of the development and administration of NAEQ, argues that NAEQ technical standards are highly aligned with those of PISA (Chai and Xiang, 2015). Song (2015) states that the development of NAEQ assessment instruments has embraced the concepts and measurement techniques of international standardised instruments such as PISA and TIMSS, with national features incorporated. Like PISA, NAEQ also focuses on assessing students' ability to comprehensively applying knowledge to solve problems, embeds real-life contexts, and uses common items to link different forms of booklets (Song, 2015).

Despite the appreciation of ILSAs (especially PISA) in the development of national educational assessments, reflective learning from ILSAs is suggested. Although Limin Liu, then Deputy Minister of Education, admits that national assessment of basic education quality has learned from assessments developed by international organisations, he argues that neither TMISS nor PISA is totally appropriate for addressing the issues in Chinese context, and therefore their standards should not be copied directly to assess our education (MOE, 2012).

In terms of the interpretation and use of PISA results, it is admitted that the performance of Shanghai does not represent that of China, and performance in PISA could not reflect the whole education system, considering PISA's target population and assessment domains (Zhang and Yan, 2015). Despite that, self-confidence in China's own education system has been increased to some extent with Shanghai's success in PISA (Zhang and Yan, 2015; Yu et al., 2017). China's performance is used as evidence to support policymakers' claims about the improvement of quality in education in China. On one of the conference series reporting the achievement during the 12th Five-Year Plan period (2011-2015), Guiren Yuan, then Minister of Education, reports that Shanghai was continuously placed top in PISA, which indicates the steadily increasing international influence of China's education system (Dong, 2015). Yandong Liu (2017), then Vice-Premier, states that China (B-S-J-G)'s higher performance compared with the OECD average in PISA 2015 reflects the improvement of educational quality in China.

3.3.2 The use and impact of PISA at China regional level

Probably because in mainland China only some regions participated in PISA, PISA results tend to be used to evaluate quality in education and identify issues by regional policymakers.

Shanghai's participation in PISA provides Shanghai an external scientific measurement of evaluating the quality in its compulsory education (Education Supervision Office of Shanghai Municipal Government, 2013). PISA results have been used for informing evidence-based policymaking in Shanghai (Yin, 2014). Although Shanghai achieved highly in PISA 2009 and PISA 2012, issues in Shanghai students' learning have been revealed as well (SHPISA, 2010; 2013). From PISA results, Shanghai has identified heavy learning burden of students and the relative weaknesses in mathematics learning, for example, students were not familiar with describing and interpreting results in mathematical language (Zhang and Huang, 2016). Previously, domestic assessment results had shown that Shanghai students compared with their peers in other provinces had heavier learning burden, higher learning pressure and lower learning self-confidence, despite that they had relatively high academic performance (Ni, 2012). These were further evidenced by PISA. Shanghai critically reflected on the cost for achieving the high performance, and pays more attention to related issues (Yin, 2014). In 2010, Shanghai was specified by the MOE as the pilot region for starting the reform on the assessment of students' academic quality in primary and secondary education (MOE, 2013b). In this context, to address the concerned issues to secure "the healthy and happy growth of students", MOE and Shanghai Municipal Government jointly promulgated "Shanghai Green Indicators for Academic Performance of Primary and Secondary School Students" (Green Indicators[2]) in 2011 (Yin et al., 2012). Construction of the assessment system of Green Indicators is strongly inspired by the concepts, framework as well as the techniques of PISA (Ji and Wang, 2012).

Amongst all the regions or local areas in mainland China that participated in PISA, Fangshan District of Beijing is the only area which had been continuously involved in three cycles of PISA from 2009 to 2015 with its locally-representative samples. Historically, as a traditional agriculture area of Beijing, Fangshan had relatively low living standards, and the quality of its education was not rated as high as that of peer districts in Beijing (Tao and Wang, 2014). After the turn of this century, Fangshan has had great development in economy and urbanisation (Cao, 2014), which requires the improvement of its local education (Tao and Wang, 2014). In Fangshan, as claimed by local educational policymakers, PISA has been used as an external measure to evaluate students' proficiency from an international perspective and examine its differences from that of students in other regions (Guo et al., 2015). It is also used "to promote the professional development of teachers, improve curriculum, and guide education reform at

---

[2] Green Indicators include ten aspects: Student Academic Level Index, Learning Motivation Index, Academic Burden Index, Teacher-Student Relationship Index, Teaching Methods Index, Principal Leadership Index, ESCS (economic, social and cultural status) Influence on Academic Performance Index, Moral Behavior Index, Well-Being Index and Progress Index (Ni, 2012).

local level”, as Guo, the former director of Fangshan Education Commission, and his colleagues state (2015, p.9). Motivated by PISA, three experimental projects, each of which focuses on one domain assessed by PISA, were successively launched by Fangshan Education Commission to improve teaching and learning (Guo, et al., 2015). The reading project, which was first launched, aims to advocate students’ reading with policy support and provision of reading materials (Gao and Zhang, 2015). The following mathematics project and the most recently-established science project aim to advance the reforms of teaching practices for promoting students’ literacy in mathematics and science (Liu, 2015; Wang, 2015; Zhang and Liang, 2015). I will return to these (i.e. the three PISA-motivated projects) in Chapter 5.

PISA’s impact has also moved beyond the regions which have formally participated in PISA. As documented in literature, it seems that in some other regions, PISA concepts are intentionally integrated into *Zhongkao* as well. For example, in one city of Zhejiang province, PISA-style items have been developed and used in mathematics examination of *Zhongkao* since 2010 (Yang, 2016), although Zhejiang had not formally participated in PISA until PISA 2018.

3.3.3 PISA facilitates China’s engagement in global communication of education policy

Due to Finland’s outstanding performance in PISA, admiration of Finnish education is no exception in mainland China. Batches of educational policymakers and practitioners went to Finland to investigate the secret of its educational success (e.g. Jiang, 2015). According to literature, some ideas in Finnish educational practices, for example, flexible curriculum, have even been incorporated into the curriculum reform in China (Zhang, 2018).

Although both Shanghai and Finland are top performers in PISA, the ways they achieved high academic performance are not the same (Shanghai Academy of Educational Sciences, 2012). In June 2012, to exchange the experience and challenges of education in China and in Finland, Shanghai Normal University, SHPISA, and Finland PISA National Centre jointly held the Sino-Finland PISA Conference “PISA results: experience and challenges in basic education” involving policymakers as well as researchers from both countries (Shanghai Academy of Educational Sciences, 2012). In this conference, the possible ways that China and Finland could learn from each other were also explored and discussed (Shanghai Academy of Educational Sciences, 2012).

Meanwhile, Shanghai's performance in PISA have drawn the attention of policymakers of other countries so that its education system has become a new reference society (Sellar and Lingard, 2013a). Since 2014, Shanghai has been involved in the "Mathematics Teacher Exchange: China-England" programme funded and managed by the Department for Education of the UK (Boylan et al., 2016). By sending primary mathematics teachers from England to Shanghai schools or hosting Shanghai teachers in the schools in England, this programme aims to learn from Shanghai mastery mathematics education and effectively apply that in England (Boylan et al., 2016). Researchers (e.g. Chen, 2016) consider that English teachers' views and comments on Shanghai teaching practices are significant for understanding and reflecting on Shanghai's own education as well.

3.4 PISA results about factors accounting for students' performance

Seen from the domestic policy responses across education systems as reviewed above in Section 3.2 and 3.3, many of the policy actions initiated in response to PISA results are simply based on students' average performance. Despite that PISA data include hundreds of indicators of policy interest, analyses on the data exploring factors accounting for students' performance is rarely seen in policymaking. As Pons (2012) suggests, PISA is mainly used as "knowledge for policy", that is, using PISA to support and legitimise one's pre-determined policy ideas; while its "knowledge for learning", which refers to learning from PISA data through deep analyses for informing policies, is often missing. Nevertheless, in seeking solutions from PISA, it is important for educational policymakers to consider students' performance as the function of various explanatory factors. Understanding the relationships between these factors and students' performance is also important in investigating and discussing PISA's impact, since it allows one to identify whether there are links between policies and PISA data, and evaluate practical impact of PISA-motivated policy initiatives on students' learning. In this section, I will review the explanatory factors that are widely discussed in the literature, and discuss how they are important in interpreting students' performance in PISA and in policymaking which draws evidence from PISA. This section has three subsections, organised according to the three categories of factors in the school effectiveness model employed in PISA framework (see Section 3.1). The factors covered here include input factors in terms of students’ demographic characteristics; processes factors such as opportunities to learn, teaching practices, and teaching quality; and non-cognitive outcomes factors such as motivation and self-beliefs.

3.4.1 Input factors

In this subsection, I will review PISA results about the relationships between students’ performance and input factors in terms of socioeconomic status (SES), students’ demographic characteristics such as gender, age, and grade.

3.4.1.1 Socioeconomic status

The relationships between students’ family SES and their academic performance have been discussed and documented in a range of studies (e.g. Hauser, 1994; Lee and Burkam, 2002; Walpole, 2003; Caro et al., 2009). According to the definition suggested by Duncan et al. (1972, cited in Sirin 2005), SES typically consists of parental education, parental occupation, and parental income, and it is usually measured as an aggregated index of these indicators.

Although student SES has been widely reported as a strong and positive predictor of academic performance (e.g. Bradley and Corwyn, 2002), some of reviews report that the relationships are not as strong as people assumed. From the meta-analysis on 101 studies covering published books, journals, and unpublished works between 1918 and 1977, White (1982) found that in students’ academic performance less than 5% of the variance could be explained by student SES, suggesting positive however weak association (r=0.2) between student SES and academic performance. He (1982) argues that the reason why significant relationships between student SES and achievement were widely reported is that nonsignificant results are usually not easy to get published. To identify whether this correlation changed over time, Sirin (2005) reviewed 58 journal articles which were published between 1990 and 2000 by replicating the meta-analysis methods used by White (1982). His findings suggest that on average the correlation had slightly decreased (Sirin, 2005). However, both White (1982) and Sirin (2005) find that the correlation between the average SES of students within schools (i.e. school SES) and academic performance was much stronger. They argue that the magnitude of the correlation between SES and academic performance is moderated by factors such as the definition and the measure of SES, and the unit (e.g. student, school) of analysis (White, 1982; Sirin, 2005).

It is also argued that it may not be the SES itself but the factors that underlie SES that influence the academic success. For example, from the perspective of neural development, Hackman et al. (2010) find that the effects of SES on children's cognitive ability on reading and mathematics could be explained by parent-child interactions as well as cognitive stimulation at home. The influence of these factors still hold even after children's intelligence quotient (IQ) is controlled for (Hackman et al., 2010).

SES has been typically used in ILSAs (e.g. PISA, TIMSS) assessment frameworks as an indicator for investigating equity-related issues, or as a control variable for discussing some other learning-related factors of interest (e.g. Mullis et al., 2009; OECD, 2004). Similar to the definition given by Duncan et al. (1972, cited in Sirin, 2005), the PISA index of economic, social and cultural status (ESCS) is aggregated with three components reported by students: highest parental education, highest parental occupation, and number of home possessions including the possession of household items which represents family wealth and the possession of books in the home (OECD, 2014a). In line with previous studies, PISA data also reveal limited effect of student-level ESCS on students' performance. For example, in PISA 2012, 5.2% of variation in student mathematics performance within schools could be explained by student ESCS (OECD, 2013b). Based on the data of 30 countries in PISA 2000, Marks (2006) also finds that controlling for students' ESCS background could only reduce limited between-school variance in achievement in reading, mathematics, and science for most countries. By analysing the data of PISA 2009 for three countries, namely, Australia, Canada, and New Zealand, Sullivan et al. (2018) find that ESCS had weaker effect on students' performance for students in rural areas compared with their peers in urban areas. As to school ESCS, aligned with the findings of the reviews of White (1982) and Sirin (2005), its stronger effect compared with that of student ESCS is consistently evidenced in PISA official results reports (e.g. OECD, 2004; 2013b) and studies using PISA data (e.g. McConney and Perry, 2010; Perry and McConney, 2010). For example, in PISA 2012, 62.8% of variation in student mathematics performance between schools could be explained by school ESCS (OECD, 2013b), suggesting socioeconomic segregation among schools (Perry and McConney, 2010).

3.4.1.2 Gender

Gender is another input variable involved in PISA framework for analysing performance of student subgroups and for identifying equity issues in education systems (OECD, 2013a). Historically, there is a stereotype that males, in general, outperform females in mathematics (OECD, 2004). Baker and Jones (1993) consider that the gender difference is caused by societal gender stratification in which female students do not face the opportunity equal to male students, and therefore unlike males, female students do not perceive that achievement in mathematics was useful for their future. In PISA 2003, the cycle in which mathematics was assessed as the major domain for the first time, males performed better in most countries (OECD, 2004). Nine years later, in PISA 2012 which focused on mathematics again, it seems that the gender difference between males' average performance and females' average performance had changed. Although males' mathematics performance was still significantly higher than that of females on average of OECD countries, only in 38 out of 65 education systems males outperformed females in mathematics, while in five education systems females even performed better (OECD, 2014b). Among those systems in which gender difference favouring males in mathematics were observed in PISA 2003, the gender gap was narrowed in some of them (OECD, 2014b). When it came to PISA 2015, only 28 out of 72 education systems showed males' advantage in mathematics (OECD, 2016a). In the Chinese context, gender difference in the mean mathematics performance was not significant for Shanghai in PISA 2012 and for China (B-S-J-G) (Beijing, Shanghai, Jiangsu, and Guangdong) in PISA 2015 (OECD, 2014b; 2016a; Guan, 2017).

The shrinking gender difference in mathematics has also been observed in other ILSAs such as TIMSS (Martin and Mullis, 2013). Therefore, instead of gender difference, "gender similarity" has begun to be argued (Hyde, 2005). By meta-analysing cross-national data of PISA and TIMSS in 2003, Else-Quest et al. (2010) find that although there was variability across nations, the mean effect sizes of gender difference in mathematics were small enough (d=0.15) to be "negligible". Marks (2008) argues that the changes in the types of mathematics problems, for example, problems which have relatively high reading demands, contribute to narrowing gender difference in mathematics performance.

It is also important to note that the shrinking gender differences are reported mainly based on comparing means of performance. Despite the officially reported non-significant gender difference, Zhu et al. (2018) argue that the gender difference in favour of males still existed for China (B-S-J-G) based on the results of multilevel analysis. Besides, it is found that, among mathematics high-achievers, males tend to have more outstanding achievement than females in China (B-S-J-G) and beyond (e.g. OECD, 2014b; Guan, 2017; Zhou et al., 2017; Zhu et al., 2018). With analysis of data in ten education systems in PISA 2003 and PISA 2012, Zhou et al. (2017) find that the gap among mathematics high-achievers were related with students' socioeconomic status, their after-school tutoring, and their attitudes towards mathematics and school. By using a multidimensional Rasch Model to identify the gender difference in USA students' mathematics performance in PISA 2000 and 2003, Liu and Wilson (2009) suggest that the gender differences were different across mathematics subareas and item types.

In contrast, the gender difference in reading performance favouring female students is commonly reported across educational systems (e.g. OECD, 2010). It is noted that the gender difference in mathematics was inversely associated with the gender difference in reading (Stoet and Geary, 2013). Based on four cycles of PISA data (2000, 2003, 2006, and 2009) across 75 education systems, Stoet and Geary (2013) suggest that this inverse relation was not only a between-system effect, but was also found within systems.

3.4.1.3 Age

As I introduced before (see Section 3.1), the specific student population targeted by PISA is those who are aged from 15 years and three months to 16 years and 2 months and enrolled in 7th grade and higher. That means, for the students involved in PISA, the age of the youngest and that of the oldest can be about one year apart. It is argued that, students born earlier than their peers in the same school year tend to achieve higher in academic performance (Massey et al., 1996; Cobley et al., 2009; Aune et al., 2018). Aune et al. (2018) consistently find this association between month of birth and achievement in Norway national numeracy test for 5th, 8th, and 9th grade students. Researchers suggest that greater cognitive maturity of relatively old students can make them advantaged in comparison to their younger peers in intellectual development (Aune et al., 2018; Vestheim et al., 2019).

Despite the recognition of the age effects, there is also some voice suggesting that for 15-year-olds, cognitive maturity usually has been stable at this age, and significant differences in cognitive maturity should not be shown between students in this age cohort (Heckman, 2000; Sprietsma, 2010). As Pehkonen et al. (2015) find, the relative age effect on Grade Point Average (GPA) is significant for 6th grade students but not for 9th grade students in Finland, indicating that the effect vanishes when students get older.

Since PISA student sampling is age-based, subject to national education policy and education system structure, PISA eligible students might be enrolled in different grade levels (see Section 3.4.1.4). Hence, it is necessary to also take into consideration the potential influence of the years of schooling that students have received when examining the age effect. After controlling for the grade level, based on the data of 31 countries in PISA 2000, Fuchs and Wößmann (2007) find significant negative relationships between students' age in month and their achievements in both mathematics and reading, while the age effects are quite small. They explained that the significant relationships might be brought by grade repetition (Fuchs and Wößmann, 2007).

3.4.1.4 Grade

Literature on PISA and some other ILSAs has documented that students in higher grades are more likely to have higher performance (e.g. Beaton et al., 1996; Fuchs and Wößmann, 2007; Melkonian and Areepattamannil, 2018). Based on PISA 2000 data of 31 countries, the significant relationships between grade and students' performance are found in all three assessment domains (Fuchs and Wößmann, 2007). On average, students in 10th grade perform 97.6, 73.2, and 106.2 score points higher than those in 7th grade respectively in mathematics, reading, and science (Fuchs and Wößmann, 2007). Marks (2006) considers that, for the same age cohort, like PISA target student population (i.e. 15-year-olds), those of higher ability might start school at a relatively young age or skip grades during schooling, and those enrolled in higher grades are exposed to more difficult and complex curriculum content. This indicates that higher-grade students might have higher chance of exposing to the content assessed by PISA, although PISA examines students' cumulated learning outcomes and does not intend to match the curriculum of grades (OECD, 2013c)

In PISA, since students are not sampled by grade, it is noted that the student composition in grades changes within countries over years, and the changes in this respect are considered meaningful for interpreting students' performance in PISA and performance trends across PISA cycles (OECD, 2016a; Aloisi and Tymms, 2017). It can be seen that the OECD has also recognised the importance of the changes in student demographics over years for interpreting students' performance trends, and documented in PISA reports the adjusted trends which account for gender, age (measured in quarters), and immigrant background (OECD, 2016a). However, it appears that the grade effect has not yet been officially included in the analysis of PISA adjusted trends. Aloisi and Tymms (2017) suggest that if student demographic composition including that in grade levels has substantive effects on students' overall performance in PISA, the adjusted scores should be underlined in results reporting.

3.4.2 Processes factors

In this subsection, I will review PISA results about the relationships between students' performance and process factors in terms of opportunities, and teaching practices and quality.

3.4.2.1 Opportunities to learn

Students' opportunities to learn (OTL) is suggested as the core of schooling (Schmidt and Maier, 2009). It is commonly considered that the concept of OTL was originally proposed by John Carroll in the early 1960s (Schmidt et al., 2013; Lafontaine et al., 2015). In his theoretical model of school learning, Carroll (1963) argues that individuals could successfully learn a given task as long as they spend enough time, and introduces OTL as a measure of allocated learning time. The concept and measures of OTL are developed along with early ILSAs (Lafontaine et al., 2015; Suter, 2017). Besides the early defined concept with regard to the time exposure which is suggested by Carroll, it has been also interpreted as exposure to or coverage of specific content. For the First International Mathematics Study (FIMS) which was launched by the IEA in 1964, Husén (1967) defined OTL as “an opportunity to study a particular topic or learn how to solve a particular type of problem presented by the test”. Schmidt et al. (2013) suggest OTL as content coverage in classes and the time spent on the specific content.

Regardless the inconsistent interpretations of OTL amongst researchers, OTL is used in ILSAs for explaining the variation in students' performance in assessments within and across countries, and informing educational policymaking in relation to national curriculum and its implementation (Yang Hansen and Strietholt, 2018). In PISA 2012, in which mathematics was assessed as the focused domain, OTL measurement was particularly included in the student questionnaire. In PISA 2012, the concept of OTL seems aligned with the definition given by Husén (1967), since it was defined as “coverage of content categories and problem types” (OECD, 2013a, p.186). It is measured by indices of students’ perceived experiences with specific mathematical tasks (e.g. exposure to pure mathematics problems, exposure to applied mathematics problems), and students’ perceived familiarity with mathematical concepts (OECD, 2013a).

Previous research shows evidence that OTL is related to students’ performance (Schmidt et al., 2001). In terms of the OTL measures in PISA, by fitting data with multilevel linear models, the official results of PISA 2012 show that students’ exposure to pure mathematics at school had significant and positive relationships with their mathematics performance across and within education systems (OECD, 2014b). Amongst the OECD countries, on average, increasing one unit of exposure to pure mathematics could bring about 50 score points (i.e. 0.5 SD on PISA international scale) higher in students’ mathematics performance (OECD, 2014b). In contrast, the relationship between exposure to applied mathematics and students’ mathematics performance is curvilinear across education systems that mathematics performance increases along with more exposure to applied mathematics and then after a certain point it decreases (OECD, 2014b). As these differential relationships suggest, it seems that mathematics principles themselves are more crucial than embedding real-life contexts in mathematics problems (Schmidt and Burroughs, 2015). Although it is not revealed that the relationship between exposure to applied mathematics and mathematics performance is as straightforward as that for exposure to pure mathematics, the official PISA report highlights that good performance in PISA mathematics is also associated with the opportunities to learn applying mathematics in real-life contexts (OECD, 2014b). The OECD suggests that both pure mathematics and application of mathematics in real-world contexts should be taught (OECD, 2014b). Complementing this view, Schmidt and Burroughs (2015) argue that, with regard to applied mathematics, “more is not necessarily better” (p.31), and caution over concentration on embedding real-life contexts in mathematics instruction.

Despite the strong relationships between OTL and students' performance officially claimed by PISA results report, some researchers argue that their relationships are rather modest (Carnoy et al., 2016; Yang Hansen and Strietholt, 2018). By reviewing OTL measures in ILSAs from 1959 to 2011, Suter (2017) concluded that the ways that OTL is measured considerably influence the magnitude of the relationships between OTL and students' performance. In PISA, OTL measures are based on students' self-report. For one of these measures, familiarity with mathematics concepts, one response option is "Know it well, understand the concept" (OECD, 2014b). Yang Hansen and Strietholt (2018) consider that this measure contains unintended construct besides of OTL, since it involves students' self-evaluation of their mathematics proficiency. They adjusted for students' mathematics self-concept for examining the OTL effect, and found that its effect on students' mathematics performance turns to be smaller.

Regarding the importance of OTL for students' learning and their achievement, a few more points maybe worth considering as well for furthering the understanding of their relationships and improving OTL for students. For example, Carnoy et al. (2016) find that the OTL effect of exposure to pure mathematics is stronger for the student subgroup who have low family academic resources (e.g. books in home, mother's educational level) and high mathematics proficiency. Hence, they caution that the various effects of OTL across student subgroups should be taken into consideration for policymaking toward improving OTL (Carnoy et al., 2016). Moreover, it is found that the variation in OTL is mainly within schools rather than across schools (Schmidt and Burroughs, 2015). This suggests, compared with educational inequality in terms of OTL across schools, variation of OTL within schools is more effective for explaining students' performance. Since PISA student population is 15-year-olds who may not definitely be enrolled at the same grade, to some extent, the within-school variation may be brought by different years of schooling, as suggested by the above reviewed grade effect (see Section 3.4.1.4).

3.4.2.2 Teaching practices and quality

In terms of teaching practices, it seems that a student-oriented (also called student-centred) approach has been widely promoted in education reforms around the world (Westbrook et al., 2013; Cheung et al., 2018). However, PISA results provide a mixed picture about the association between the frequency of employing this approach in teaching and students' performance.

Participating countries and regions ranked high on the league table do not necessarily also report high frequency of using student-oriented approach (OECD, 2013d; Schweisfurth, 2013). By employing data of 62 education systems, Caro et al. (2016) find inconsistent relationships between this factor and mathematics performance across systems. As to the picture within systems, for example, based on PISA 2012 data of Turkey, Demir (2018) finds negative effect of student-oriented teaching practices on students' mathematics performance. It is suggested that student-oriented teaching approach could explain student's motivation and self-concept in mathematics rather than their mathematics performance (Caro et al., 2016; Scherer et al., 2016). Caro et al. (2016) suggest that this teaching approach may have differential effectiveness for students in different contexts of classrooms, schools, and education systems, and for students of different characteristics. Regarding the influence of students' characteristics on the effectiveness of this kind of teaching practices, students' own learning strategies can matter. For example, through interviewing secondary school students, Campbell et al. (2001) find that compared with students with surface learning strategy, those with deep learning strategy tend to have better recognition and understanding of the potential learning opportunities offered in student-oriented teaching practices. Hence, it is suggested that, along with the employment of student-oriented teaching practices, the skills about learning from the student-oriented classroom activities are necessary to be taught to students as well (Campbell et al., 2001). Researchers argue that the artificial division between student-oriented and teacher-centred approaches, and advancing one over the other, oversimplify the dynamic and complex classroom contexts (Cuban, 1983; Noyes, 2012). Cuban (1983) suggests that it is hard to assess the effect of teaching practices in this regard on students' performance, since they are usually dynamic rather than stable over time, and often embody diversity of strategies.

As to teaching quality, Klieme et al. (2001) suggest cognitive activation as one of its dimensions, and find cognitive activation positively associated with students' mathematics performance. Cognitive activation is featured with cognitively challenging and highly structured tasks (Baumert et al., 2010; Keller et al., 2017), in which students are encouraged to apply their acquired knowledge to different contexts, seek different solutions, communicate and reflect on their thoughts and methods (Klieme et al., 2001; OECD, 2014a).

It is suggested that the extent to which students can benefit from cognitive activation instruction is subject to the pedagogical content knowledge of their teachers (Baumert et al., 2010; Kunter et al., 2013). Pedagogical content knowledge allows teachers to be able to make instruction content accessible to students (Shulman, 1986; 1987; Kleickmann et al., 2013), anticipate and respond to the problems that students would encounter (Keller et al., 2017). However, the provision of the potential learning opportunities in cognitive activation does not mean the actual use of these opportunities by students. It is argued to distinguish these two conceptions and proposed that students' use of the opportunities mediates the effect of cognitive activation on their performance (Praetorius et al., 2018). By comparing video observation and students' reported perception, researchers find that students may not necessarily perceive or recognise that they are cognitively activated (Hugener et al., 2009). Mayer (2004) writes that teachers need to provide focused goals and enough guidance to students to allow them build knowledge through appropriate cognitive activation. Additionally, supporting and scaffolding students is considered to be essential for motivating students to be engaged in the process of cognitive activation (Baumert et al., 2010).

Demir (2018) finds that Turkish students' perceived frequency of their teachers' use of cognitive activation in mathematics classes was positively associated with their mathematics performance in PISA 2012. Similarly, the positive relationships are also found in Korea and Singapore (Yi and Lee, 2017). However, Caro et al. (2016) find that the relationships between the extent to which students are cognitively activated and students' mathematics performance are positive yet curvilinear across 62 education systems in PISA 2012. They propose that there is no longer association between the level of cognitive activation and students' mathematics performance when cognitive activation is of quite high level (Caro et al., 2016).

Researchers argue that it is problematic to use a single indicator for the complex construct of cognitive activation (as PISA does) to predict students' performance, since the possible level of cognitive activation can be different in different types of lessons (e.g. introduction lessons vs practice lessons), depending on the content of lessons and how it is implemented (Baumert et al., 2010; Praetorius et al., 2014; Praetorius et al., 2018).

3.4.3 Non-cognitive outcomes factors

Non-cognitive outcomes, included in PISA framework as the outcomes of the input and processes of learning, are also considered to have influence on the cognitive outcomes of learning (e.g. mathematics performance). In the following, I will review the PISA results about non-cognitive outcomes in terms of motivation and self-beliefs, which have been widely discussed in the literature.

3.4.3.1 Motivation to learn mathematics

Motivation, as one category of non-cognitive factors, is basically distinguished as two types, that is, intrinsic motivation (i.e. interest and enjoyment) and extrinsic motivation (i.e. instrumental motivation), according to the reasons and goals that motivate an action (Deci and Ryan, 1985; Ryan and Deci, 2000). Usually, a student's motivation to learn is a composition of both (Husman and Lens, 1999). Both of these two types of motivation are employed in PISA framework to investigate to what extent gaining joy from learning activities (i.e. intrinsic) or believing learning activities are good or important for studies or future career (i.e. extrinsic) motivates students (OECD, 2013d). Previous empirical research has shown that intrinsically motivated students tend to learn better, especially on the heuristic tasks or those requiring deep understanding (Ryan and Deci, 2009). Interest and enjoyment could affect the degree of students' engagement in learning, the depth of understanding (Schiefele, 2009) and even the types of careers students would like to pursue (Reeve, 2012). By contrast, extrinsic motivation is considered to be influential in students' choices to involve in mathematics courses (Stevens et al., 2004), and is linked to students' occupational choices (Eccles, 1994). Discussion on the association between students' learning motivation and their academic performance have been widely documented (e.g. Wigfield and Eccles, 2000; Lepper at al., 2005). The explanatory power of intrinsic and extrinsic motivation varies amongst education systems, as revealed by PISA (Shin et al, 2009).

In PISA 2003 and PISA 2012 in which mathematics were assessed as the focus domain, the intrinsic motivation to learn mathematics had significant correlation with mathematics performance for the average of OECD countries (r= 0.19, r=0.22 respectively) (OECD, 2013d). The strength of the relationship between intrinsic motivation and mathematics performance varied across education systems that participated (OECD, 2004; 2013d). On OECD average, intrinsic motivation explained 5% of variance in mathematics performance in PISA 2012, and one-unit increase in this factor was associated with 19 score points improvement in mathematics performance (OECD, 2013d). However, this positive relationship was not observed in some systems (OECD, 2013d). It is also noteworthy that some East Asian education systems, such as Shanghai-China, Hong Kong-China, Chinese Taipei, and Singapore, were the exceptions to those showing the relationships for highest-achieving students stronger than for lowest-achieving students (OECD, 2013d).

Compared with intrinsic motivation, instrumental motivation has a more ambiguous relationship with academic performance (Gabriel et al., 2018), since some studies identified it as a positive predictor (e.g. Barron and Harackiewicz, 2001) while others (e.g. Lepper et al., 2005) showed its negative effects. Researchers argue that instrumental motivation may be a positive predictor for college students' performance, but it may be less predictive for students in elementary and secondary schools (Lepper et al., 2005). In PISA, generally, instrumental motivation has a weaker relationship with performance than intrinsic motivation does (Lee and Stankov, 2013). Instrumental motivation of OECD average has a relationship with mathematical performance with r=0.14 and 0.20 respectively in PISA 2003 and PISA 2012, which is weaker than intrinsic motivation (r=0.19, 0.22) (OECD, 2013d). The relationship varied across education systems as well. On OECD average, difference in students' instrumental motivation to learn mathematics explained 4% of variance in their mathematics performance (OECD, 2013d).

3.4.3.2 Self-beliefs in learning mathematics

Self-beliefs are also considered as important non-cognitive outcomes factors influencing students' performance. The extent to which students are confident in their own ability of solving specific tasks (i.e. self-efficacy) and students' overall perception of their own ability (i.e. self-concept) "have an impact on learning and performance on several levels: cognitive, motivational, affective and decision-making" (OECD, 2013d, p.88).

Although self-efficacy is a judgement of what one can do with the skills one has, and may not necessarily reflect the skills one actually possesses, it is considered that self-efficacy is an important factor influencing performance (Bandura, 1992; 1997). Researchers argue that students with high self-efficacy are more likely to endeavour to solve problems or even challenge difficult problems (Bandura, 1977; Marsh et al., 2006; Schunk and Pajares, 2009; Hoffman, 2010). Self-efficacy is also the outcome of this endeavour.

Performance accomplishments, depending on whether they are successes or failures, could in turn increase or reduce self-efficacy (Bandura, 1977).

Consistent with a range of literature that shows that students' self-efficacy is one of the strongest non-cognitive factors predicting students' academic performance (e.g. Lee and Stankov, 2013; Şahin and Yıldırım, 2016; Lee and Stankov, 2018), self-efficacy is identified as a strong predictor in PISA as well (OECD, 2004; 2013d). In most education systems, it had a moderate or strong positive correlation with mathematics performance (OECD, 2013d). In PISA 2012, for example, self-efficacy could explain 28% of variance of mathematics performance of students across OECD countries (OECD, 2013d). One-unit increase in self-efficacy was associated with 49 score points higher in mathematics performance (OECD, 2013d). For most education systems, the association between self-efficacy and mathematics performance was greater amongst high-achieving students. However, this is not the case for Belgium and some Asian education systems such as Shanghai-China, Hong Kong-China, Korea, Macao-China, Chinese Taipei, and Singapore, in which one-unit increase in self-efficacy was associated with larger difference in mathematics performance of low-achieving students than that of high-achieving students (OECD, 2013d).

In terms of self-concept, like extrinsic motivation as relative to intrinsic motivation, it has a weaker effect than self-efficacy on students' performance in PISA (OECD, 2004; 2013d). It is suggested that compared with self-efficacy, self-concept could predict better for affective-motivational variables such as interest and anxiety rather than academic performance (Ferla et al., 2009). As reported in PISA 2012, across OECD countries, one-unit increase in mathematics self-concept was positively associated with a difference of 37 score points in mathematics performance (OECD, 2013d). Although in general self-concept is positively correlated with performance, it does not necessarily mean that education systems with high average mathematics self-concept also achieve high in mathematics (OECD, 2013d). It is observed that East Asian areas such as Shanghai-China, Hong Kong-China, Macao-China, Chinese Taipei, Japan, and Korea, which were ranked high in PISA performance (including mathematics) had relatively low mathematics self-concept compared with Western countries (OECD, 2004; 2013d). This phenomenon is consistently found in many other international assessments (Ho, 2009).

As indicated in the measures of self-concept, for example, “I have always believed that mathematics is one of my best subjects; in my mathematics class, I understand even the most difficult work.”) (OECD, 2013a), self-concept has normative nature because it greatly reflects students’ perception of themselves compared to their peers around (Schunk and Pajares, 2005). Students in high-performing schools tend to have lower self-concept than those with equal ability in low-performing schools, which is called “Big-Fish-Little-Pond effect (BFLPE)” (Marsh and Parker, 1984; Marsh et al., 2008). BFLPE is used to explain the paradox in East Asian areas that students have relatively low self-concept but have high performance (Ho, 2009). Besides in East Asia, BFLPE in relation to self-concept has been supported in studies comparing regional performance within countries (e.g. Seaton et al, 2011). Its generalisability across cultures has also been evidenced (Marsh and Hau, 2003). Response style is another possible reason used to explain East Asian students’ relatively low self-concept, considering that students in East Asia tend to choose the middle point in the Likert-scales which are typically used in self-concept measures (Cheung et al, 2018). After adjusting for response styles, high levels of mathematics self-concept are identified in students of Singapore and Shanghai-China (Cheung et al, 2018).

Mathematics anxiety is another category of self-beliefs in learning mathematics that is investigated in PISA. It refers to “thoughts and feelings about the self in relation to mathematics, such as feelings of helplessness and stress when dealing with mathematics” (OECD, 2013d, p.88). It is suggested that students’ mathematics anxiety is caused by the composite effect of high mathematics anxiety of people around students (e.g. parents, teachers) and students’ own cognitive predisposition (Beilock and Maloney, 2015). Alpert and Haber (1960) distinguish that anxiety could be facilitative or debilitative. Facilitative anxiety could urge students to make efforts to achieve their learning goals, while debilitative anxiety may make students avoid challenges and therefore hinders their learning (Ho, 2009). Previous studies have widely documented the negative effect of anxiety on performance (Hembree, 1990).

In terms of its negative effect, anxiety would make students avoid engaging in mathematics (Ashcraft and Krause, 2007; Luttenberger et al., 2018). Besides this, anxiety also hinders mathematics performance through impairing working memory (Hopko et al., 1998). When doing a mathematics task, for students with high mathematics anxiety, much of their attention is on negative emotion caused by anxiety, while working memory resources for processing mathematics are greatly impaired (Ashcraft and Krause, 2007).

Consistent with previous studies, PISA results reports also have demonstrated that students' mathematics anxiety is negatively associated with their mathematics performance (OECD, 2004; 2013d). Following mathematics self-efficacy, mathematics anxiety is considered as another strong predictor of mathematics performance (Lee and Stankov, 2018). In PISA 2012, for example, differences in anxiety of OECD students could explain 14% variation in their mathematics performance (OECD, 2013d). Across OECD countries, one-unit increase in anxiety was related to 35 score points decrease in mathematics performance in PISA 2003 and 34 score points decrease in PISA 2012 (OECD, 2004; 2013d).

3.5 PISA limitations

Although PISA provides rich data that allow for analysing features of education systems worldwide (Fuchs and Wößmann, 2007) and identifying relationships between input factors, processes factors, and learning outcomes (Schleicher and Zoido, 2016), reflective use of PISA data is suggested (Dolin and Krogh, 2010). It is worth noting that, PISA, like many other ILSAs, also has limitations which are important to keep in mind when using and interpreting its data. In the following, I will review its limitations in measurement (Section 3.5.1), sampling (Section 3.5.2), translation (Section 3.5.3), content coverage (Section 3.5.4), and causality inference (Section 3.5.5) which have been mainly discussed by researchers.

3.5.1 Measurement model

In pre-2015 cycles, the Rasch model (Rasch, 1960), in which the probability of correctly responding to an item is modelled on the item difficulty parameter and the respondent's latent ability parameter, was employed in the estimation of item difficulty parameters. One of the assumptions of the Rasch model is parameter invariance. That is, equivalence of item difficulty parameters across respondent populations is assumed. In early cycles of PISA, under this assumption, international item parameters were estimated based on the random sample selected only from each of OECD countries (so called "international calibration"), to which item parameter estimates for all PISA-participating education systems were then scaled (OECD, 2002;

2005; 2009). However, with empirical investigation, this assumption has been found hard to hold across student populations of various systems (Kreiner and Christensen, 2014; Rutkowski et al., 2016). Researchers argue that violation of this assumption may weaken the accuracy of rankings of performance, since estimation of the differences in students' performance between education systems might be biased in this approach (L, Rutkowski and D, Rutkowski, 2016). It is noted that the technical team of PISA has been making efforts to reduce bias brought by the heterogeneity of item difficulty parameters across populations. For example, from PISA 2009 which gave countries an option to use easier booklets compared with standard ones, a subsample selected from the countries using easier booklets were added into the international calibration sample (OECD, 2012). In PISA 2015, data from all participating systems in 2015 and previous cycles (2006-2012) were further added into international calibration, with the consideration of population differences across systems and over cycles (OECD, 2017).

Another measurement limitation of PISA that has been debated is with respect to the unidimensionality assumption of the Rasch model (Rasch, 1960). It assumes that there is only one single ability dimension for each scale (e.g. mathematics literacy). The items which fail to meet this assumption at national level are treated as "dodgy items" which would be excluded from international scaling (e.g. OECD, 2014a). Although this assumption allows PISA to render league tables of students' performance on each scale of assessed domains (Bonnet, 2002), it is argued that, like the parameter assumption, the unidimensionality assumption may also not hold in the case of ILSAs such as PISA (Bonnet, 2002; Goldstein, 2004; Goldstein, 2018), considering the linguistic and cultural influence on the test (Bonnet, 2002). Eliminating the "dodgy items" would make interesting information which might also be meaningful for understanding country differences removed as well (Goldstein, 2004). Hence, adopting multidimensional statistical models which retain this information is suggested (Goldstein, 2004).

3.5.2 Sampling

PISA is criticised for its limitations in target population comparability. As we know, PISA targets students of a given age (i.e. 15-year-old). However, the percentage of 15-year-olds enrolled in schools varies across education systems depending on national policies (e.g. policies on grade repeating, school-starting age) or students' maturation (Prais, 2003; Bracey, 2004; Duru-Bellat and Suchaut, 2005). This means, for some systems, some pupils have finished their compulsory education and left schools by age 15, and therefore would not be captured by PISA sampling. In addition, as I reviewed previously in Section 3.4.1, due to various school-starting ages across systems, students at the same age, 15-year-old, for example, would have received different years of schooling (Leung, 2014). Because of this, some researchers (e.g. Prais, 2003) prefer grade-based sampling which is used in other ILSAs such as TIMSS. Yet, this approach is contestable as well, since it is not problem-free either considering various school-starting ages (Goldstein and Thomas, 2008; Leung, 2014).

The uncertainty in students' demographic distribution can hinder the accurate estimation of students' performance trends across cycles (Gebhardt and Adams, 2007; Aloisi and Tymms, 2017), as I reviewed in Section 3.4.1.4. For example, by reweighting Portuguese samples in PISA 2009 and PISA 2012 in terms of students' distribution in grades, tracks, and school types to reflect the changes of population composition over this period, Freitas et al (2016) find that the trends in each of the three primary assessment domains (i.e. reading, mathematics, and science) were notably different from those reported in PISA official reports.

3.5.3 Translation

The international comparability of PISA assessment instruments has been questioned in terms of the linguistic and cultural equivalence across various language versions. Since the first cycle of PISA was administered in 2000, the number of participating education systems has been increasing (see Figure 3.1 in Section 3.1). With the launch of its extension programme, PISA for Development, even more and more developing countries have been getting involved in this programme. In developing assessment instruments, PISA uses English and French as two language sources, from which national versions are translated and adapted (e.g. OECD, 2014a). The expanded participation of education systems of various languages and cultures has brought PISA the challenge of developing assessment instruments of various national versions of different languages. It is admitted that the OECD has made great effort with systematic and rigorous procedures during translation and back-translation to assure the linguistic and cultural equivalence of these versions across systems (Grisay, 2003; McQueen and Mendelovits, 2003; Grisay et al., 2007). However, some researchers argue that it is hard to achieve full comparability in international assessments of various language versions (Brown et al., 2007).

There are still concerns of possible bias caused by linguistic and cultural differences across countries (Goldstein, 2004; Grisay et al., 2007; Goldstein, 2018). By comparing the English version of PISA cognitive tests with other language versions such as Finnish, Irish, and German, Eivers (2010) finds that translation into different languages caused changes to various item text lengths which would bring the issue of test speededness, considering that participating students in all systems have the same time limits on the test. Moving the comparison to worldwide language and cultural context, by examining the item difficulties across participating jurisdictions in the first three cycles of PISA, Grisay et al. (2007) find that PISA test instruments tended to favour students in Western countries in which Indo-European languages are used.

3.5.4 Content coverage

As already discussed, PISA indicators inform educational policies (Section 3.1). Due to the policy-driven aim, PISA focuses on the content that is "of high value or interest to educational policy-makers or practitioners" across participating education systems (Adam, 2003, p.379). Those constructs that can be measured and assessed internationally in terms of feasibility with existing technology are selected (Gorur, 2016) and operationalised in assessment (OECD, 1999; Adam, 2003).

For PISA cognitive tests, the primary domains assessed in each cycle of PISA are decided to be limited to reading, mathematics, and science literacies (OECD, 1999). As a senior PISA official stated in an in-depth interview, "reading, science and maths are there largely because we can do it. We can build a common set of things that are valued across the countries and we have the technology for assessing them" (Interview transcript, senior PISA official, cited in Gorur, 2011, p.83). To keep up with social development, since PISA 2012, more aspects of cognitive content, such as financial literacy (OECD, 2013a), collaborative problem-solving (OECD, 2016d), and global competence (OECD, 2019b) have been successively included in PISA assessment frameworks as optional domains.

Data collected through background questionnaires suffer debates on content limitations as well. PISA questionnaires collect a range of data with regard to educational effectiveness including students' background, educational policies and classroom practices through self-reports of students and school principals (OECD, 2013a). Researchers suggest that not all the key factors leading to high educational achievement are covered in PISA. For example, Wu (2014) considers that, like many educational surveys, PISA does not comprehensively capture the education environmental factors (e.g. coaching schools, parental pressures) about students' lives outside of schools, which may also be influential in leading to the success of education outcomes. Feniger and Lefstein (2014) argue that from PISA data, one cannot learn how policies have actually been enacted and the impact of actual classroom teaching and learning practices. Goldstein (2004) further argues that, like many other ILSAs, as a cross-sectional survey, PISA has limitations in comparing the effectiveness of education systems since it does not have longitudinal data to allow social and other differences, in addition to differences across education systems, to be accounted for.

3.5.5 Causality inference

By building understandings of the characteristics of an education system, PISA could be an empirical knowledge base serving policy formulation (Gustafsson, 2008; Yore et al., 2010). However, to assume that a study could provide a comprehensive understanding of a complex social system such as education is seen as unrealistic (Gist, 1998). Mainly because of PISA's limitations in content coverage, cautions on drawing causal inferences from the association between students' performance and learning-related factors identified in PISA data are raised by a number of researchers (e.g. Duru-Bellat and Suchaut, 2005; Goldstein and Thomas, 2008; Gustafsson, 2008; Rutkowski and Delandshere, 2016).

It is considered that for each education system educational outcomes are generated by the overall structures of its societal coherence (Duru-Bellat and Suchaut, 2005). There might be mediating variables having effects on the relationships between the given factors and students' performance (Gillis et al., 2016), while some of them, as reviewed in last section, are absent in PISA assessment frameworks. As stated in PISA official results reports (e.g. OECD, 2010, p.18), "PISA cannot identify cause-and-effect relationships between inputs, processes and educational outcomes". Statistics alone could not sufficiently evidence causal relationships (Wu, 2014). Therefore, admittedly, PISA has limitations in answering what we would like to know about how to improve education (Buckingham, 2012; Jerrim, 2015; Baird et al., 2016; Gillis et al., 2016). To form good policies with PISA data, an amount of reasoning and other evidence beyond statistical analyses are still needed (Wu, 2014).

With regard to additional evidence necessary for disclosing causality relationships, Goldstein and Thomas (2008) suggest the need of controlling for potential confounding factors such as prior achievements and social background, which could be identified from in-depth qualitative investigations (Egelund, 2008). Since longitudinal data are considered more powerful in controlling for these factors through repeated observation on the same students and education systems over time (Gustafsson, 2008), incorporating longitudinal data, collected from qualitative methods or mixed methods, is proposed (Goldstein, 2004; Egelund, 2008; Gustafsson, 2008) to reveal causal relationships. It is suggested that PISA participating education systems could additionally collect these kind of data for their national purposes by adapting and extending the design of PISA international assessment instruments (Gustafsson, 2008).

In summary, the literature as reviewed above (Section 3.2-3.5) displays the overall picture about extant discussions and debates on PISA's policy impact in domestic education contexts, and points out the importance of understanding PISA data and limitations for making evidence-based policies, and also for discussing and evaluating PISA's impact in educational policies and practices. Based on the synthesis of the literature, I will develop the conceptual framework by further employing a hybrid of theories about washback (also called backwash[3]) effect (e.g. Alderson and Walker, 1993; Hughes, 1993, cited in Bailey, 1996) and ecological systems theory (Bronfenbrenner, 1979) as theoretical underpinnings in the next section.

3.6 Developing a conceptual framework

Regarding the impact of educational assessments, Bachman and Palmer (1996, p.39) consider it "can be defined broadly in terms of the various ways in which test use affects society, an education system, and the individuals within these". They suggest that it operates at a micro level (i.e. the influence on the individual test takers and test users), and a macro level (i.e. the influence on the educational systems or society). Wall (1997) also defines it

---

[3] As Hughes (2003) writes, the term "backwash" can be found in dictionaries, while "washback" cannot. However, "washback" is more prevalently used (Bachman and Palmer, 1996).

broadly. According to her point of view, impact is “any of the effects that a test may have on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole” (Wall, 1997, p.291). From the above literature review, we see that PISA’s impact on educational policies in a range of education systems has been widely documented. When PISA-motivated policy initiatives are put into practice, they would also have influence on schools, teachers and students (Sellar et al., 2017). During the process of shaping and enacting these policies, various actors such as educational policymakers and practitioners would be involved in interpreting and translating PISA concepts and data, and thereby influencing PISA’s impact on educational practices in schools. Therefore, it is reasonable to assume that, due to various policy actors’ agency, the practical impact of the so-called governing power of PISA on students’ learning may be dynamic. However, research on empirically investigating the impact that PISA has brought to students’ learning via the enactment of policy reforms is still rare to see, and how PISA impacts students’ learning at schools remains unclear. To appropriately use PISA for informing educational policies and practices, it urges the investigation of the impact PISA has brought to students’ learning and how its impact occurs. To do that, theories about washback effect in combination with the ecological systems theory provide the foundation, serving to understand the “what” (i.e. process and outcomes) as well as the “how” of the impact of PISA.

3.6.1 Washback effect

The concept of washback has been considerably discussed and its existence is asserted most widely in the context of language testing or high-stakes educational assessments (Alderson and Walker, 1993; Hughes, 2003; Cheng et al., 2004). For example, the washback effect of *Zhongkao* and *Gaokao* has been discussed extensively (e.g. Gao, 2010; Cong, 2014; Zhang, 2016). Although PISA is not explicitly high-stakes for schools and students, since it does not report individual students’ scores, it is considered high-stakes for policymakers in PISA participanting education systems (Sellar et al. 2017). The concept of “washback ” has also been introduced to discuss PISA effects as well (e.g. Baird et al., 2016; Maddox and Addey, 2016).

Regarding washback, at the time when it emerged to be a research interest, it specifically refers to test impact on teaching and learning (Buck, 1988; Hughes, 1989; Shohamy, 1992; Hamp-Lyons, 1997). Since discussions on washback most focus on the processes of teaching and learning in which not only individuals, but also educational and societal systems, are involved, some researchers consider washback as “effects on individuals, educational systems, and society at large” (Bachman and Palmer, 1996, p.30). Along with the development of studies on washback, now the latter’s point of view has been widely recognised (Hughes, 2003). Sometimes, washback is used as a synonym of impact as well (Wall, 1997). For instance, Scott (2014) classifies washback into two types: “micro washback” which directly affects individuals; “macro washback” which directly affects institutions and systems in which individuals are influenced subsequently. This definition is similar to that stated by Bachman and Palmer (1996) for “impact”. It is generally argued that washback could be either positive or negative, depending on the administration and use of the test within a particular educational context (Cheng et al., 2004).

To shape the structure of this empirical investigation on the impact of PISA on student learning, I employ the basic washback model suggested by Hughes (1993, cited in Bailey, 1996, p.262) as the basic frame. In his unpublished manuscript[4], Hughes (1993, cited in Bailey, 1996, p.262) distinguishes the washback on participants, process and product in teaching and learning, believing that all three could be influenced by the nature of a test. His trichotomous washback model (Hughes 1993, cited in Bailey, 1996, p.262) is as Table 3.1 shows.

Table 3.1 Hughes’ washback model

Trichotomy	Signified
participants	students, classroom teachers, administrators, materials developers and publishers, whose perceptions and attitudes towards their work may be affected by a test
process	any actions taken by the participants which may contribute to the process of learning, including materials development, syllabus design, changes in teaching methodology, the use of learning and/or test-taking strategies, etc.
product	what is learned (facts, skills, etc.) and quality of the learning (fluency, etc.)

Hughes (1993, cited in Bailey, 1996, p.262) notes that participants’ perceptions and attitudes towards teaching and learning would be firstly

[4] I contacted Bailey by email, enquiring whether the copy of Hughes’ (1993) manuscript is still available. However, she said “I don’t have this paper any longer”.

affected by test nature, their changed perceptions and attitudes may then affect what participants do in their work so as to affect the process of students' learning, and subsequently influence the learning outcomes. He considers that improved learning is the ultimate product of positive washback. This model suggests that to investigate PISA's impact on students' learning: (1) participants who are involved in the process of students' learning through taking actions in response to PISA shall be recruited; (2) their perceptions of PISA, the actions or initiatives motivated by PISA should be identified; (3) the changes in the process of students' learning and subsequent changes of students' learning outcomes could be seen as the evidence of the existence of PISA's impact on students' learning.

As this basic model suggests, the process of students' learning is complex since various participants are involved, albeit at different magnitudes and in different ways. Hence, employing qualitative data and quantitative data is suggested to be used to triangulate and support a balance of evidence for investigating washback (Bailey, 1999). This underpins the significance of employing mixed methods in my research. For example, besides qualitative data on the perceptions of participants about PISA's impact on the process of student learning and learning outcomes, evidence based on quantitative data are also needed for triangulation. From policy initiatives motivated by PISA to school-implementation of these initiatives, different participants involved in the process of students' learning may have different perceptions and interpretations of PISA and its related initiatives. Therefore, it also suggests the need of accounting for different stakeholders' perspectives (Cheng et al., 2015).

To examine PISA's impact on the processes and outcomes of students' learning, it is necessary to further clarify the possible aspects of the changes brought by PISA and PISA-motivated initiatives. From the perspective of empirically examining how washback occurs in the process of teaching and learning, Alderson and Wall (1993, p.120-121) propose a list of possible washback hypotheses, as shown below.

(1) A test will influence teaching.
(2) A test will influence learning.
(3) A test will influence what teachers teach; and
(4) A test will influence how teachers teach; and therefore by extension from (2) above:
(5) A test will influence what learners learn; and

(6) A test will influence how learners learn.
(7) A test will influence the rate and sequence of teaching; and
(8) A test will influence the rate and sequence of learning.
(9) A test will influence the degree and depth of teaching; and
(10) A test will influence the degree and depth of learning.
(11) A test will influence attitudes to the content, method, etc., of teaching and learning.
(12) Tests that have important consequences will have washback; and conversely
(13) Tests that do not have important consequences will have no washback.
(14) Tests will have washback on all learners and teachers.
(15) Tests will have washback effects for some learners and some teachers, but not for others.

The above hypotheses provide insights into my research in terms of identifying possible categories and themes evidencing PISA's impact, for example, learning content, teaching practices, teaching quality, attitudes towards learning, and cognitive performance, which can also be classified as processes and outcomes as defined in the education effectiveness model employed in PISA assessment framework (OECD, 2013a). They also suggest that the impact may be different across teachers and students due to individual differences between them (Alderson and Wall, 1993).

As discussed previously (see Section 3.2.4), the interpretation and translation of PISA data in domestic education systems involves a process of de-contextualisation and re-contextualisation, with the agencies in different contextual levels (e.g. national, local, regional and international) (Carvalho and Costa, 2015). It can be further contextualised during school enactment (Braun et al., 2010; Braun et al., 2011). Similarly, washback effect also suggests that, interpretations and uses of a test "are value laden and situated in political, social, cultural, and educational contexts" (Cheng et al, 2015, p.464). These various contextual factors may mediate the washback effect and should be evaluated when investigating the impact of a test (Bachman and Palmer, 1996; Cheng et al., 2004; Shih, 2009; Cheng et al., 2015). Hence, it is reasonable to conceive that contextual considerations are also essential for investigating and understanding PISA's impact on students' learning brought by the interpretation and use of PISA in a specific context. Although the above listed hypotheses about washback particularly focus on the impact of test at the micro level, that is, effects on teaching and learning, washback effect can be more complex than those listed, with test nature, educational context, and decisions made according to the test results involved (Alderson and Wall, 1993).

Regarding the contextual factors which may mediate washback, Bachman and Palmer (1996) suggest the goals and values of the society and the educational system wherein the test is used. Cheng et al. (2015) classify them as two types: micro-context factors (e.g. school setting) and macro context factors (e.g. society and educational system). Despite the fact that the studies on washback effect have noted the importance of contextual considerations for investigating the impact of a test, neither the basic model of washback mechanism itself nor the washback hypotheses explicitly include and specify contextual factors in detail to allow for empirically investigating how they link with the trichotomy (i.e. participants, process, products) and how they work in mediating washback. Therefore, to investigate PISA’s impact on students’ learning, the mechanism of washback which is employed as the framework needs to be expanded to build into contextual considerations.

3.6.2 Ecological systems theory

Ecological systems theory offers “a foundation for building context into the research model” (Bronfenbrenner, 1979, p.21). This theory suggests that the ecology of human development involves nested contexts with concentric structures denoting micro-, meso, exo-, and macrosystems (Bronfenbrenner, 1979). For human development, for example, students’ learning, it is influenced not only by the immediate settings (i.e. microsystems) (e.g. school, home) in which the developing individual students are present, and the interconnections between these settings (i.e. mesosystems), but also by larger contexts (i.e. exosystems and macrosystems) in which the immediate settings are embedded (Bronfenbrenner, 1979).

According to ecological systems theory, the first layer of contexts that developing individuals are embedded in is microsystem which refers to “a pattern of activities, roles, and interpersonal relations experienced by the developing person in a given setting with particular physical and material characteristics” (Bronfenbrenner, 1979, p.22). For students, the typical microsystems can be school settings and home settings. During the process of students’ learning at school, students interact with peers and teachers through classroom activities arranged by teachers. Teaching practices and quality that are experienced and perceived by students might contribute to the development of students' learning in terms of, for example, students' attitudes toward and beliefs in learning, and also their performance.

The mesosystem denotes "the interrelations among two or more settings in which the developing person actively participates" (Bronfenbrenner, 1979, p.25). For students, the typical settings that they may move from one to another would be school and home. So far, whether PISA's impact on students' learning through PISA or PISA-motivated initiatives implemented at schools could be enhanced or impeded by home context factors is not yet clear, since related research is rarely documented. I will retain this category of context layers in the framework for now and will probe this in the empirical research when it is possible.

Moving beyond the immediate systems and the interconnections among these systems is the exosystem, referring to "one or more settings that do not involve the developing person as an active participant, but in which events occur that affect, or are affected by, what happens in the setting containing the developing person" (Bronfenbrenner, 1979, p.25). As introduced above (see Section 3.3.2), in the local context focused in my research, a number of PISA-motivated initiatives have been launched with the interpretation of PISA concepts and results underpinned to improve the quality of local education. Although students do not directly participate in the launch of these initiatives, they might also have reciprocal relations with these local PISA-motivated initiatives. On the one hand, their performance perceived by policymakers will affect the formulation of these initiatives; on the other hand, students will be affected by these initiatives through school-level implementation of these initiatives on teachers' teaching.

Despite differences among settings within and between micro-, meso-, and exosystems, it is suggested by the ecological systems theory that, within a given society or social group, all these lower-order systems share some consistencies in terms of common culture, belief or ideology of an even broader context, namely, the macrosystem (Bronfenbrenner, 1979). Like Cheng et al. (2015) suggest, the education system in the national context would be one of the important settings at macro-level. Factors in this system, for example, national curriculum and assessment, would be consistently of high value for local educational policymakers, schools leaders, teachers, parents, and students. They might each play a role mediating the PISA's impact on student learning as well, subject to local education policymakers and school-level practitioners' perceptions of the importance of these factors and PISA.

In summary, the conceptual framework developed as above provides insights into my research in terms of shaping the methodological design. It will also serve as the theoretical underpinning for my research data collection, data analyses, and as the basic framework for conceptualising my findings. With the data analysis, this framework will be further refined and developed to portray the mechanism of PISA's impact on students' learning in the local context focused in my research.

3.7 Summary

This chapter has overviewed the background of the launch of PISA, pointed out its policy-driven attributes and increasing global reach (see Section 3.1). It then covered extant discussions and debates on PISA's impact in domestic education contexts, and the issues that arise along with domestic policy responses, such as policy borrowing and PISA's expanding governing power over global education (see Section 3.2). More specifically, this chapter has provided a review of how PISA has been used in educational policies and practices at national and regional level in mainland China, and the impact PISA has brought in Chinese context (see Section 3.3). Following these discussions on PISA's policy impact, this chapter then dicussed PISA results in relation to the factors associated with students' performance (see Section 3.4), and reflected on the limitations in PISA (see Section 3.5), both of which are essential for making policies based on robust evidence drawn on PISA data, and also for evaluating the link between the claimed policy responses and PISA data. Building on the literature review, this chapter has developed the conceptual framework, employing theories about washback effect and the ecological systems theory for underpinning and informing the empirical investigation of my research.

The literature reveals that the policy impact of PISA in a range of domestic education systems has been widely claimed. As to China, for example, PISA's impact on education has manifested not only in the national context but also in regional and local contexts. The policy responses across education systems as illustrated before suggest that, the extent to which these systems have reaction to PISA are various and dynamic, although it appears that there is some policy convergence (e.g. developing national testing system, incorporating PISA-like content into national curriculum). We also see that there have been concerns about the negative consequences in terms of focusing on a narrow range of education contents. However, as the literarure argues, for the policy responses claimed as drawing on lessons from PISA high performing systems or taking measures to align with PISA framework, internal factors may recontextualise the external contents and adapt them to fit domestic context. Moreover, the literature suggests that the policies claimed as responses to PISA are often only loosely linked to PISA data, or even just use PISA as a tool providing sound legitimation for policy reforms. It is noted that the rankings on PISA's international league tables which are genereated with students' average performance seem to be the major concern of the education systems that have pariciapted in PISA, and are mostly referenced in policymaking. This is not surprising, considering the inherent attribute of PISA which focuses on policy-driven interests, and performance in PISA is considered by policymakers as one of the indicators of potential global competitiveness. However, the complexity of potential factors underliying students' performance and the limitations of PISA itself signal that understanding and interpreting the effectiveness of an education system simply based on the average of student performance is problematic and may even be misleading for informing policymaking.

These implications raise the question that evaluating PISA's impact according to the texts or speeches of policies which are claimed as responses to PISA may not be sufficient to understand how PISA is used and influences domestic education in practice. This suggests that it is essential to look at policies and beyond with contextual considerations when analysing the impact of PISA in a given domestic context.The increasing impact of PISA on global education policymaking, the further expansion of PISA reach to school level, and the concerns about the possible negative consequences brought by PISA stress the need for investigating the impact of PISA on students' learning in domestic contexts. However, extant studies mainly focus on PISA's impact at policy level, while few of them take a further look the more fine grained local/school/student level. The synthesis of the literature, as summarised above, provides the rationale for my research focusing on this under-researched area. It also provides the basis for building the conceptual framework which further employs theories about washback effect and ecological systems theory to underpin the investigation into PISA's impact on students' learning with consideration of factors in different contextual levels. The conceptual framework contributes to shaping the methodogical design, and informing data analyses and interpretation of my research. In the next chapter, I will focus on presenting the methodology of my research.

Chapter 4 Methodology

This chapter includes eight sections respectively introducing my research aims and research questions, research paradigm, research design, methods, data collection, data analyses procedures, ethical considerations, and trustworthiness.

4.1 Research aims and research questions

My research aims to empirically investigate PISA's consequences in terms of its impact on students' learning in schools in a local Chinese context, that is, Fangshan District of Beijing. It attempts to first understand the overall picture of the impact of PISA on students' learning and how this impact occurs in the local context; and then further investigates in detail its impact on students' learning in a specific learning domain, mathematics. Under the conceptual framework (see Section 3.6), the research was guided by addressing the following three research questions:

RQ 1. How is PISA used by Fangshan educational policymakers and practitioners in the local context?

RQ 2. What are the perceptions of Fangshan educational policymakers and practitioners of the impact of the use of PISA in educational policies and practices on mathematics learning?

RQ 3. To what extent have the PISA-motivated initiatives actually influenced students' learning outcomes in mathematics, as evidenced by PISA data?

For RQ 2 and RQ 3, as shown above, only mathematics was focused. This is because, amongst the three local PISA-motivated projects, the mathematics project focuses on improving learning through improving teaching practices and also had been undertaken for several years at the time of my research data collection (see Section 3.3.2). Compared with other subjects, mathematics is considered as more suitable for investigating the effects of OTL, since the process of learning mathematics concepts basically takes place in schools (Törnroos, 2005).

For the aim of portraying the impact of PISA in Fangshan on students' learning, RQ 1 is to identify how Fangshan makes use of PISA, and how PISA data, such as PISA results, databases, assessment frameworks and released items, are interpreted and translated in the local PISA-motivated initiatives. As suggested by the trichotomy washback model (Hughes 1993, cited in Bailey, 1996), the perception of participants involved in the test washback would be first influenced by the test, and would then mobilise participants to undertake corresponding actions (e.g. making changes to teaching practices). Moreover, as indicated by the ecological systems theory (Bronfenbrenner, 1979), when policies are put into practice, they may be moderated by contextual factors in school settings. Furthermore, informed by the ecological systems theory (Bronfenbrenner, 1979), consideration of potential influence of the ongoing reforms in national education context (see Section 2.2 and 2.3) is taken. Hence, to seek the answer to RQ 1, the below four sub-questions were addressed:

1a. How do Fangshan educational policymakers perceive PISA?
1b. What kind of initiatives have been motivated by PISA due to these perceptions?
1c. How are these initiatives implemented at school level?
1d. What is the influence of national education contextual factors on the formation and implementation of PISA-motivated initiatives?

Following RQ 1, RQ 2 is used to particularly explore the perceived impact of PISA on the process of student mathematics learning in schools and learning outcomes. As indicated by theories about washback effect, any actions taken by participants would influence the process of teaching and learning in terms of what and how students learn, and this would subsequently influence learning outcomes (Alderson and Walker, 1993; Hughes, 1993, cited in Bailey, 1996). Therefore, to address the RQ 2, answers to the below three sub-questions were explored.

2a. What are their perceptions of the impact of the use of PISA in educational policies and practices on the process of mathematics teaching and learning in schools?
2b. What are their perceptions of the impact of the use of PISA in educational policies and practices on mathematics learning outcomes?

2c. What is the perceived role of national education contextual factors in the impact on mathematics learning?

RQ 3 is used to further triangulate the perceptions about the development of students' mathematics performance brought by PISA's impact, and the factors contributed to the development, by seeking evidence from quantitiative data of PISA. It is also used to expand the understanding of the use and translation of PISA data in educational policymaking in the local context.

4.2 Research paradigm

My research was carried out on the basis of dialectical pragmatism (Teddlie and Johnson, 2009; Johnson and Christensen, 2017). Pragmatism allows researchers to focus on and practice "what works" for best seeking answers to research questions in designing and conducting research (Johnson and Onwuegbuzie, 2004; Johnson and Christensen, 2017), rather than being "the prisoner of a particular [research] method or technique'' (Robson, 1993, p. 291). Pragmatism acknowledges that relationships that follow structural regularities and patterns are moderated by human elements (Yvonne Feilzer, 2010). As a type of pragmatism, dialectical pragmatism also suggests that researchers should consider quantitative as well as qualitative viewpoints (Tashakkori and Teddlie, 2010). It provides "a supportive philosophy for mixed methods research" (Tashakkori and Teddlie, 2010, p.88). The qualifier "dialectical" specifically suggests that in this type of pragmatism, not just ideas from multiple forms of data are embraced, but also perspectives of multiple research disciplines and participant groups are valued (Johnson and Christensen, 2017). This paradigm concurs with the theoretical foundation of my research which combines washback effect in the discipline of language testing with the ecological systems theory in the discipline of psychology. It philosophically underpins the mixed methods design employed in my research which drawn on ideas from both qualitative and quantitative data and different groups of participants for understanding the impact of PISA on students' learning. In the following, I will present my research design in detail.

4.3 Research design

A mixed-methods design which involved qualitative and quantitative data was adopted in my research, according to the insightful conceptual framework in which different types of data are suggested to be collected to triangulate and support a balance of evidence that the washback does occur (Bailey, 1999). It is known that both qualitative and quantitative methods have their strengths and weaknesses. For example, with large sample size, quantitative methods are advantaged in terms of generalisation, while qualitative methods can capture depth, detail and meaning although they typically employ a small sample (Patton, 2002). Johnson and Christensen (2017) consider that the use of multiple research methods is a strength in educational research and that qualitative and quantitative methods can overcome the weaknesses of each other and complement each other.

There are several types of designs for mixed methods research. Choosing a design depends on the purpose of using mixed methods (Johnson and Christensen, 2017) and the feasibility of data collection. In my research, the main purpose of the employment of both qualitative and quantitative data is for triangulation, that is, comparing and contrasting results of different data strands and seeking convergence and corroboration between them (Greene et al., 1989; Creswell and Clark, 2011). Therefore, the convergent design which allows one "to obtain different but complementary data on the same topic" (Morse, 1991, p.122) was used. In line with this design, in my research, qualitative data and quantitative data were collected or obtained independently and concurrently. Data analyses of the two data strands were also conducted separately as well, yet, they were informed by each other to some extent. For this design, the two strands were merged at the stage of interpretation and discussion (Creswell and Clark, 2011). Qualitative data were employed to investigate how PISA is used in Fangshan, and also to investigate local educational policymakers' and practitioners' perceptions of PISA's impact on student mathematics learning at schools. To triangulate these perceptions, quantitative data were employed to examine students' mathematics performance trends and to explore explanatory factors explaining students' mathematics performance. The results of the two data strands were synthesised based on the synonymous themes identified from both strands to understand the impact that PISA has brought to students' mathematics learning in the local area. The technique of Pillar Integration Process (PIP) (Johnson et al., 2019) was employed for the synthesis. This technique allows for identifying the consistency and contradiction between what people perceived about PISA's impact on students' learning and what quantitative data tell about students' performance, and also for expanding the understanding of one strand of data from the perspective of the other strand (Johnson et al., 2019). Under this technique, categories identified from each of these two strands are listed, matched, checked, based on which the synonymous themes were built. The overall mixed methods design of my research is as shown in Figure 4.1 below.

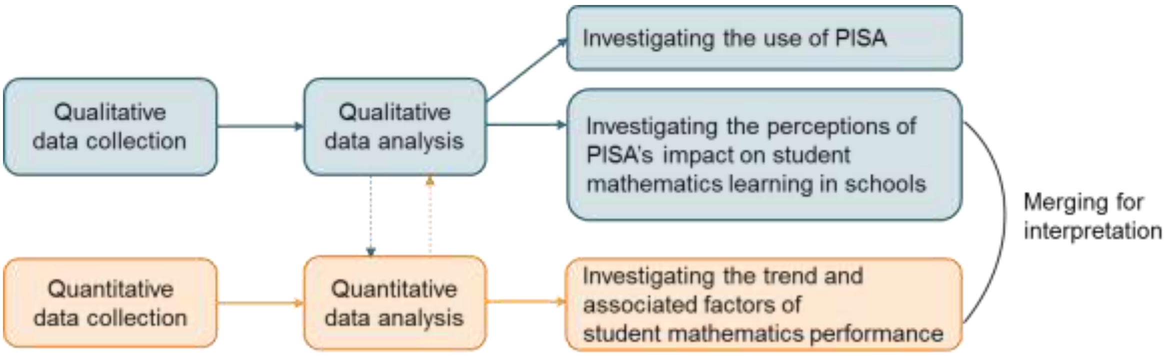


Figure 4.1 Mixed methods design

4.4 Research methods

To address the first two research questions, interviews were used. Interviews are more adaptable than questionnaires in investigating motives and feelings, and could collect information (e.g. facial expression) that a written response would not reflect (Bell, 2005). In my research, the interview participants were sampled purposively. That is, individuals who could inform the answers to the research questions were selected (Creswell, 2007). In terms of the sample size, the qualitative approach usually recruits a small number of participants who could provide in-depth information (Creswell, 2007). According to the practices of research into washback, various groups of participants are suggested to be recruited for data triangulation (Cheng et al., 2004). In addition, the ecological systems theory also indicates that participants in different contextual layers can play a role in influencing students' learning (Bronfenbrenner, 1979). With these considerations, my research recruited local educational policymakers as well as practitioners. For RQ 1, one local educational official who is responsible for the administration of PISA in Fangshan and deeply engaged in the education reforms at the local area was recruited in addition to the key informants of the three PISA-motivated projects. From these interview participants, first-hand information about how Fangshan interpreted PISA data, and translated them into local educational policies can be obtained. Besides, school leaders and mathematics teachers were recruited to identify the enactment of the local PISA-motivated initiatives at schools. The interviews with these educational policymakers and practitioners were also used to understand the perceived impact on mathematics learning for addressing RQ 2. As introduced in Section 3.3.2, literature has briefly described what Fangshan has initiated in light of PISA, and I have got clear topics for the interviews. Hence, a semi-structured approach which covers specific topics and is flexible in pursuing rich information (Bryman, 2016) was adopted for all interviews.

For the students who experienced the local PISA-motivated initiatives through school enactment, by the time when I did field work, autumn of 2017, they had upgraded from lower-secondary schools to upper-secondary schools, and most of them had completed their secondary education. Therefore, I did not investigate students' perception of the impact of these initiatives on their learning. I also did not conduct classroom observation since the comparable data of observing the classroom teaching and learning collected before the initiation of the local PISA-motivated projects were not available.

To address RQ 3, quantitative data drawn from Fangshan PISA databases were employed, including student performance on cognitive tests and student questionnaire data, to estimate the change of students' mathematics performance across the three cycles of PISA that Fangshan had participated in, and to identify the factors associated with the mathematics performance. Domestic longitudinal data which include comprehensive indicators for tracking and explaining the development of individual students' performance are not available, while PISA at least provides rich data for identifying the performance trends of student population and the effects of explanatory factors on students' performance. As I introduced before (see Section 2.3), PISA China Trials only involved students from academic schools rather than those from all eligible schools. Hence, my research only focused on academic school students.

4.5 Data collection

In this section, I will first introduce the qualitative data collection, then I will introduce the obtainment of Fangshan PISA data.

4.5.1 Qualitative data collection

4.5.1.1 The context of the fieldwork site

The fieldwork of my research was conducted in Fangshan District of Beijing in China. Fangshan District, located in the southwest of Beijing, has an area of 2019 square kilometres, with mountains, hills, and plains each accounting for one third of its total area (Beijing Fansghan government, 2017). Fangshan has a long history with a number of historic cultural sites (e.g. the Peking Man Site at Zhoukoudian) and scenic places. It has a resident population of 1.046 million (Fangshan District Bureau of Statistics and Fangshan Survey Team of National Bureau of Statistics, 2016).

By the end of 2015, in the geographical area of Fangshan, there were 56 secondary schools with 40631 enrolled students (Fangshan District Bureau of Statistics and Fangshan Survey Team of National Bureau of Statistics, 2016). The details of schools and students in basic education in Fangshan are shown in Table 4.1.

Table 4.1 Schools and students in Fangshan (including Yanshan Subdistrict)

School Type			Number of Schools	Number of Enrolled Students
Primary			108	48192
Secondary	Academic	Lower-secondary	32	16679
		Upper-secondary	15	9139
	Vocational	Upper-secondary	9	14813
Total			164	88823

There are several reasons for choosing Fangshan as the local area focused upon in my research. First, it had continuously participated in

PISA for three cycles since PISA 2009 China Trial (Wang et al., 2017) and it published its results (Wu, 2015). In PISA 2009 China Trial and PISA 2012 China Trial, it took part in this assessment with samples representing eligible student population in this local area. And then, with the student sample from all its eligible 46 schools, Fangshan participated in PISA 2015 as a part of Beijing. Additionally, Fangshan local policymakers explicitly claimed that its local PISA scores and PISA assessment concepts had been actively used in motivating and informing a number of policy initiatives for improving teaching and learning in mathematics, science and reading in its local context (Wu, 2015). Guo, the former director of Fangshan Education Commission and colleagues stated that PISA is used in this local area “to promote the professional development of teachers, improve curriculum, and guide education reform at local level” (Guo et al., 2015, p.9).

4.5.1.2 Piloting

Considering that some of the potential interview participants are specific individuals, for example, the local educational official and the key informants of the local PISA-motivated projects, I was not able to do a full pilot of interview guides (see Appendix A.1). Instead, I did an informal pilot by interviewing two PhD students who had had rich experience of conducting interviews to familiarise myself with the interview procedure and to improve the interview guides. For example, according to their suggestions, I revised the interview guides in terms of using more general questions rather than detailing the specific aspects intended to probe. Before I started to do interviews in the field site, I additionally piloted the interview guide for mathematics teacher participants through interviewing a mathematics teacher of an upper-secondary school in another district of Beijing. Based on this pilot, I further modified the warm-up question by asking about teaching tasks on the day or work experience in the school for icebreaking; and took his advice on the specific wording to show my appreciation of teaching profession to make mathematics teachers more engaged in the interview.

4.5.1.3 Sampling of interview participants

There are 16 interview participants in total involved in my research, including one local educational official, the reading project leader, the mathematics project leader, the science project assistant, five school leaders, one reading teacher, and six mathematics teachers. Their basic information and the codes/pseudonym names used for anonymisation is shown below in Table 4.2.

Table 4.2 The participants of interviews

Participant	Participant code/pseudonym	School affiliated	School type	Expertise
A local educational official	Local educational official	NA	NA	Educational assessment
The reading project leader	RL	NA	NA	Reading, Chinese
The mathematics project leader	ML	NA	NA	Mathematics
The science project assistant	SA	NA	NA	Science
A leader of school 1	SchL1	School 1	Rural area, Upper-secondary	Mathematics
A leader of school 2	SchL2	School 2	Town, Upper-secondary	Geography
A leader of school 3	SchL3	School 3	City, Lower-secondary	Chinese
A leader of school 4	SchL4	School 4	Town, Lower-secondary	Reading, Chinese
A reading teacher of school 5	Sch5-R	School 5	Rural area, Lower-secondary	Reading, Chinese
A leader of school 6	SchL6	School 6	City, Upper-secondary	English
A mathematics teacher of school 1	Linyun	School 1	Rural area, Upper-secondary	Mathematics
A mathematics teacher of school 2	Mobai	School 2	Town, Upper-secondary	Mathematics
A mathematics teacher of school 3	Yunlian	School 3	City, Lower-secondary	Mathematics
A mathematics teacher of school 4	Wenting	School 4	Town, Lower-secondary	Mathematics
A mathematics teacher of school 5	Zhushan	School 5	Rural area, Lower-secondary	Mathematics
A mathematics teacher of school 6	Mingyi	School 6	City, Upper-secondary	Mathematics

Note: NA refers to Not Applicable.

As shown in Table 4.2 above, a local educational official who is responsible for the administration of PISA and research on utilising PISA data in Fangshan was recruited. With the recommendation of this official, the reading project leader (RL), the mathematics project leader (ML), and the science project assistant (SA) were recruited. Regarding the science project, it had been led by a new leader since January 2017, while the new science project leader was not involved in the previous work of this project. Compared with the new leader, SA had been involved in the work since the preparation stage of this project and also engaged in its research activities. Therefore, my research involved SA rather than the leader.

Besides the involvement of various groups of participants, I also took into consideration the potential influence of characteristics of schools and teachers in interviewee recruitment. Variants of their characteristics were intentionally covered by selecting schools of different educational levels and school locations and selecting teachers of different degrees of involvement in PISA-related work. At school level, according to school characteristics (i.e. city/town/rural area, lower/upper secondary) and the extent of schools' involvement in PISA-motivated initiatives, six secondary academic schools recommended by the local educational official, RL, ML, and SA were recruited. By the time I was recruiting participants, there was no longer upper-secondary schools in Fangshan rural area due to school merger. So the upper-secondary programme of one rural complete school (i.e. School 1) was selected instead. School leaders were selected from those who have been working in the affiliated schools as a leader for years and have some knowledge about the local PISA-motivated initiatives. In five of the six schools, I interviewed one school leader of each school. For the remaining one school (i.e. School 5), the eligible school leader was just transferred to another school in August 2017, before my data collection. Although I could not recruit a school leader in this school who can contribute rich data meaningful for answering my research questions, I interviewed one reading teacher (i.e. Sch5-R) recommended by the school contact, considering that this participant was engaged in PISA-motivated reading initiatives. In each of the six schools, I also interviewed one mathematics teacher recommended by school leaders. The mathematics teacher participants had taught in their corresponding affiliated schools for years. Five mathematics teacher participants, to various degrees, were involved in Fangshan PISA-motivated initiatives, while one mathematics teacher participant seemed not have awareness of the local PISA-motivated mathematics initiatives.

I invited each of the interviewees to participate in my research by phone (see the phone scripts in Appendix A.2) at his or her convenience, after the attempts to approach them and introduce myself and my research via text messages. To recruit school-level interviewees, I also sent a letter to the school gatekeepers (see Appendix A.3), requesting permission to conduct research in their schools.

4.5.1.4 The process of interviews

My field work took two months from September to October in 2017. Interviews with the local educational official, RL, SA, ML, school leaders (or the reading teacher), and mathematics teachers were conducted successively. All interviews were semi-structured, carried out individually and face to face in quiet rooms in their working places. Generally, for the same school, I interviewed the school leader (or the reading teacher) first and then interviewed the mathematics teacher. According to participants' responses and their available time, the duration of each interview was different. Generally, the interview with the local educational official, RL, SA, or ML took about 1 hour to 1.5 hours, while each interview conducted at school level took between 40 minutes and 1 hour.

During the fieldwork, the interview guides were sent to the interviewees of corresponding groups a few days before the interview. To collect basic demographic information for allowing me to get to know some background of interviewees, a brief questionnaire (see the sample in Appendix A.4) was prepared, and given to each interviewee to complete before the commencement of the interview. Although the interview guide for all school leaders is the same, during the interviews, I adjusted the order of some questions according to school leaders' discipline background by starting from something they were familiar with.

All interviewees are Chinese and all the interviews were conducted in Mandarin. With the approval of the participants, all the interviews were audio recorded. In each interview, I used two recorders in case there would be something wrong with the recorder in the fieldwork site. Before I went to one site I checked the power of the recorders. I also checked the recordings as soon as I was back from one site. During interviews, I made some field notes about the keywords or important ideas mentioned by interviewees. However, to keep eye-contact with interviewees and engaged in the discussion, I did not spend an amount of time on note-taking.

The written consent was obtained from each interviewee. When contacting each school for the first time, I introduced my research project and how school leaders and mathematics teachers participate in it. After inviting them to participate in my research, I immediately sent the electronic version of the information sheet (see the sample in Appendix A.5) and the consent form (see the sample in Appendix A.6) to allow the potential participants to have sufficient time to consider their participation and ask me questions. A few days later after the invitation, I contacted them again and asked about their willingness to participate and arranged the interview time according to their availability. In the fieldwork site, I introduced my research to the interviewee again, explained the research ethics requirement and gave him/her chance to ask me questions before he/she signed on the consent form.

4.5.2 Quantitative data obtainment

According to the confidentiality agreement about PISA China Trial signed by the OECD and the NEEA, the authority where China PISA National Centre was based, the NEEA takes the responsibility of managing databases of PISA 2009 China Trial and PISA 2012 China Trial, any use of these data needs the approval of the NEEA. For PISA 2015, although the data of China (B-S-J-G) were released by the OECD for public use, data of specific regions could not be identified from the public database files due to the sampling design and anonymity; while in the database files which were exclusively delivered to the National Centre, Fangshan data could be identified and split from the whole database. Therefore, I applied for the use of Fangshan PISA data of the three cycles by emailing China PISA National Centre with an application form attached. In the application form (see Appendix A.7), I included detailed information to allow China PISA National

Centre clearly understand what my research is about, why I needed the databases of Fangshan I was applying for, and how I would use these data. Since Fangshan data of the three cycles of PISA are not in the public domain, use of the data needs to follow the confidentiality requirement. Therefore, I signed a confidentiality statement for getting the approval for using these data. In this confidentiality statement (see the statement in Appendix A.8 and its English version in Appendix A.9), I agreed to not disclose or distribute these data to any individuals or institutions.

In PISA 2009 and 2012, as required by PISA sampling standards (OECD, 2012; 2014a), 35 eligible students were expected to be randomly selected within each sampled school; for the schools which have less than 35 students, all eligible students would be sampled. In PISA 2015, this required value had been changed to 42 for the countries adopting computer-based assessment mode which was the case for China (OECD, 2017). Table 4.3 shows the sample size of Fangshan in each PISA cycle.

Table 4.3 Fangshan's sample size in PISA

Fangshan's participation in PISA	School sample size	Student sample size	Mean assessed students per school	Median assessed students per school	Min assessed students per school	Max assessed students per school
PISA 2009 China Trial	25	610	24	27	11	34
PISA 2012 China Trial	23	624	27	32	14	35
PISA 2015[5]	42	1407	34	42	4	42

4.6 Data analysis procedures

In this section, I will first describe the procedure of analysing interview data with the thematic analysis approach (Braun and Clarke, 2006; Braun et al., 2016). Following that, I will present the procedures of doing quantitative analyses of PISA data, including the trend analysis for identifying students'

[5] In PISA 2015, besides the students from academic schools which are displayed in Table 4.3, 156 students from 4 vocational schools participated in this assessment as well.

performance trends, and multilevel modelling for exploring factors accounting for students' performance.

4.6.1 Thematic analysis of the interview data

My research aimed to explore the patterns of the perceived impact of PISA in a specific educational context, namely, Fangshan District of Beijing, and describe and interpret the patterns within this context. Therefore, thematic analysis which identifies, describes and interprets patterns (themes) across data (Braun and Clarke, 2006; Braun et al., 2016) was employed in analysing the interview data. Compared with other methods such as narrative analysis, thematic analysis is flexible because it is not tied to a particular theoretical framework or research paradigm (Braun and Clarke, 2006; Robson, 2011; Braun et al., 2016), and can be used to answer a wide range of types of research questions (Braun et al., 2016). For example, it can be employed to understand participants' actions and practices, and their views and perspectives (Clarke and Braun, 2017). Its flexibility allows me to seek answers to RQ 1 and RQ 2 with interview data, and to synthesise the results from the qualitative strand and the quantitative strand, and interpret the results in detail.

In terms of the process of applying thematic analysis, Braun and Clarke (2006) propose a step-by-step guide which includes six phases: familiarising with data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. They note that the process of thematic analysis is recursive rather than linear (Braun and Clarke, 2006). Perhaps because of this characteristic of the analysis process, later, Braun, et al. (2016) further developed the classification of the six phases, proposing phases 1-2 “Familiarization and coding”, phases 3-5 “Theme development, refinement and naming”, and phase 6 “Writing up”. They additionally suggest that, at the phase of familiarization, one reads the data analytically and identifies ideas that are helpful for addressing research questions. It is also admitted that writing is along with the analytic process and is an integral part of it (Braun and Clarke, 2006; Braun et al., 2016). Hence, phase 6 involves “compiling, developing, and editing existing analytic writing, and situating it into an overall report” (Braun et al., 2016, pp.200-201). In my research, I basically followed these phases for analysing interview data.

Firstly, I familiarised myself with the data. I transcribed the audio recordings of all 16 interviews manually by myself, and also captured some non-language behaviours such as laughs and pauses, which I considered meaningful, in transcripts. To annotate the details in the speech, I created a transcription notation system (see Appendix A.10). For example, “{ }” is used to indicate speakers’ nonverbal expression and behaviour, and **bold font** indicates emphasised words. During transcription, I made comments and marks on transcripts when I noticed information of potential importance for addressing my research questions. After I completed transcription, I checked all the transcripts against the audio recordings and then anonymised transcripts. Through transcription I have become familiar with the data to some extent. However, that process focused more on the language of segments in speech, rather than the content of whole data. Therefore, I read and reread each transcript to further familiarise myself with the data, with attention on the meanings and patterns, and the potential meanings of non-language behaviours like laughs and pauses in the data. From this step, I started to write memos (see Appendix A.11) for recording the development of my thoughts (e.g. initial ideas about the patterns in the data) during interview data analysis.

Secondly, I moved on to generate codes. A code is a label that identifies an interesting aspect in the data, which is potentially relevant to the research questions (Braun et al., 2016). At the second step, I systematically identified and labelled (so-called “code”) important data extracts which would be potentially useful for answering my research questions. To assure the coherence of coding, as suggested by Braun et al. (2016), after the initial codes were generated, I read through the whole data for the second time to review all the codes and made them coherent. Since my research employed mixed methods design that the qualitative data analysis results were to be synthesised with quantitative results, after I got the quantitative results, I read the interview data and reviewed the codes again to identify whether I omitted any codes which are corresponding with some variables in quantitative analyses.

Following coding, the third step was to group codes, develop and refine themes. At this step, first, I looked at the relationships amongst codes and their relevance to my research questions. Then I clustered the codes which share same broader patterned meaning or concept in relation to the use of PISA, perceived impact on the process of teaching and learning, perceived impact on learning outcomes (non-cognitive and cognitive). After I had got candidate themes, I read all the transcripts again and moved on to review these themes to check: (1) each of the themes captures the ideas in the data well; (2) they are distinct with each other; and (3) they are tied to my research focus, coherently telling a story about the data. The themes, each of which has a distinctive focus and scope were finally identified, when they did not need great changes for addressing the research questions. At this step, I also gave each theme an informative and concise name.

During thematic analysis, I also analysed whether there are links between the themes and whether the themes vary in the data of different groups (e.g. educational policymakers versus practitioners) of interviewees and in the data of different interviewees in the same group (e.g. school leaders; mathematics teachers), as suggested by Bryman (2016).

Coding and developing themes was conducted in QSR International's NVivo 11, a qualitative data analysis software (see the screenshot of categorising codes in Appendix A.12). Memos were written in NVivo 11 as well. Compared with manual analysis on paper, employing this software in data analysis is more efficient to manage data extracts and visualise data results. The data extract used for inter-reliability check and the quotes employed in presenting and discussing findings in the thesis were all translated from Mandarin to English.

4.6.2 Quantitative data analyses on PISA data

Quantitative data analyses were conducted to examine students' mathematics performance trends over three PISA cycles from 2009 to 2015, and also to explore the factors associated with students' mathematics performance in PISA.

4.6.2.1 Trend analysis

Firstly, the observed performance trend analysis was conducted. That is, calculating students' average mathematics performance in each of the three cycles, and then identify the changes of means over the three cycles. Secondly, the trend performance after adjusting the changes in the composition of target student population over years was calculated. As I discussed in Literature Review (see Section 3.4.1), changes in population composition, for example, differences in the proportion of age groups, gender, and grade may have influence on interpreting performance trends (Vista, 2013; Aloisi and Tymms, 2017; Melkonian and Areepattamannil, 2018). Table 4.4 shows the composition of Fangshan student samples in terms of these aspects in the three PISA cycles.

Table 4.4 Demographic composition of Fangshan's student sample in PISA

Categories		**Unweighted (%)**			**Weighted (%)**		
		2009	2012	2015	2009	2012	2015
Age	15 years and 3-5 months	17.2	18.3	15.1	16.6	19.1	21.6
	15 years and 6-8 months	24.9	27.1	16.6	24.0	28.1	20.9
	15 years and 9-11 months	27.3	27.5	22.0	27.4	27.0	23.5
	16 years–16 years and 2 months	30.6	27.1	46.3	31.9	25.8	34.0
Gender	Girl	47.9	46.8	48.8	45.1	50.1	50.4
	Boy	52.1	53.2	51.2	54.9	49.9	49.6
Educational level	Lower-secondary	37.9	38.1	67.8	51.6	20.9	45.6
	Upper-secondary	62.1	61.9	32.2	48.4	79.1	54.4

The target population of PISA is defined as students aged from 15 years and 3 months to 16 years and 2 months (commonly known as 15-year-olds) who are enrolled in 7th grade and higher (OECD, 2013a). As shown in Table 4.4 above, students' age was divided by three-month interval according to the way the OECD did this in PISA reports. In Fangshan, students of this population are enrolled in lower-secondary (7th-9th grade) or upper-secondary (10th-12th grade) schools, and most of them are in 9th grade and 10th grade. Because of this, PISA 2009 China Trial and PISA 2012 China Trial only included students in these two grades. Therefore, instead of grade, educational level is used in addition to age and gender for adjusting performance trends. In Table 4.4, values in the "Unweighted (%)" columns and "Weighted (%)" columns respectively refer to the demographic composition of student sample, and that of student population represented by the sample. From this table, we see that the composition in age, gender, and especially in educational level in the student population was not stable across the three PISA cycles. To adjust the changes in the population composition, in line with the PISA approach to adjusted trend analysis (OECD, 2017), the composition of Fangshan academic school students in PISA 2015 was used as a reference, based on which Fangshan student data in PISA 2009 China trial and PISA 2012 China Trial were reweighted. With the reweighted databases of these two cycles and PISA 2015 database, adjusted performance trends which are the average mean performance in each cycle were then calculated.

In PISA 2009 and 2012, for each assessment domain, five plausible values (PVs) were drawn for each student as performance estimates from the estimated ability distribution (OECD, 2014a); in PISA 2015, ten PVs were generated per student (OECD, 2017). As introduced before (see Section 3.1), on PISA international scale, PVs of each domain have a mean=500 and SD=100 for OECD countries' average in the cycle in which the domain was the majority domain for the first time (OECD, 2014a). All the whole sets of PVs, replicate weights which account for measurement errors, and student final weights which account for sampling errors in the three cycles of Fangshan PISA data were involved in trend analysis. This analysis was conducted by using IDB IEA Analyzer 4 (IEA, 2018) in conjunction with SPSS 23 (IBM Corp, 2015).

4.6.2.2 Multilevel modelling analyses

In PISA data, students are nested in schools since they were sampled within selected schools (see Section 3.1). Taking this hierarchical structure of data into consideration, two-level modelling was conducted to identify the factors associated with students' mathematic performance. Level 1 is student level, while level 2 is school level. Since amongst the three cycles in which Fangshan had participated, PISA 2012 focused on mathematics, Fangshan data in PISA 2012 China Trial were employed for multilevel modelling (MLM). This section first describes the explanatory variables involved in multilevel analysis, and then introduces the approaches of data processing, and model building.

4.6.2.2.1 Explanatory variables

As introduced before (see Section 3.1 and Section 3.4), according to the school effectiveness model used in PISA assessment framework (Purves, 1987; Scheerens, 1990; Scheerens and Bosker, 1997; OECD, 2013a), explanatory variables could be classified as input which is related to student background or school background, processes of teaching and learning (e.g. OTL, teaching practices, and teaching quality), and outcomes (e.g. non-cognitive outcomes such as motivation and self-beliefs). As suggested by Hughes' (Hughes, 1993, cited in Bailey, 1996) washback model (see Table 3.1 in Section 3.6.1), participants' actions would influence the process of teaching and learning so as to influence students' learning outcomes. Therefore, as the conceptual framework indicates (see Section 3.6), processes variables and non-cognitive outcomes variables were employed in MLM. According to the conceptual framework developed earlier (see Section 3.6), neither the school effectiveness model nor the washback model explicitly suggests students' learning outcomes in reading as variables explaining mathematics performance. However, students' reading performance was still employed as an explanatory variable in MLM, since interview data indicate that students' reading ability may have effect on their mathematics performance in PISA (see Section 5.2.2.2.1). Student and school background were involved for controlling for background effects in interpreting results of MLM.

Selection of explanatory variables also took into consideration the potential influence of students' response styles. The processes variables (e.g. exposure to applied mathematics, cognitive activation) and non-cognitive outcomes variables (e.g. interest, self-efficacy) involved in PISA student questionnaires are generally constructed based on Likert-type scales which are commonly used in a range of surveys. On this kind of scale, students are asked to place themselves into a category of the level of their agreement with the statements about their mathematics learning process or non-cognitive outcomes. It is argued that this question format poses the issue of interpersonal incomparability, since the ways that individuals understand the 'same' question might be vastly different and the perceived cut-points between categories (e.g. strongly agree/agree/disagree/strongly disagree) may differ across individual respondents (Brady, 1985). This incomparability may not only exist across individuals, but also across cultures. For example, it has been evidenced that East Asian students compared to students from other background are more likely to show midpoint responding (MPR) style (Chen et al., 1995; Cheung et al., 2018). Since response styles, such as acquiescence response style (ARS), extreme response style (ERS) and MPR, make observed scale scores deviate from true scores, they can bring bias estimation of relationships between variables (Baumgartner and Steenkamp, 2001).

To adjust the possible bias caused by these issues, PISA 2012 anchored some of the variables with the anchoring vignette technique introduced by King et al. (2004) (OECD, 2014a). In this technique, along with the self-reports, students are additionally given vignettes which describe hypothetical individuals (e.g. teachers) and asked to assess their level of agreement with the statements about these hypothetical individuals with the same scale as used in student self-report items (King et al., 2004; King and Wand, 2007;

Hopkins and King, 2010; OECD, 2014a). Students' responses to the vignette assessments which reflect individual differences in interpreting scales and response styles are then used to adjust for students' self-report responses (King et al., 2004; OECD, 2014a). As vignettes were involved in two of the three forms of PISA 2012 student questionnaire, anchored variables are available for those constructed with the items in these two forms (OECD, 2014a).

With investigation of the validity of anchoring vignettes by using PISA 2012 data of all 64 participating education systems, He et al. (2017) find that students' responses to the anchoring vignettes were valid for representing individual and cultural differences. The official analyses of PISA 2012 also showed that compared with original (unanchored) variables, anchored variables tend to have higher correlations with students' performance (OECD, 2014a). Besides, the paradoxical phenomena, for example, students in high-performing education systems tended to have relatively low mathematics self-concept, were no longer found when anchored variables are used (OECD, 2014a). In consideration of the above, amongst the explanatory variables, for those having corresponding anchored indicators in Fangshan PISA 2012 China Trial database, the anchored ones instead of the original ones were employed in the MLM analysis in my research.

Regarding students' responses to the measures of their familiarity with mathematics concepts, it is considered that students may overstate what they know about the concepts due to their response tendency (OECD, 2014a). The overclaiming technique was therefore used in PISA 2012 particularly to address this issue (OECD, 2014a). This technique employs some questions asking students about their familiarity with concepts that do not exist to adjust students' response tendency to overclaim (OECD, 2014a). In my research, the adjusted variable of familiarity with mathematics concepts was employed in MLM analysis.

The explanatory variables and their descriptive statistics are shown by level 1 and level 2 respectively in Table 4.5. I will further describe them in the following.

Table 4.5 Descriptive statistics of explanatory variables

Catergories		Variables	Description	N	Min	Max	Mean (SD)	Missing (%)
Level 1								
Input		Centred_ESCS	Grand mean of ESCS	614	-2.291	2.469	0 (0.816)	0
		Centred_AGE	Grand mean of AGE	614	-0.753	0.577	0 (0.269)	0
		male	Dummy variable of gender 1=male, 0=female	614	0	1	0.528 (0.500)	0
		highersec	Dummy variable of educational level 1=Upper-secondary, 0=Lower-secondary	614	0	1	0.616 (0.487)	0
Processes	OTL	EXAPPLM	Experience with applied mathematics tasks at school	413	-2.987	3.204	0.095 (1.014)	32.74
		EXPUREM	Experience with pure mathematics tasks at school	413	-2.733	0.796	-0.068 (0.937)	32.74
		FAMCONC	Familiarity with mathematical concepts (adjusted)	413	-3.440	2.920	0.481 (1.163)	32.74
	Teaching practices	TCHBEHFA	Teacher behaviour: formative assessment	411	-2.392	2.630	0.601 (1.033)	33.06
		TCHBEHSO	Teacher behaviour: student orientation	410	-1.600	3.311	0.605 (1.209)	33.22
		TCHBEHTD	Teacher behaviour: teacher-directed instruction	410	-3.653	2.563	1.021 (1.116)	33.22
	Teaching quality	TEACHSUP	Teacher support in mathematics classes	409	-2.885	1.577	0.594 (0.861)	33.39
		ANCCOGACT	Cognitive activation (anchored)	403	-3.067	3.460	0.363 (0.828)	34.36
		DISCLIMA	Disciplinary climate	411	-2.508	1.870	0.629 (0.969)	33.06
		ANCMTSUP	Mathematics teacher support (anchored)	403	-2.749	2.638	0.725 (1.020)	34.36
		ANCCLSMAN	Classroom management (anchored)	403	-2.822	2.973	0.446 (0.969)	34.36
Non-cognitive outcome		ANCINTMAT	Mathematics interest (anchored)	194	-1.696	2.998	0.584 (0.965)	68.40
		ANCINSTMOT	Instrumental motivation for mathematics (anchored)	194	-2.217	2.531	0.388 (0.887)	68.40
		MATHEFF	Mathematics self-efficacy	404	-1.535	2.298	0.860 (1.064)	34.20
		ANCSCMAT	Mathematics self-concept (anchored)	403	-2.013	2.807	0.064 (0.796)	34.36
		ANXMAT	Mathematics anxiety	411	-2.243	2.467	0.223 (0.914)	33.06
Cognitive outcome		pv1read	Reading performance	614	308	712	530 (69)	0
Level 2								
Input		SCH_ESCS	Student mean ESCS within schools	614	-0.536	0.733	0 (0.301)	0

Note: missing is almost caused by design (see Section 4.6.2.2.2).

As displayed in Table 4.5 above, the input variables involved in multilevel analysis include students’ family socioeconomic status (ESCS), students’ age, gender (male), educational level (highersec), and student mean ESCS within schools (SCH_ESCS). Among them, male and highersec are dummy variables of gender and educational level respectively. Processes variables of interest in this research include opportunity to learn content (EXAPPLM, EXPUREM, FAMCONC), teaching practices (TCHBEHFA, TCHBEHSO, TCHBEHTD), and teaching quality (TEACHSUP, ANCCOGACT, DISCLIMA, ANCMTSUP, ANCCLSMAN). Outcomes variables contain students’ non-cognitive outcomes, that is, interest (ANCINTMAT) and instrumental motivation (ANCINSTMOT) to learn mathematics, self-beliefs (MATHEFF, ANCSCMAT), and students’ cognitive outcomes in terms of reading performance.

Except the two dummy variables, male and highersec, the variables listed in Table 4.5 are all continuous variables. ESCS has a mean=0 and SD=1 for OECD countries’ average on PISA international scale (OECD, 2014a). AGE was calculated based on students’ year of birth and month of birth. To make the intercept of regression meaningful or observable, ESCS and Age, as labelled as “Centred” in Table 4.5, were linear transformed by subtracting their respective means from their values. This approach of centring is called grand mean centring. It does not change the interpretation of explanatory variables’ effects and is commonly used in multilevel analysis (Hox, 2010). With this transformation, Centred_ESCS=0, for example, refers to the average ESCS background. For the processes variables and non-cognitive outcomes variables listed above were all scaled with the mean=0 and SD=1 for OECD countries’ average (OECD, 2014a). As I introduced in Section 4.6.2.1, cognitive outcomes in each assessment domain of PISA 2012, for example, reading, mathematics, and science, five sets of PVs were drawn from the distribution of the estimation of students’ ability, representing their performance in the domain (OECD, 2014a). PVs of each domain have a mean=500 and SD=100 for the average of OECD countries in the cycle in which the domain was assessed as the majority domain for the first time (OECD, 2014a). pv1read which is displayed in Table 4.5 is the first PV of students’ reading performance. Here I just listed pv1read as an example. However, in the following multilevel analyses all the five sets of reading PVs were involved.

4.6.2.2.2 Missing values and multiple imputation

As an international large-scale assessment, PISA usually experiences the issue of missing values due to students' non-response on or non-reach of some questions. There are also a number of missing values in Fangshan PISA 2012 China Trial database employed for MLM. As shown in the fifth column in Table 4.5, the biggest number of observed student cases, is 614 rather than 624 which was displayed in Table 4.3 (see Section 4.5.2). This is due to the removal of ten cases having missing values in ESCS. ESCS was comprised based on three other indices which reflect students' background in terms of home possessions, parents' occupations and parents' educational levels (OECD, 2014a). When missing is observed in one of these three indices, regression on the other two indices would be conducted to impute the missing value (OECD, 2014a). In Fangshan PISA 2012 China Trial database, these ten students who had missing values in ESCS had missing values in two or even three of the indices. Due to the lack of information to predict the values of ESCS, these ten student cases were excluded from analyses, the student sample size of the database therefore turns to be 614.

From Table 4.5 we also see that there are missing values in each of the non-cognitive outcomes variables and processes variables. In PISA 2012, three forms of student questionnaire were used to cover more mathematics-focused issues of policy interest, with common items linking with each other. Due to this rotation design of the student questionnaire, each student only received one of the three questionnaire forms. Hence, on each of mathematics-focused items, missing values brought by this design would theoretically account for around one third of the student sample. This is the case for the processes variables and non-cognitive outcome variables which were constructed on these items. For the anchored variables ANCINTMAT and ANCINSTMOT, the percent points of missingness were even larger, as Table 4.5 shows above, since fewer students were delivered items with anchoring vignettes. Table 4.6 shows the patterns of missingness below.

Table 4.6 Patterns of missingness

Variables																Percent (%)
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	<1
1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	34
1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	33
0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	31
0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	<1
0	0	0	1	1	1	1	1	0	1	0	0	0	0	0	0	<1
0	0	0	1	1	1	0	0	1	1	1	1	1	1	1	1	<1
1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	<1
																100%

Note: variables are (1) EXAPPLM (2) EXPUREM (3) FAMCONC (4) ANXMAT (5) DISCLIMA (6) TCHBEHFA (7) TCHBEHSO (8) TCHBEHTD (9) TEACHSUP (10) MATHEFF (11) ANCCLSMAN (12) ANCCOGACT (13) ANCMTSUP (14) ANCSCMAT (15) ANCINSTMOT (16) ANCINTMAT

In this table, “1” in the patterns represents complete data, and “0” represents missingness. Reading across the table, we see that for almost all cases there are missing values in at least three variables. For example, in the second row, missingness in variables MATHEFF, ANCINSTMOT, and ANCINTMAT accounts for 34%. Ignoring the cases with missing values would be a waste of information and also hinder the analyses to involve all these variables of interest. According to the missing mechanism I explained before, these missing values were assumed as missing at random (MAR). I used multiple imputation (Rubin, 1987) which is suitable for addressing this kind of missing data. Because the missing values are arbitrary continuous values, I imputed them by using the Monte Carlo Markov Chain (MCMC) approach (OECD, 2014a).

It is suggested that all the variables, including the dependent variable (i.e. mathematics PVs), involved in the analysis model, and additional variables that could predict the missing values to be imputed should be employed in the imputation model (Acock, 2014). However, this does not mean that the more involved the better in the case that the variables themselves used to impute missing values have a number of missing values (Acock, 2014). In my research, with the observed dataset, bivariate correlation analysis on the variables of concern was conducted to identify potential predictors of the variables having missing values. As Table 4.7 shows below, amongst processes variables, most bivariate correlations were significant, and this is the same for the bivariate correlations amongst non-cognitive outcomes variables. For most processes variables, their correlation with non-cognitive outcomes variables were non-significant or weak. Although the processes variables ANCCOGACT, ANCMTSUP, and ANCCLSMAN had medium correlation with most non-cognitive outcomes, it is notable that there were no students who had responses both on the processes variables EXAPPLM, EXPUREM, FAMCONC and on the non-cognitive outcome variables ANCINTMAT, ANCINSTMOT. Involving processes variables in imputing missing values for non-cognitive outcomes variables would be problematic and vice versa. Moving to look at the reading performance, its correlation with most processes variables and non-cognitive outcomes variables was quite weak or not significant, and therefore it was not suitable to be involved in imputation. In consideration of these, imputation of missing data was conducted for processes variables and non-cognitive outcomes variables separately. Input variables, all the processes variables, and mathematics PVs were involved in imputing missing values for processes variables. Input variables, all the non-cognitive outcomes variables, and mathematics PVs were involved in imputing missing values for non-cognitive outcomes variables. Regarding whether it is necessary to include weights in imputation, it seems still inconclusive (Goldstein, 2014; Kim et al., 2014). Besides, techniques for employing weights in imputing multilevel data are still rarely developed (Rutkowski and Zhou, 2014). Hence, in my research, weights were not involved in multiple imputation.

Table 4.7 Bivariate correlation matrix of variables

	pv1math	centred _ESCS	centred _AGE	male	highersec	SCH_ESCS	EXAPPLM	EXPUREM	FAMCONC	TCHBEHFA	TCHBEHSO	TCHBEHTD	TEACHSUP	ANCCOGACT	DISCLIMA	ANCMTSUP	ANCCLSMAN
pv1math	1.000																
centred_ESCS	0.069	1.000															
centred_AGE	-0.029	-0.028	1.000														
male	-0.015	0.052	0.031	1.000													
highersec	0.407*	0.039	0.114*	-0.148*	1.000												
SCH_ESCS	0.395*	0.336*	-0.065	-0.014	0.101*	1.000											
EXAPPLM	-0.151*	0.134*	0.024	-0.029	-0.072	0.008	1.000										
EXPUREM	-0.016	0.084	0.032	0.011	-0.032	-0.053	0.411*	1.000									
FAMCONC	0.048	0.050	-0.083	-0.002	0.071	0.026	-0.056	0.111*	1.000								
TCHBEHFA	-0.202*	0.061	0.026	0.135*	-0.179*	-0.116*	0.312*	0.268*	0.022	1.000							
TCHBEHSO	-0.329*	0.060	0.022	0.179*	-0.265*	-0.126*	0.301*	0.195*	0.050	0.724*	1.000						
TCHBEHTD	-0.203*	0.058	0.057	0.006	-0.174*	-0.109*	0.208*	0.190*	0.099	0.644*	0.522*	1.000					
TEACHSUP	-0.050	0.042	0.048	-0.121*	-0.038	-0.093	0.132	0.160*	0.045	0.448*	0.310*	0.619*	1.000				
ANCCOGACT	0.097	0.033	0.065	-0.083	0.083	-0.011	0.092	0.081	0.039	0.206*	0.120*	0.283*	0.201*	1.000			
DISCLIMA	0.053	0.037	0.119*	-0.127*	-0.016	0.020	-0.013	0.082	-0.023	0.162*	0.079	0.308*	0.329*	0.112*	1.000		
ANCMTSUP	0.090	-0.084	0.072	-0.242*	0.141*	-0.085	0.086	0.114	-0.012	0.093	-0.048	0.155*	0.198*	0.565*	0.068	1.000	
ANCCLSMAN	0.083	-0.063	0.049	-0.206*	0.073	-0.059	-0.019	-0.045	-0.051	0.065	0.002	0.170*	0.206*	0.653*	0.238*	0.638*	1.000
ANCINTMAT	0.180*	-0.100	0.146*	-0.055	0.105	-0.069	.	.	.	0.175*	0.043	0.228*	0.181*	0.456*	0.157*	0.527*	0.606*
ANCINSTMOT	0.113	-0.132	0.098	-0.173*	0.104	-0.057	.	.	.	0.096	0.012	0.173*	0.145*	0.564*	0.166*	0.677*	0.681*
MATHEFF	0.407*	0.123*	-0.031	0.149*	0.127*	0.198*	0.221*	0.013	-0.015	0.125	-0.014	0.137	0.178*	0.109	0.131	0.053	0.035
ANCSCMAT	0.289*	-0.065	0.005	-0.067	0.110*	0.032	0.042	0.049	-0.005	0.056	-0.044	0.114*	0.121*	0.578*	0.123*	0.460*	0.578*
ANXMAT	-0.223*	0.031	0.102*	-0.081	-0.001	-0.027	-0.031	-0.122	0.104	-0.049	-0.020	-0.073	-0.150*	-0.084	-0.228*	-0.051	-0.084
pv1read	0.800*	0.042	-0.005	-0.249*	0.429*	0.338*	-0.171*	0.012	0.086	-0.252*	-0.405*	-0.200*	-0.051	0.115*	0.063	0.136*	0.093

	ANCINTMAT	ANCINSTMOT	MATHEFF	ANCSCMAT	ANXMAT	pv1read
ANCINTMAT	1.000					
ANCINSTMOT	0.763*	1.000				
MATHEFF	0.300*	0.110	1.000			
ANCSCMAT	0.740*	0.697*	0.268*	1.000		
ANXMAT	-0.255*	-0.201*	-0.273*	-0.326*	1.000	
pv1read	0.086	0.098	0.306*	0.181*	-0.067	1.000

Note: For illustration, pv1math and pv1read were used in correlation calculation. Significance at 95 confidence level is marked with ”*”.

In multiple imputation, a number of datasets rather than one single database are typically generated to address the uncertainty of imputation. Usually, five imputations are considered sufficient to obtain valid results (Rubin, 1987; van Buuren et al., 1999). In the field of international large-scale assessments (ILSAs), for example, PISA (pre- PISA 2015) and TIMSS, five imputations are commonly used. To align with the five sets of PVs of students' performance, in my research, five imputed datasets were generated. The MI procedure in Stata 13 (StataCorp, 2013) was employed for multiple imputation. It is suggested to compare the completed dataset in which missing values are imputed with the observed dataset for evaluating the imputation (Acock, 2014). According to the imputation results in my research, imputed values were generally consistent with observed values. To save space, imputation results of one of the variables having missing values, that is, ANXMAT (Mathematics Anxiety), are displayed below in Figure 4.2.

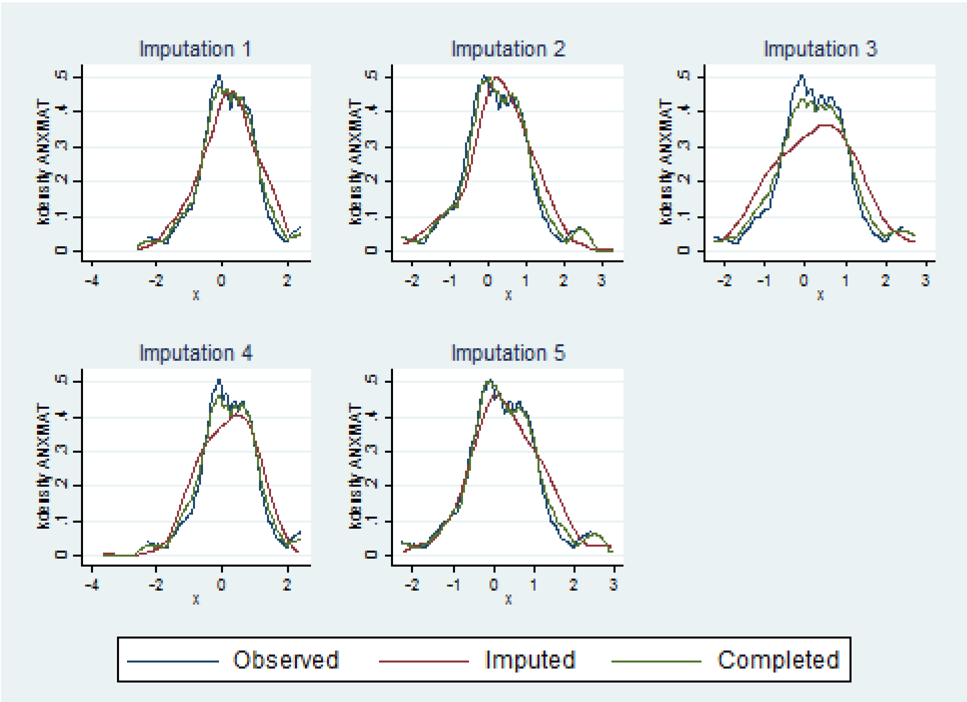


Figure 4.2 Imputation results for ANXMAT

4.6.2.2.3 Modelling strategy

Strategies for building multilevel models could be "top-down" or "bottom-up" (Hox, 2010). For the former type, the model with comprehensive variables of interest is built and analysed first, and then modified by dropping non-significant effects from the model. For the latter type, analysis starts from a simple model, models are then built on this by gradually adding variables of different levels, allowing random effects of explanatory variables or interaction between variables. In my research, the "bottom-up" strategy was adopted. Firstly, I started from the null model (also called empty model) to investigate the between-school variance in students' mathematics performance. Secondly, I added input explanatory variables to look at what kind of input predictors were associated with mathematics performance. With the control for input predictors, I further added processes and outcomes variables to identify the factors having relationships with mathematics performance.

For the multilevel model which includes both processes variables and non-cognitive outcomes variables, completed datasets of these two groups of explanatory variables were combined for analysis. In the MLM analysis, each of the five datasets was analysed together with one of the five mathematics performance PVs. With Robin's (1987) rules, the results were then combined to be the final reported results.

As the aim I mentioned earlier is to explore the factors that account for students' mathematics performance, student rather than school is the analysis unit in multilevel analysis. Hence, final student weights were used and scaled so that they sum to student sample size within schools. MLM analysis was conducted with the joint employment of the MIXED procedure and MI procedure in Stata 13 so that analysis results from the multiply imputed data could be combined efficiently without the need of manual calculation.

Besides the fixed-effect MLM, random-slope MLM modelling with the variables that were identified as having significant effects in fixed-effect models was also conducted. However, the variances of random slopes were found to be too small (near zero), indicating that the random slope for each of these variables had very small deviation from the corresponding average slope for each of these variables. Therefore, the results of the random-slope MLM model was decided not to be presented in the following chapters.

4.7 Ethical considerations

Since my research employed interviews with people who work in educational authorities and institutions, and employed PISA data, a number of ethical issues were considered in this research. I had obtained the approval (AREA 16-156) from the Faculty Research Ethics Committee (see the letter in Appendix A.13) before I started fieldwork which was then conducted following the ethical standards. In the following I will present the ethical issues that were considered in my research.

4.7.1 Informed consent

Regarding doing interviews, one of the ethical concerns is the informed consent of interviewees. During the process of recruiting interviewees, I provided potential interviewees with the electronic version of information sheets (see the sample in Appendix A.5). On the information sheets, I clearly introduced the aims of my research, and how I would use and protect the interview data to ensure they adequately understand the necessary information. Interviewees were informed that they had the entitlement to refuse to participate or to withdraw data. They were also informed that any withdrawal of data should be proposed by the interviewee within two weeks after he/she was interviewed, so that I could recruit another participant in time for collecting data. The information sheets also include my contact information, in case potential participants would like to know more about their participation. Before the start of each interview, I introduced my research and interviewees' participation again, gave a copy of the information sheet to the interviewee to read and allowed them to ask any related questions. The consent form (see the sample in Appendix A.6) was then provided to the interviewee for getting the interviewee's formal written consent of participating in my research. The written consent was obtained from all the interviewees involved in my research. In the end of my data collection, no interviewee withdrew.

4.7.2 Confidentiality and anonymisation

Another ethical consideration is about confidentiality and anonymity. For interview data, all the interview transcripts were anonymised before being imported into NVivo for data analysis. In interviews, some interviewees mentioned the names of some schools or teachers which may make their own schools or themselves identifiable. This kind of information was also anonymised in transcripts. All the quotes of interviewees' words employed in the thesis are extracts of anonymised transcripts. However, for the local educational official and the three key informants (i.e. RL, ML, SA) of the local PISA-motivated projects, even after anonymised, they might be still identified by people who are familiar with the PISA-motivated initiatives in Fangshan. This possible situation was made clear to these interviewees. Interviewees were informed that some segments of their anonymised data may be looked at by my supervisors for supervising my PhD. However, they were assured that their data would not be shared with others, and their personal details which might reveal their identities would not be disclosed either. Since the interviews focus on investigating the interviewees' perception of the impact of PISA on students' learning, rather than the interviewees themselves, I did not collect their sensitive personal data.

For PISA databases, any sensitive information which could make specific individual students or schools identifiable had been suppressed and anonymised during the process of generating databases. Therefore, the databases which I obtained from the NEEA do not include this kind of information. However, considering the confidentiality requirement of using Fangshan PISA data, I need to comply with the confidentiality statement which I signed and will not share these data with any individuals or institutions.

4.7.3 Data management and protection

Management and protection of research data is also of concern. In line with the University policy of research data management, all the interview data and the backups are stored in my own network storage area in the University, that is, M: drive. Physical materials were sorted in types and are saved in a locked place in my office. Data processing and analysis were conducted in M: Drive. Following the suggestion I obtained from the University research data management team, Fangshan PISA databases were encrypted and transferred to me by China PISA National Centre via emails, while the encryption password was separately sent to me in another way. These databases and their backups are stored in my M: Drive for data processing and analyses. All digital data files, either process ones or the up-to-date ones, were named in an orderly manner according to the conventions I specified in my research data management plan so that the original outputs of any analyses can be quickly located. After the completion of my PhD, these data are to be securely erased from M: Drive with the help of the University IT team.

4.8 Trustworthiness

Regarding trustworthiness, Lincoln and Guba (1985, p.290) consider that the basic issues are:

> How can an inquirer persuade his or her audiences (including self) that the findings of an inquiry are worth paying attention to, worth taking account of? What arguments can be mounted, what criteria invoked, what questions asked, that would be persuasive on this issue?

For judging the trustworthiness of research, Guba (1981, p.80) proposes four aspects including truth value, applicability, consistency, and neutrality. For qualitative inquiries and quantitative inquires, he suggests to use two different sets of terms appropriate to these four aspects, as shown in Table 4.8 below.

Table 4.8 Terms appropriate to the four aspects of trustworthiness

Aspect	Term for qualitative inquiries	Term for quantitative inquiries
Truth Value	Credibility	Internal Validity
Applicability	Transferability	External Validity Generalizability
Consistency	Dependability	Reliability
Neutrality	Confirmability	Objectivity

Note: this table is adjusted from the Table 1 in Guba (1981, p.80).

To increase the credibility of the findings and interpretation of the perceived impact on students' learning, I adopted the techniques of prolonged engagement and triangulation (Guba, 1981; Lincoln and Guba, 1985). In terms of prolonged engagement which aims to learn the context and build trust (Guba, 1981; Lincoln and Guba, 1985), I took time in advance to search information and learn about recently-released Fangshan educational policies and reports, the culture, history and profile of each of the six schools involved in my research. As to triangulation, I collected interview data from various groups of interviewees whose multiple perspectives could triangulate the perceptions of PISA's impact. Besides, the mixed methods design was also employed for triangulation.

As to establishing transferability, it is suggested to describe the context and time in which findings are drawn on (Lincoln and Guba, 1985). Compared with many other local areas, Fangshan is a particular case in terms of its engagement in PISA. Hence, to make transferability possible, my research provided the description of the context of Fansghan and the duration of its participation in PISA (see Section 4.5.1.1), to enable readers who are interested in transferring the findings of this research to consider the particular context of Fangshan in addition to their target context.

In terms of dependability, it is in relation to "both factors of instability and factors of phenomenal or design induced change" (Lincoln and Guba, 1985, p.299). Inter-coder reliability was conducted between myself and my co-supervisor to check the consistency and appropriateness of my coding. High agreement of our coding results has been achieved.

Confirmability refers to “the extent to which the data and interpretations of the study are grounded in events rather than the inquirer’s personal constructions” (Lincoln and Guba, 1985, p.324). To ensure this, as I mentioned in Section 4.6.1 “Thematic analysis of the interview data”, the appropriateness of candidate themes were checked against interview data to confirm that findings are grounded in the data. Moreover, the analysis and interpretation of interview data in my research accommodate various data evidence, either positive or negative, about the use of PISA data and perceptions of PISA’s impact (see Chapter 5). PIP technique (Johnson et al., 2019) was used for clearly displaying the process of synthesising findings obtained from interview data and PISA data (see Section 4.3 and Section 6.3). In addition, my writing pieces were regularly given to my supervisors for judging and discussing the appropriateness of the inferences and interpretations I made based on the data.

Regarding the quantitative strand, Fangshan PISA databases were employed in my research for secondary data analyses. The reliability and validity of PISA assessment instruments have been examined and confirmed by the OECD (OECD, 2012; 2014a; 2017). To confirm the process of data analyses transparency, a reflexive journal (see Appendix A.14) was written along with data analyses to document the decisions I made for choosing the approaches of data processing and analyses and the justification supporting those decisions.

Chapter 5 Findings

5.1 The use of PISA in the local context

With the participation in two cycles of PISA China Trials and PISA 2015, Fangshan has been exposed to PISA assessment frameworks, the approach of PISA coding (so called scoring), and received its local reports of PISA results from the national authority. The interview data show that PISA has been influential in the local educational policymaking. The various types of PISA data have been actively utilised by educational policymakers and practitioners in Fangshan to inform local educational policies and practices. To investigate and describe the impact of PISA on students' learning, it is necessary to identify how Fangshan makes use of PISA and how PISA data are interpreted and translated in the local PISA-motivated initiatives; that is, seek the answer to RQ1.

RQ1. How is PISA used by Fangshan educational policymakers and practitioners in the local context?

In this chapter, I present the findings of thematic analysis on addressing RQ1 by answering the following sub-questions.

1a. How do Fangshan educational policymakers perceive PISA?

1b. What kind of initiatives have been motivated by PISA due to these perceptions?

1c. How are these initiatives implemented at school level?

1d. What is the influence of national education contextual factors on the formation and implementation of PISA-motivated initiatives?

5.1.1 PISA is perceived as a new perspective for reviewing the quality of local education

The interview data show that local educational policymakers in Fangshan perceived PISA as an assessment with a new perspective which is different from domestic examinations. They considered that PISA allows educational policymakers to evaluate the quality of local education from an international perspective, and PISA concepts shed light on their understanding of teaching.

5.1.1.1 Evaluating the quality of local education

As an administrative district of Beijing, Fangshan's involvement in PISA relies on national participation in this assessment. As reported by one local educational official who has been responsible for the local administration of PISA in Fangshan, Fangshan was involved in PISA 2009 China Trial by chance, rather than intentionally participating in this assessment on its own initiative for evaluating its local educational quality. This official said:

> Regarding the participation in PISA 2009 (China Trial), at that time, we were working with NEEA to research on examinations as measures for evaluating regional education. During the process of research and discussion, we got to know that NEEA was conducting this international assessment programme, that is, PISA. … Finally, NEEA considered that Fangshan's data could be representative to a certain extent. In addition, due to the geographical reason, going to Fangshan to do investigation or research is more convenient (for NEEA) than going to other provinces or cities. Hence, in 2009, NEEA assigned Fangshan District as a pilot area of national educational quality evaluation framework and one of the areas for PISA China Trial. (Local educational official, turn2[6])

It seems that Fangshan educational policymakers were not quite familiar with PISA at the very beginning of Fangshan's first participation in PISA China Trial. However, after tasting this international assessment which has a different flavour from domestic high-stakes examinations such as *Zhongkao* and *Gaokao*, Fangshan got to know this assessment. The local educational official said:

> After participating in PISA 2009 (China Trial), indeed we felt that it is an international perspective reviewing the current state of Fangshan education. This is an absolutely new perspective. (Local educational official, turn3)

The local educational official, the reading project leader (RL), and the mathematics project leader (ML) explicitly stated that PISA was used for identifying the issues of students' performance, and reflecting on local

---

[6] The transcript was numbered by turn.

education with the comparison to PISA assessment framework. For example, the local educational official said:

> PISA is an internationally recognised and popular international assessment. We think it's very meaningful to review Fangshan education through this perspective. … In addition, from PISA, we found that our reading performance lagged behind the international average performance. … What is assessed in PISA is not in conflict with our national curriculum. However, it has a different entry point and perspective. Hence, after participation in PISA, some issues have been identified, in other words, PISA assessment framework brought us a new perspective. With this perspective, we reviewed our previous education and found that we did fairly well in same aspects. … and were short in some other aspects. (Local educational official, turn3)

This quote indicates Fangshan's belief in PISA for its worldwide recognition and popularity, and this perceived credibility and reliability of PISA made Fangshan local eductaional policymakers consider that it was valuable to evaluate its local educational quality with PISA. By comparing with the performance of other education systems participating in PISA, Fangshan has had understanding of its local educational quality on the international scale. As the quote shows above, from PISA 2009 China Trial, Fangshan identified its relatively low reading performance by referencing PISA "international average performance" (i.e. the average performance of OECD countries) as the benchmark.

As stated by this official, although PISA assessment framework was perceived as "not in conflict" with national curriculum, there is still some difference between these two. Because PISA has a different perspective, the official perceived that some issues in Fangshan education were identified.

In response to the issues such as weaknesses in reading identified from PISA results, Fangshan local educational policymakers, using the PISA "mirror", reflected the local education by comparing it with PISA assessment framework. The reading project leader RL stated in detail about their reflection on teaching and learning of reading after the local poor reading performance was recognised from the local results report of PISA 2009 China Trial.

> In 2009, we participated in that assessment … and found that our reading was very poor, the achievement was very poor. Actually,

> in reality, students' reading volume, reading range, and reading objects were all relatively limited, not enough. So, this is a weakness. (RL, turn11)

> We got to know these from the report of assessment analyses provided to Fangshan. We've read that. Reading range (is limited), in particular. Speaking of our poor achievement, it's mainly because of the text formats, is it? The formats that we used more were continuous texts, classical belles-lettres. While the materials employed in PISA and what is needed for future society are more practical reading, including graphs, tables, that is, non-continuous texts. (RL, turn12)

In Fangshan, or even in most areas of China, reading was not set as an independent course at secondary schools since national curriculum had no such requirement (see Section 2.2). Traditionally, high-stakes examinations in secondary education such as *Zhongkao* and *Gaokao* also did not evaluate students' reading ability, especially practical reading ability, with a particular test. Usually, reading was instructed within the Chinese course. The above two quotes indicate that the poor reading performance identified from PISA 2009 China Trial motivated Fangshan to reflect on the possible reasons causing this performance. As RL stated, it was "limited" time allocation to reading and "limited" content coverage of reading that accounted for the unsatisfactory performance. She compared the text formats which students were usually exposed to in schooling with those typically involved in PISA. Based on this, she further interpreted that the poor reading performance was mainly caused by students' lack of opportunities for developing practical skills of reading non-continuous texts.

Students' weak performance in reading non-continuous texts in PISA also impressed the mathematics project leader ML. He said:

> After participation in this assessment, from the feedback what did we found about students' weaknesses? Reading charts, reading charts. … In 2012, *Gaokao* in Beijing happened to involve a problem with a chart. Each year after *Gaokao* finishes, we can obtain all the stuff of *Gaokao* analyses result from Beijing Education Examinations Authority. Having a look at that, I was shocked that students' responses on that problem were extraordinarily poor. … PISA actually reflected this issue to some extent as well, that is, our students were weak in reading charts. …They were extraordinarily consistent. (ML, turn28-30)

In the above quote, ML reported the issue of students' mathematics performance regarding reading charts in mathematics problems. It suggests that this issue was recognised by using PISA information in conjunction with the information from domestic high-stakes examinations. With *Gaokao* and PISA, students' weakness in solving mathematics problems which include charts, a type of non-continuous texts, was cross-verified.

Gender difference favouring girls in mathematics performance is another issue perceived by local educational policymakers from Fangshan's results in PISA 2012 China Trial. ML said "The difference between boys and girls (in favour of girls), I didn't realise that before." This indicates that PISA made Fansghan local educational policymakers notice gender difference in mathematics, an issue which they did not realised before. ML further stated how the gender difference was on the overall mathematics scale as well as on mathematics subscales. He said:

> There're some differences between boys' performance and girls' performance. … Boys were weaker on the whole mathematics scale. In some dimensions boys were weaker, which made them weaker than girls on the whole scale. It's really poorer. In terms of scores, (boys performed) a few scores less. This gap we then realised. (ML, turn79)

This quote also shows that mean scores of boys and girls in PISA 2012 China Trial were used in Fangshan as measures for evaluating gender difference, based on which, local educational policymakers perceived that boys achieved lower than girls in some subareas of mathematics and the overall mathematics.

Although Fangshan's results in PISA 2012 China Trial made the local educational policymakers see the gender difference in mathematics performance for the first time, it seems that not all the issues perceived from PISA drew these policymakers' attention immediately. As ML said:

> From the *Gaokao* 2016 analyses on mathematics, **once again**, I found this issue. I didn't recognise this issue until I received PISA 2012 results. Last year after *Gaokao* analyses were completed, this issue was still revealed. Therefore I've been aware of it. We've requested teachers to pay attention to boys in class, to specifically pay attention to boys. (ML, turn77)

Regarding gender difference, this quote suggests that it is *Gaokao* that mainly motivates local actions to address this issue, since the issue was identified again from *Gaokao* results several years after PISA 2012.

5.1.1.2 Enlightening educational policymakers on teaching

With the perception that PISA is a new perspective for evaluating the quality of local education, for the local educational policymakers, PISA was considered to be meaningful for enlightening them on the understanding of teaching. For example, the local educational official said:

> From the participation of PISA 2009 (China Trial), we felt that its perspective, its framework is indeed quite helpful as a guide for improving our teaching. Set PISA 2009 as an example, its focus was reading. After our participation in PISA 2009 (China Trial), we found that what PISA highlighted in reading was not exactly the same with the reading we focused in daily teaching. In terms of the texts of reading, reading in our previous classroom teaching was mainly limited to belles-lettres, just words, or a lot of words, while PISA reading does not just involve words, it also includes tables, figures, charts, etc., that is, multiple text formats. This is the shortage in our previous teaching practices. (Local educational official, turn3)

This quote shows that from the participation in PISA 2009 China Trial, in which reading was the majority domain, Fangshan found out the limitation of text formats of reading involved in its previous teaching practices. As this official reported, compared with the text formats used in PISA reading assessment, reading non-continuous texts such as tables and charts was rarely involved. RL held the same view regarding the perceived difference between the text formats of reading in PISA and what students were exposed to in schooling. As I have presented in Section 5.1.1.1, reading was traditionally instructed in the Chinese course by using Chinese textbooks. The content included in the textbooks would affect what is instructed in teaching practices. RL pointed out this.

> The reading in PISA actually reflects the stuff that would be needed for reading in life after they enter into society. Textbooks themselves are lacking in this aspect, that is to say, textbooks mainly embody reading of literary selections, continuous texts, belles-lettres. (RL, turn21)

This quote suggests that what is highlighted in PISA reading was rarely covered in the textbooks; however, that in PISA is appreciated since it is perceived as valuable for preparing students for their future life in society. The local educational official explained that their understanding of reading underlies the shortage of reading PISA-like texts in teaching practices. He said:

> Conceptually, the concept of reading that we perceived was a bit narrow. Just belles-lettres. Speaking of reading, we mainly referred to classics, belles-lettres, modern literature and so on. Probably our reading mainly focused on literature. Yet from PISA 2009 we realised that reading actually is an instrumental ability of collecting information, rather than the reading that we referred to in our previous daily teaching. Reading in PISA is closer to practical use. …Thus, our perspective is expanded. (Local educational official, turn4)

From this quote, we see that perceiving reading as appreciating literature brings limited coverage of reading text formats in teaching practices. Yet, as indicated in this quote, reading as defined in PISA 2009 framework refreshed the local educational policymakers' understanding of the concept of reading. PISA has broadened their understanding of reading and made them realise reading's instrumental function.

PISA not only shed light on local educational policymakers' understanding of reading, but also enlightened their understanding of mathematics and science teaching in terms of the content students are exposed to. The local educational official gave a few examples in the below quote.

> For example, transforming applied problems to academic problems or subject problems, validating solutions in a context after solving a subject problem. Besides, we previously focused on the instruction in within-subject knowledge, while lacked the instruction in enquiry methods. Therefore, we should say that PISA, this perspective, has complemented our understanding of teaching. (Local educational official, turn14)

This indicates that learning from PISA, Fangshan local educational policymakers perceived that instruction of some content such as applied mathematics problems and scientific enquiries were not sufficient and should be involved more than it used to be in teaching. Although the concepts of applied mathematics problems and scientific enquiries are not new in domestic context of secondary education, in light of PISA, the local educational official perceived that they were still short in instruction. He said “Say, applied problems, we have taught this type of problems since primary education; however, it seems that there was lack of this aspect, compared with PISA”. The science project assistant SA gave an example explaining why he perceived that scientific enquiry was also not done well in teaching practices.

> The phrase “scientific enquiry” is not new in our country. Set the lesson about combustion again as an example. Previously in the scientific enquiry, a set of related materials were given for students; students were then asked to enquire into them, or to verify them. … We perceived that their previous practices in enquires were a kind of pseudo-enquiry. What is pseudo-enquiry? That is, in the lesson plans for students or for teachers, teachers have already explicitly displayed something. For instance, the chemical materials such as matches and white phosphorus are given, students are then asked to do experiments, to do “enquiry”. (SA, turn16)

In the scientific enquiry activity as introduced above by SA, all related materials are ready and given to students, there is no need for students to explore them. Students therefore have few opportunities to discover the necessary conditions for combustion to take place. Although in previous teaching practices this kind of activities were called “scientific enquiry”, SA perceived that they were not real enquiries.

With the understanding enlightened by PISA, interviewees claimed that they would take initiatives accordingly to improve teaching. For example, ML said:

> Whether from our regular tests or from PISA, students’ issues identified are actually consistent. The assessment perspective of PISA is more distinctive. It’s more distinctive. For many aspects of it which I think worthwhile to use for reference, I think those are what we’d like to learn from. Grasp those and use its concepts as a guidance to address the issues of mathematics teaching and students’ learning. (ML, turn61)

This quote simply claims again that compared with domestic examinations, PISA’s perspective is distinctive, which is perceived to be valuable to learn from. And it also shows that ML and his peers would use what they value of PISA to improve teaching and learning. However, it does not mean that the instruction content which was focused in previous teaching practices would be belittled. As stated by the local educational official, the aspects that were recognised as weak from PISA would be supplemented to the existing instructional content. The local official said that, based on the complemented understanding of teaching, “We make complements to previous stuff. It doesn’t mean that the previous one is not good. After exposure to PISA, PISA enlightened us. We just additionally develop these abilities (of students) in teaching.” In the next section, I am going to present the initiatives motivated by PISA.

5.1.2 PISA has catalysed various initiatives for improving the quality of local education

In addition to using PISA as a perspective for reviewing the quality of local education, motivated by PISA, Fangshan local educational policymakers have initiated various policies and actions for improving the quality of local education, in most of which PISA results and concepts are used as a reference. The local educational official said “It indeed provides us with an international view and international perspective for improving education. We’ve got the direction or reference to meet the requirements for future talents in international development”. The interview data show that since the participation in PISA 2009 China Trial, a number of initiatives have been launched successively by Fangshan Education Commission[7] to improve teaching and learning. The reading project “Enhancing students’ reading ability”, the mathematics project “Under PISA concepts, research and practice on unit-based instructional design (UBID) for the improvement of students’ mathematics literacy”, and the science project “Enhancing science literacy” are the main initiatives. Even at the school level, some teachers have actively taken their own initiatives to bring PISA concepts into their teaching. In the following, I am going to present the initiatives of each of the three projects motivated by PISA.

5.1.2.1 Increasing students’ exposure to reading

As presented in Section 5.1.1, from PISA 2009 China Trial, Fangshan realised students’ lack of exposure to some text types of reading, for example, practical reading which usually involves non-continuous texts such

[7] Fangshan Education Commission is a local government department responsible for local education in Fangshan District of Beijing.

as tables and graphs. It is also found that Fangshan students achieved lower than their peers in OECD countries. Motivated by these results in PISA, the reading project was initiated by Fangshan Education Commission. The local educational official said:

> From PISA, we found that our reading performance lagged behind the international average performance. Hence, we addressed the issues of reading by promoting reading. We use the concepts of PISA 2009 assessment framework as a reference to improve our reading. Through the establishment of the project “Enhancing students’ reading ability”, we’ve conducted various training and activities. (Local educational official, turn3)

This indicates that the way Fangshan seeks to improve students’ reading performance is to use reading framework of PISA 2009 as a reference. As reported by this official and RL, the reading project which was launched in July of 2011 was the first PISA-motivated project in Fangshan. RL said:

> After the results of PISA 2009 (China Trial) were available, Fangshan decided to establish several research projects. Among reading, mathematics, and science, the Education Commission specified that the reading project shall be first started. The other projects were initiated two years, three years or four years later. The reading project was launched early. On the one hand, because reading was perceived more important since solving either mathematics or science problems needs reading abilities. On the other hand, it was considered that reading was easier for teachers to get started. (RL, turn6)

This quote suggests that one reason for the early launch of the reading project is that Fangshan local policymakers perceived that reading abilities are not only crucial for students’ reading performance but also important for students’ performance in mathematics and science. The following quotes show in detail of how Fangshan developed a specific reading material for students, making reference to PISA.

> For this material, we rarely chose poems, novels, fables or stories. It mainly involves practical reading, non-continuous reading, for instance, an instruction of using a microwave, a description of medicines, and a bus route map. It focuses more on practical reading in life. (RL, turn21)

> Actually, we selected stimulus aligned with the text formats of PISA released items. First, we referenced their text formats. The design of questions (in this material) is also different from our regular Chinese tests or reading tests. We imitated PISA in terms of the design of questions as well. (RL, turn29)

As I presented in the last section, students' poor reading performance in PISA 2009 China Trial was attributed to their limited exposure to PISA-style reading texts, that is, practical reading, in schooling. The above two quotes show that Fangshan has been highlighting reading for use in real-life contexts, which is one of the main features of PISA reading. To compensate students' weakness in reading non-continuous texts, or, practical reading in real-life contexts, the reading project developed a specific reading material, in which the PISA style was "imitated" in terms of text selection and question design. This material has been further developed into a "textbook". RL said:

> Before 2014, we had got a reading material with rather fixed content. This reading material has been used by students at 7th grade and 8th grade across the whole district. Especially for students at 8th grade, it is required to be used as one of school-based textbooks. (RL, turn21)

Sch5-R, a reading teacher who was one of the reading project members stated that the development and use of the reading textbook aroused the attention of Fangshan local policymakers and even Beijing municipal policymakers, who then provided expert support to develop and edit this textbook. Besides the development of the reading textbook, to guarantee the time for practical reading, reading has become an independent subject separated from Chinese, and been added into local curriculum for 8th grade students, required by Fangshan local policy.

As stated by the local educational official and RL, Fangshan Education Commission has required all concerned schools to provide one reading lesson per week for students in 8th grade. As explained by RL, considerations of the textbook difficulty, students' cognitive level, and the priority of teaching and learning in 9th grade (i.e. preparing for *Zhongkao*) were taken in making the decision of targeting students in 8th grade. The available reading textbook and the reading course indicate that, at policy level, at least for the 8th grade students, their opportunity to learn practical reading is assured.

Besides, a final examination assessing students' performance in practical reading has been specifically developed. RL said:

> For that reading textbook, we have a local-wide unified examination, since reading has been a subject. We have a final exam which has a unified exam paper developed by us. … This is to promote the initiatives at schools. If we just send the copies to schools, we think they may not have high motivation. This is for hauling students to read. (RL, turn56)

The final reading examination which is conducted at the end of an academic year, to some extent, is high-stakes for schools, teachers, and students, since performance in it is an important part of students' academic records at schools. This quote suggests that examinations are used as a measure to drive school implementation of instruction on practical reading and drive students to read.

In addition to promoting students' exposure to practical reading and guaranteeing the time for this exposure, the reading project has also made efforts to extend the range of books recommended for students to read. RL reported that the *Recommended bibliography for middle school students* (hereinafter *Recommended bibliography*) had started being developed since 2013 and in recent years it has been further improved. Different from the reading textbook, the *Recommended bibliography* orients students in both lower-secondary and upper-secondary education. RL stated "We have provided each new student at 7thgrade (i.e. the first year in lower-secondary education) and 10thgrade (i.e. the first year in upper-secondary education) one copy since 2013". This initiative is motivated by PISA as well. The local educational official said:

> Enlightened by PISA, we now first expand the range of reading. Not only recommend literature, the type we emphasised before, we've also involved various types such as science fictions, popular science, biographies, and even economic books in our *Recommended bibliography*. (Local educational official, turn4)

From this quote we also see that various types of books are recommended in this bibliography. The same as this official, RL also stated that the recommended books are no longer limited to literature. She introduced the categories involved for the two different educational levels, as the below quote shows.

> Personally, I think its greatest feature is that the selection of books is not limited to literature. …For the part for lower-secondary school students, the categories of books which our teachers recommended include literature, comic, philosophy, law, history, religion, military, natural science, ethical, and others. … The part for upper-secondary school students additionally includes social observation. (RL, turn47)

Like RL said, the broader coverage of text types is the “greatest feature” of the *Recommended bibliography*, and the content in it is quite different from what was used to be commonly recommended to read. The local educational official said:

> This has changed the state that we mainly recommended literature for reading. In fact, it expands students’ view of reading that as long as they read, whatever the content is, reading would be helpful for their life development. (Local educational official, turn4)

This quote indicates that Fangshan has been highlighting students’ exposure to different text types of whole-book reading through the *Recommended bibliography*. It also suggests that Fangshan local educational policymakers believe in their reshaped perception of reading and they have been trying to deliver the concept of broad reading which is perceived by them to students.

Moreover, a number of reading activities, such as holding lectures at schools, have been undertaken by the reading project to increase students’ interest in whole-book reading. RL reported that they have invited experts or outstanding teachers to schools and given students lectures on reading, meanwhile, schools have been encouraged and financially supported to invite experts and organise lectures by themselves according to their own needs.

Besides these reading activities, writing is a measure perceived by local educational policymakers as helpful for students to exercise and solidify what they have absorbed from reading. RL said:

> We've got a journal called *Longxiang Xinya*. It has two versions respectively for lower-secondary school students and upper-secondary school students to submit for publication. That is, the combination of reading and writing. Without writing, reading could not solidify what is got from reading, could it? It would be hard to deepen the impression (of what has been read). Therefore we've launched this journal. (RL, turn61)

The specifically developed journal provides a platform for students to share what they have read and their thoughts gained from reading in the way of writing. Writing is perceived as an approach to promoting the quality of reading.

5.1.2.2 Integrating PISA concepts into teaching to improve mathematics and science learning

Following the launch of the reading project, the mathematics project and the science project were launched successively. As ML stated, the preparation of the mathematics project had been started before PISA 2012 China Trial result was available and this project was formally launched in January 2014. SA reported that the science project started from October 2014, the time before the test dates of PISA 2015 in which science was the focus domain. The launch times of these two projects suggest that the local performance in PISA was not the main reason motivating Fangshan to conduct specific initiatives to improve students' learning in these subjects. This is different from that of the reading project which aims to address the weakness in reading ability as identified in PISA 2009 China Trial. As captured in the below two quotes, the mathematics and science projects were catalysed by PISA due to the perception of local educational policymakers that PISA is helpful for improving local education.

> From PISA 2009 to PISA 2012, we mainly worked on improving reading. From that we found that referencing the perspective of PISA is quite helpful for improving teaching. Hence, after PISA 2012, we continued to carry out the project for improving students' mathematics literacy. ... For science, we increase experiments, increase the practice of developing students' scientific enquiry ability and pay attention to develop students' ability of learning by analogy. (Local educational official, turn13)
>
> After saw the assessment method of PISA, many of us felt that it's novel and interesting. Our domestic assessments on secondary education, for example, *Zhongkao* and *Gaokao*, honestly, for mathematics, the stuff that we assess basically has standard

> answers, unique answers, while problems in PISA would have no standard answer keys, PISA uses codes (to mark the categories of responses). … On the condition that the score is the same, different codes indicate different ways of thinking. So we thought this is rather good. Secondly, we felt that problems used in PISA were not as dull as those we use. … Every problem in PISA was rich in contextual information. And it would have diverse answer keys instead of a standard one. Therefore, after the discussion together with directors, we decided (to initiate the mathematics project). (ML, turn7)

The above two quotes indicate that Fangshan local educational policymakers appreciate the way PISA assesses students. They also suggest that the perceived success of PISA-motivated initiatives of highlighting students' exposure to reading made policymakers believe that it would also be helpful by taking the perspective of PISA to improve students' mathematics and science learning.

5.1.2.2.1 Involving teachers to improve students' mathematics and science learning

The PISA-motivated initiatives are carried out according to the characteristics of subjects. Different from the reading project, in the mathematics project and the science project, initiatives are conducted to first improve teachers' teaching, rather than target students' improvement directly. Teachers' role is perceived to be crucial in these initiatives for improving students' learning of mathematics and science. The local educational official considered that the initiatives of mathematics and science require more of teachers' dedication.

> The ways to develop the three literacies are not exactly the same. …For these two aspects, we consider that the role of teachers cannot be ignored. Its influence is stronger than that in developing reading literacy. It does not mean that reading does not need teachers to lead. In reading, in terms of the time to spend, the time that students need to spend is more than that for teachers. For these two projects, teachers shall start first. (Local educational official, turn27)

For mathematics, as reported by ML, after the participation in PISA 2012 China Trial, the members of the mathematics project conducted classroom observation in Fangshan secondary academic schools, from which they found out that some issues in learning were connected with the issues in teaching. He said:

> The issues in students' learning are consistent with the issues in teachers' teaching. Some of the issues are indeed related with teachers' teaching. Hence, we address these two strands of issues together by starting from solving the issues of teachers. (ML, turn21)

This quote shows that due to the perception that factors of teaching practices influence students' learning, the mathematics project started improving students' learning from improving teachers' teaching practices. The science project also aims to improve students' learning through the influence of teachers.

> If it was me who design a lesson and give a lesson, the lesson would not be better than that given by teachers, because teachers know about their own students. …Besides, students' learning is a gradual and cumulative process. If we would like to translate PISA concepts into practices to change students, we must do it through teachers. It's impossible that students' ability could be improved by an assessment or asking them to solve lots of PISA-style problems. That is unrealistic. We must combine some content of PISA with the content we currently have. …There's another reason we considered. Besides the specific lessons involved in our project, teachers also give lessons with regard to other content. We wanted teachers to bring our ideas and our practices to their daily teaching. (SA, turn45)

This indicates that the cumulative process of students' development is acknowledged, while teachers are considered as the best vehicle bringing PISA concepts into teaching practices and gradually influence their students' learning during this process.

To make use of teachers' role in the process of students' learning, as reported by ML and SA, some teachers from different schools in Fangshan are involved in the mathematics projects or the science project. They have participated in various initiatives such as developing materials, giving demonstration classes. While a wider range of teachers across schools get involved in these initiatives by receiving these materials and related training.

5.1.2.2.2 Improving mathematics learning through improving instructional design

PISA 2012 assessed mathematics literacy as its focus domain; therefore it also provides information on students' performance in subareas of mathematics literacy (OECD, 2013a). As defined in PISA 2012 framework, mathematical content subareas include change and relationships, space and shape, quantity, uncertainty and data (OECD, 2013a). Mathematical process subareas contain formulating situations mathematically; employing mathematical concepts, facts, procedures, and reasoning; and interpreting, applying and evaluating mathematical outcomes (OECD, 2013a). The interview data show that the aim of the PISA-motivated mathematics project was not primarily to improve students' performance in one or some specific mathematics subareas according to Fangshan's performance in PISA mathematics subareas.

> The sub-dimensions of mathematics are not independent with each other. In terms of mathematical abilities, actually, they are hierarchical, rather than at the same level……With regard to the abilities, just emphasising some subarea may not definitely bring an improvement to this subarea. We do by treating them as a whole. We did not do by targeting some specific subarea, rather, based on daily teaching we complement those not done well enough. It could not simply say that we specifically take measures targeting some specific subarea. (Local educational official, turn21)

This quote suggests that with the perception of interconnection and hierarchical relationships amongst mathematics abilities, the initiatives of the mathematics project are taken to improve students' overall mathematics performance. For those in which students used to be weak, more efforts are made in teaching.

ML stated that the initial name of the mathematics project is "Emphasising the development of students' mathematics literacy in mathematical teaching". As stated by ML, the project proposed seven categories of mathematics literacy to be emphasised in teaching, namely, mathematical concepts, mathematical methods, mathematical language, mathematical thought, mathematical communication, mathematical modelling, and mathematical computation. He reported that "at that time, honestly, the conception 'literacy' was rarely proposed, while 'quality' and 'subject essence' were popularly referred to". We see from the below quote that, however, these seven categories were proposed based on the project members' own research and understanding of mathematics literacy for students.

> Our members and I read a lot of materials. Through the research on these, we've got seven categories. Because you propose students' mathematics literacy, you should know about the categories it mainly includes. (ML, turn12)
>
> I reviewed a lot of literature … about mathematics literacy. I cannot remember whose articles I've read; however, the most impressive one for me was written by Mr Jianyue Zhang. Then I looked back of what I read before, I considered that these seven categories could almost cover all the important aspects of concern. (ML, turn14)

As stated by ML, "the most impressive" article for him is written by the person who is responsible for the editing of mathematics textbooks. These two quotes suggest that the proposal of the seven categories is motivated by PISA, but it focuses more on the content of daily teaching.

Although the proposed categories of mathematics literacy are not shaped by directly borrowing the concept of mathematics literacy and categorisation of mathematics subareas as defined in PISA, the mathematics project has consciously integrated into them some of PISA concepts, for example, highlighting applied mathematics problems. ML said:

> Why did I propose "mathematical modelling"? Because we found that previously Fangshan students had issues in this aspect, once they met applied problems. With this project, we are striving to resolve this. … Seen from PISA, it has high requirements on applying knowledge to solving problems, and rarely includes problems purely assessing mastery of knowledge. That kind of literacy in PISA is actually what I think worthwhile for learning from. (ML, turn30)

This indicates that the local proposal of mathematical literacy took into consideration students' issues in mathematics modelling as well as the perceived PISA's feature in terms of highlighting "applying knowledge to solving problems". Embedding mathematics problems in real-life contexts is another feature perceived by Fangshan local educational policymakers from PISA. Therefore, in the PISA-motivated mathematics initiatives, for example, visiting classes and giving feedback to teachers, teachers are suggested to employ real-life contexts that students are familiar with in designing problems. ML said:

> I once visited a class at a lower-secondary school. The teacher was lecturing on the cartesian coordinate system. He/she used the National Theatre as an example, say, set one chair in one row as the origin and create a coordinate system, and then mark someone's position with coordinates. … Students must feel confused. After the class, I talked to the teacher calmly, because I need to point out the issue. … I asked the teacher whether he/she had been to the National Theatre. He/she said no. I also asked him/her how many students of his/her had been there, he /she then told me that only two of them had been there. I suggested that the example he/she used was not from everyday life, and it's too far away from students. Students don't know about it at all. Why did you use it as an example? ... In students' life there are a lot of available examples, for instance, we have a flagpole in the playground, to see it, students just need to turn their heads. (ML, turn54)

As presented previously in Section 5.1.1.2, the perceived feature of PISA with regard to applied mathematics problems embedded in real-life contexts shed light on local educational policymakers' and practitioners' understanding of teaching. This quote demonstrates an example that, to improve teaching, the mathematics project has been educating teachers to employ contexts that are close to students' life.

Considering that teachers had weak awareness of designing instruction for a lesson from the view of mathematics knowledge system, the project then uses unit-based instructional design (UBID) as the means to address issues in teaching so as to improve students' learning. The name of the mathematics project was then updated accordingly. ML said:

> As our research continues, we still work on those initiatives; however, we changed the project's name. Because we've found the means, we've got the name updated. Now the name is longer. It's named with "Under PISA concepts," because our project originates from PISA, we could not forget PISA. PISA has provided us the direction of research, "research and practice on unit-based instructional design for the improvement of students' mathematical literacy". We directly focus on the classroom. (ML, turn15)

ML reported that the proposed mathematics literacy and the concepts such as employing real-life contexts are embedded in the UBID. There are two main outputs of UBID, as ML stated. The *Unit-based instructional design* is a series of materials which were cooperatively composed by mathematics teaching teams from various schools under the guidance of the mathematics project. In the materials, for each lesson of a unit in a textbook, teachers design instruction from the view of the unit or broader mathematics knowledge system. The other output *Cases of teaching practices* documented the practices of UBID which were conducted at schools and have been polished based on three rounds of lesson preparation, instruction delivery, and different people's feedback. The designs which have been piloted at schools were then promoted to mathematics teachers across the whole district through local-level research lessons or demonstration classes. Some of the edited materials have been printed out and delivered to mathematics teachers across secondary academic schools in Fangshan. ML further explained the aim of developing the UBID materials:

> Many of our teachers lack resources for preparing lessons, indeed. I think they really lack resources. So besides the aim to address the issues in our district, we'd like to apply the quite good concepts of PISA to our classroom teaching, we also want to accumulate some classroom teaching resources of rather good quality. We are proofreading the remaining part (of the materials), and will print it out afterwards. The printed copy, together with its electronic version, is available for every teacher, since it is financially supported in our district. However, we have a requirement that when you prepare a lesson, you should not directly use it without any adaptation. Because your students are specific, and the requirements of teaching and assessment have also changed. You could learn from these resources, but you cannot copy them. (ML, turn24)

This shows that through UBID the mathematics project aims to address the issues in teaching and learning, the shortage of teaching materials, and also aims to bring the appreciated PISA concepts into teaching practices. As the quote indicates, these materials are recommended to be used as a reference when teachers prepare lessons. In addition, teachers are also reminded that the use of these materials should be adapted to their own students' needs and current requirements of teaching and assessment.

Moving to the school level, amongst the six mathematics teacher participants, there are two who participated in PISA coding. It is reasonable to believe that compared with the other four mathematics teachers, these two mathematics teachers know more about PISA, since they have had direct exposure to PISA and received related training. Their perceptions of PISA would motivate them to make changes in their own teaching. As Mobai, one of these two mathematics teachers said:

> I feel that PISA is open in terms of its association with each discipline, and the various scores that could be assigned to students' responses. Response like this could receive two score points, while response like that could receive three score points. So I thought, evaluating a student should also be multidirectional. Therefore, you view the child from different perspectives. These might what PISA coding enlightened me. For the mistakes students made in class, I would analyse more. … I would think more. (Mobai, turn69)

This indicates that the approach of PISA coding that Mobai perceived inspired her on how to evaluate a student. Because of this perception, she has applied this thought to her own teaching practices for analysing students' mistakes made in solving problems.

5.1.2.2.3 Translating PISA science framework into teaching practices

The launch of the PISA-motivated science project started in October of 2014. SA explicitly stated that its aims are to translate the scientific competencies defined by PISA 2015 into science teaching practices. He said "The launch of our project aims to explore the way to realise some concepts of PISA or some abilities defined in PISA in our classes under our current curriculum system." As he stated, the project focuses on developing students' competencies, that is, "explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically", which are defined in PISA 2015 science framework. The science project members consider that the science literacy defined in PISA is internationally recognised.

> We'd like to do this exploration, because PISA concepts and some of the abilities it emphasises for developing students are internationally common, that's for sure. We'd like to look at whether it could be fused with our curriculum system. (SA, turn9)

This quote indicates that, because of this perception, the science project is committed to merging PISA science literacy into domestic curriculum. As SA reported, the project achieves its aim mainly by modifying teachers' teaching methods under the goal of developing students' competencies as defined in PISA. With the researchers' interpretation of PISA framework and researchers-teachers discussion on integrating PISA framework into specific lessons, teachers make modifications to their teaching methods to highlight developing the proposed competencies of students. SA said:

> We mainly make modifications to teaching methods, or make some changes in teaching strategy. … During the process, we help teachers link the content in our curricula with that in PISA. This is one part of what we are doing. Another part is to help teachers with modifying their teaching, including the modifications of the sequence of instructional content delivery. (SA, turn41)
>
> For example, on the same topic, whether first introducing the concepts or displaying the phenomenon or proving evidence, the order is logical. We modify it based on the goal of competency development. (SA, turn42)

Besides the modification to the sequence of instruction content, the science project, like the reading project and the mathematics project, also perceived the feature of PISA in terms of contextualisation and has brought this concept into the initiatives.

> Another aspect that we'd like teachers to adjust is the stimulus used in teaching for the consideration of contextualisation. PISA also particularly emphasises contextualisation. Contextualisation is a feature of PISA commonly perceived around the world. Therefore, with regards to the selection of contexts, what kind of contexts should be chosen, daily life contexts, pure science contexts, or scientific experimental contexts, we have this consideration. (SA, turn42)

This quote shows that in the initiatives of the science project different kinds of contexts defined in PISA 2015 science framework are concerned during the process of helping teachers with modifying their teaching methods.

Besides improving teaching methods, the science project has also been carrying out other initiatives translating PISA concepts into teaching practices. For example, improving teachers' ability of developing PISA-style items of tests.

> Moreover, we also would like teachers to be able to develop a unit of items for a test or some items by themselves by imitating PISA released items. We hope teachers to achieve some improvement in this aspect. (SA, turn42)
>
> During the process of our project's research on improving teaching, we use pre-tests and post-tests. One part of the items comes from daily teaching. In daily teaching, certainly there are some practice items. We select some of them for using in the tests. The other part of the items is developed by ourselves specifically and creatively. How do we develop these items? First, teachers draft them. We'd like to see whether teachers could draft contextualised items by **imitating PISA** items. Then, based on the draft items, we make modifications together with the teachers. … Our experts have discussion with the teachers on these items and help them to revise. (SA, turn43)

In the above quotes, the emphasis on "imitating PISA" clearly shows that during the process of developing teachers' ability of developing items, teachers are encouraged and educated to learn from PISA released items.

Table 5.1 displays the brief summary of the PISA-motivated initiatives that have been taken in the local context.

Table 5.1 Summary of the local PISA-motivated initiatives

Projects	Themes/ categories	Sub-categories	In light of PISA results/concepts	Initiatives
The reading project	Increasing students' exposure to reading	Promoting practical reading	√	• Setting the reading course • Developing the reading textbook • Introducing the reading examination
		Promoting whole-book reading		• Developing *Recommended bibliography* • Conducting reading activities • Launching a student-targeted journal
The mathematics project	Improving mathematics learning through improving instructional design	Highlighting applied mathematics	√ (partly)	• Promoting seven categories of mathematics literacy • Developing the UBID materials • Educating mathematics teachers through research lessons or demonstration classes
		Highlighting mathematical language, communication, and computation, etc.		• Promoting seven categories of mathematics literacy • Developing the UBID materials • Educating mathematics teachers through research lessons or demonstration classes
		Highlighting the employment of real-life contexts in mathematics problems	√	• Developing the UBID materials • Educating mathematics teachers through research lessons or demonstration classes
		Addressing gender difference	√	• Improving boys' mathematics learning in teaching
		Promoting to design instruction with the view of mathematics knowledge system		• Developing the UBID materials
		Addressing the shortage of teaching materials		• Developing the UBID materials
The science project	Translating PISA science framework into teaching practices	Developing students' science literacy as defined in PISA	√	• Helping teachers with making modifications to teaching methods
		Highlighting the employment of real-life contexts in science problems	√	• Helping teachers with making modifications to teaching methods
		Improving teachers' ability of developing PISA-style items of tests	√	• Developing tests following PISA style

5.1.3 Implementation of PISA-motivated initiatives varies across schools

Data of school-level interviewees show that school implementation of the local PISA-motivated initiatives varies in teaching practices across schools, depending on school resources, school-level educational practitioners' perceptions of these initiatives, and their different degree of involvement in these initiatives.

The initiatives of the reading project, such as adding a reading course into the curriculum, have been commonly implemented at school level. Although the reading course is only formally required by Fangshan local policy for students at 8^{th} grade, some schools (e.g. School 1, School 2, School 3, School 5) have set the reading course for students at other grades as well. School 2, an upper-secondary school in Fangshan, is one of the first batch of base schools getting involved in the reading project. Because of this, it has added a reading course for 10^{th} grade students. One leader of this school, SchL2, said:

> Our reading instruction is really mobilised by PISA. Because at that time Fangshan District was promoting the reading project, our school was one of the first batch of schools applying for being its base schools. We aimed to strengthen students' reading experience and expand the range of their reading in students' spare time and school-based courses. First, we have specifically added the reading course, using one class hour in the school curriculum for this course. (SchL2, turn49)
>
> To promote reading, we take measures by assuring class hours, purchasing copies of classics for students to read, and assigning teachers. We have specific teachers for teaching reading. Besides, we also have held lectures. We invited people from local institutions for giving lectures on appreciating classics. (SchL2, turn53)

These indicate that by taking the opportunity of involving in the PISA-motivated reading project, this school not only has set reading as an independent subject, but also supplies related resources for it. In addition, other activities for promoting reading, such as holding lectures, which are encouraged by the reading project, have also been conducted in the school. In contrast, in School 1 which is also a base school in the reading project and has set the reading course for 10^{th} grade students, Chinese teachers take tasks of the teaching in the reading course in addition to Chinese course. As SchL1 stated, this school had to enact like this due to teacher shortage.

In some cases, school implementation of PISA-motivated initiatives is linked with schools’ own policies, as one leader of School 3 said below.

> We had reading classes in Chinese course for students at each grade before. This year we started to separately set two reading classes at 7th grade. This reading course is separately instructed in the reading room. (SchL3, turn18)
>
> Previously the librarian played the role of our reading teacher. However, this year our school specifically issued a guidance on improving reading literacy, meanwhile, designated a full-time Chinese teacher, who is quite skilled, to specifically take charge of the reading course for (students in) 7th grade. (SchL3, turn19)

School 3 is a lower-secondary school enrolling students from 7th grade to 9th grade and is also a base school of the reading project. From these quotes we see that “reading classes in Chinese course” were commonly available for students at each grade, and the school learning time for reading is even more than required by the local policy. However, for 8th grade students, there is no specific reading course for them. The school-level policy which was issued in 2017 suggests that the perceived importance of reading has been increased at this school. Yet, this school policy targets students in 7th grade rather than those in 8th grade.

Regarding the instructional content in the reading course, as I presented before (see Section 5.1.2.1), for lower-secondary schools, especially for students in 8th grade, the reading textbook developed by the reading project is required to be instructed, as introduced by RL. SchL3 said “They are probably using a textbook called *Junior middle school reading textbook* which was developed by Fangshan. It’s a local textbook”. However, this textbook, in other words, practical reading, seems not be dominant content in reading classes. For example, SchL3 commented that “in reading classes, currently we are still focusing on reading classics”. Nevertheless, the local examination appears to have motivated schools to cover this kind of content to some extent. The interview data show that, due to the reading examination, schools indeed have been motivated to use the reading textbook and deliver instruction on PISA-style reading.

> Specific instruction in reading non-continuous texts is not systematic. We would instruct this as a theme. Students actually know that texts such as medicine instructions and graphs are non-continuous. (SchL4, turn18)

> Just take some time from the Chinese course, for example, one class is just used for teaching reading non-continuous texts, focusing on it as one theme. We do in this way a bit more. Our teachers at 8th grade develop students' reading ability in this aspect a bit more. Let me find the book for you {finding the reading textbook}. Because "Enhancing (students') reading ability" is an examination-required subject for (students) in 8th grade. (SchL4, turn19)

Although the local policy of setting reading course is originally to instruct students' reading of various formats of texts by using the reading textbook, the above two quotes suggest that instruction in practical reading was perceived to be not necessary to take the time of a whole course since students do not have a big problem in this kind of reading. Instruction in practical reading is not even in the content of the reading course, as stated by SchL4. However, this content is not skipped in the school curriculum. Students are instructed in this in one class of Chinese course for preparing the reading examination. He frankly explained that 8th grade students do it because they have an examination on it. SchL4 commented that at his school the reading course focuses on the instruction in reading classics. This is not a single case, as interview data reveal. Looking at reading instruction beyond that for grade 8 students, the main content of it widely focuses on literature reading amongst schools, regardless of schools' educational levels.

By contrast, mathematics initiatives are not mandatory as required by local policies to be implemented at school level. Teachers' direct participation in these initiatives, for example, developing and editing the UBID materials, influences the school implementation of PISA-motivated mathematics initiatives. In school 2, mathematics teacher Mobai was involved in the PISA-motivated mathematics project as one of the key members, which motivated this school to have active response to the initiatives of the local PISA-motivated mathematics project. SchL2 said:

> This project has selected some teachers from schools and involved them in the training and research on mathematics thinking. This (selected) teacher (of our school) then leads our mathematics teachers to implement the corresponding initiatives of that project. ... Due to her participation, our school itself has

> been conducting a project on developing mathematical thinking in recent three or four years. This is the whole mathematics teaching team's project. In this project, they research on developing students' mathematical thinking based on their teaching and then enact it in their teaching. (SchL2, turn12)

Through this school's own project in response to local PISA-motivated mathematics initiatives, all mathematics teachers have been involved in transforming teaching practices to highlight students' mathematical thinking and communication. Yet, this is not always the case in every school. As I presented in Section 5.1.2.2.2, UBID, the main outcome of Fangshan mathematic project motivated by PISA, has been developed for Fangshan mathematics teachers as teaching materials. Regarding its use at school level, Zhushan, a mathematics teacher in a lower-secondary school who also participated in the development of UBID, said:

> Every mathematics teacher has got one copy of it. You could read it, use it for reference, or employ it. However, that is up to individuals. For example, when I'd like to refer to it, I would have a look at it. Uh, the initial goal of ML, I think, is that after it is developed, every teacher would know how to teach once receive it. If you open it, you should feel quite easy to prepare a lesson, and there's no need to reference others. To get to know the stuff for this chapter, such as instruction requirements, it's sufficient to just read this copy. That original intention (of the UBID initiative) is good, but I think whether it is applied by teachers we are not very clear {laughs}. (Zhushan, turn34)

The laughs at the end of this quote indicates that Zhushan perceived that, in school-level real practices, the usage of the UBID materials might be different from the mathematics project's expectation.

Although Zhushan acknowledged that some content in the UBID is novel and considered "beneficial to children to master knowledge", he only referenced some content of it sometimes. He attributed the infrequent use to the format of UBID which is edited in the style of a lesson plan for teachers. For his school, mathematics teachers are only required to develop the lesson plans for students to use, while teachers themselves use the same plans with additional notes or comments on instruction. In this case, teachers would feel not quite convenient to adopt the materials of UBID in their teaching, since “the lesson plan is edited from students’ perspective” and “it is indeed not quite the same with the plan for being used by teachers”.

Schools’ own teaching mode would also affect the implementation of PISA-motivated initiatives at school level. Mathematics teacher Mobai stated that she had used the UBID for four or five years until recently the school itself launched a new teaching mode which not only features student-oriented teaching but also underlines students’ reading of mathematics textbooks.

The UBID is commonly treated by mathematics teachers as a reference. In practice, when preparing teaching materials for each class, teachers do not stick to UBID, rather, they consider that students’ needs matter more. SchL1, also a mathematics teacher, said:

> Just take a look. Students are not exactly the same. Schools are different with each other. Even for the same content, we always need modification when teaching different students, so I think others’ stuff could only be used for reference, rather than be applied directly. For our own students, some knowledge they may not understand, while students at other school may already have mastered it. So we need to add more about this knowledge. It should be impossible to be lazy {laughs}. If I use it directly in instruction, students certainly cannot accept a lot of the content. (SchL1, turn38)

This indicates that when referencing the UBID materials, students’ cognitive level and needs are taken into consideration by mathematics teachers. The local PISA-motivated mathematics project involved mathematics teachers in the development of the UBID, which might be helpful for bringing the projects’ ideas into school-level implementation, as discussed above. However, the output of teachers’ cooperation in developing the materials may not be suitable for students of every school. Mingyi who participated in this cooperative work stated that “well, the content written by some teacher may be more suitable for that school’s students. We have been discussing on this issue. … What I wrote may not be applicable for teachers from other school as well.” This is again because of teachers’ attention to their students’ needs.

For those schools which consider the difficulty of mathematics problems included in UBID materials are not suitable for their students, or they have their own materials, they may not adopt the UBID materials, as SchL1 stated below.

> Students' proficiency difference is rather big (across schools). For example, students in schools like School N would not use these problems. Those schools tend to follow the schools in Haidian District[8]. While for schools like School 2 and School 6, as I know, they have their own materials. For us, we usually select some from the UBID materials. We would not use the difficult problems. (SchL1, turn30)

In some cases, it is the ideas embedded in UBID rather than the UBID physical materials that are adopted by teachers in their teaching practices, especially for those involved in PISA-related work, for example, PISA coding, and developing UBID. As to the UBID physical materials, Mingyi said "I think I did not use it a lot". However, she also reported that her instructional design has been in line with the ideas proposed by the mathematics project.

Compared with the reading project and the mathematics project, the science project which promotes the integration of PISA science literacy into curriculum is a newly launched one. Amongst the school leader interviewees, some of them reported the school implementation of the science initiatives. For example, SchL 2 said:

> First, a group of experts came to our school and delivered training. Second, they embody (the science literacy) through classes. They mainly research how to embody the science literacy through classes or experimental classes. (SchL2, turn40)

5.1.4 The influence of national education contextual factors on PISA-motivated initiatives

In this section, I will present the findings about the influence of national education contextual factors on the formation and school implementation of the local PISA-motivated initiatives. National curriculum and high-stakes examinations are the two main national contextual factors identified from interview data.

5.1.4.1 National curriculum's role in PISA-motivated initiatives

National curriculum plays an important role in the PISA-motivated initiatives. As introduced before in Section 2.2, the national curriculum which was released in 2011 "further highlights the education of Chinese fine cultural

---

[8] Haidian District is an urban district of Beijing.

traditions” (MOE, 2012). Cultural elements were intentionally integrated into text selection for the reading textbook. RL gave an example by introducing one of the selected texts, “Quadrangle Courtyards in Beijing”, which describes the architecture of Beijing traditional courtyards, and the cultural symbols in this kind of architecture. She considered that the inclusion of cultural elements is another feature of the reading textbook which is different from Chinese textbooks.

> We added and highlighted something of culture, {pointing at the text names listed in the content table of the reading textbook} for example, this Quadrangle Courtyards in Beijing. … There was no this kind of texts in previous textbooks. (RL, turn39)

One of the aims of developing the UBID is to address the issue that “mathematics teachers put the curriculum on the shelf without reading it”, as commented by ML. Therefore, in the involvement of teachers in developing the UBID, teachers were asked to read the curriculum and retrieve the related content from the curriculum for designing the instruction of a lesson. ML also demonstrated that attention to student individual differences as required by the curriculum (see Section 2.2) has been considered in the UBID to “allow different students have different achievements” by providing students mathematics problems of various levels of difficulty.

Although the science project centres on developing students’ scientific competencies as defined in PISA through teachers’ teaching, SA also stated that “during this process, we help teachers to associate the knowledge in our curriculum with that in PISA”. Rather than simply copy PISA competencies to teaching practices, the science project had been working on merging PISA 2015 science framework with national science curriculum.

> We are merging the competency and knowledge system defined in PISA with our knowledge system and our three-dimension curriculum goals. On the basis of merging, we emphasise the abilities that are promoted by PISA, so that we have a focus in **our** classes, meanwhile, we are not losing what **our** current curriculum aims to develop students. (SA, turn10)

His emphasis on the word “our” indicates the importance of national curriculum that is perceived by the science project team. As he stated, accomplishing the teaching tasks as specified in the national curriculum is still taken seriously during the process of translating PISA concepts into practices.

5.1.4.2 High-stakes examinations’ role in PISA-motivated initiatives

The requirements of *Zhongkao* and *Gaokao* are also considered in launching local PISA-motivated initiatives. Besides developing the reading textbook and promoting adding reading as an independent course into local curriculum, the reading project moves beyond the aim of developing students’ ability in reading PISA-style texts. It has undertaken a number of reading activities promoting whole-book reading.

> We’ve also conducted various reading activities. Regarding reading activities, I think we do them on the basis of the broad concept of reading, not only for practical reading or PISA. We promote reading from the aspects of students’ future development and their soul. In fact, it also could be interpreted as the kind of reading texts required by *Zhongkao* and *Gaokao*. Our activities work more on whole-book reading. (RL, turn45)

According to interviewees, some classics have been explicitly included as potential assessment content in *Zhongkao* and *Gaokao* in recent years. In terms of the text type of these classics, they are all books. This quote indicates that initiatives are also made to improve students’ ability of reading the texts of the types required by *Zhongkao* and *Gaokao*. Based on the reading project’s informal survey on students from different schools in Beijing, RL reported that Fangshan students’ reading ability was “weak in base” due to the limited range of their reading. Hence, the object of the whole-book reading is to “first meet the requirements of *Zhongkao* and *Gaokao*”, and then to expand more. As RL stated, in the *Recommended bibliography* which includes various types of reading texts, the lists of books required by *Zhongkao* and *Gaokao* are labelled as “mandatory”.

In the mathematics PISA-motivated initiatives, efforts are also made to address the issues of students’ learning which were shown in *Zhongkao* or *Gaokao*. ML reported that “for mathematics, **whether** it’s at lower-secondary or upper-secondary, **no matter** in which run of the examination, students always meet problems in two aspects, one is computation, the other is communication”. The emphasis on “whether” and “no matter” suggests that issues in computation and communication are commonly seen in students’ performance in *Zhongkao* and *Gaokao*. He described symbolic operation and mathematical communication as “disaster areas”. He said that because of these issues, in the mathematics project, developing “mathematical computation” and ” mathematical language” are proposed.

High-stakes examinations influence school implementation of PISA-motivated initiatives as well. As I presented previously in Section 5.1.3, what students mainly read in reading classes is classics. These classics in fact are almost the famous literature required by *Zhongkao* or *Gaokao*, and of course are marked as “mandatory” in the *Recommended bibliography*. Reading instruction at schools focuses more on reading of these books, especially for the final-year (i.e. 9th grade or 12th grade) students. Sch5-R, a reading teacher and also a Chinese teacher, said that when it comes to 9th grade, the final year of lower-secondary education, reading teaching tends to focus on the classics required by *Zhongkao*.

> Actually, we usually pay more attention to the must-read books, because they are related to *Zhongkao*. After all, you have to attend that examination. If there was no (such requirement of) *Zhongkao*, (which shall be focused) would have not mattered. (Sch5-R, turn86)

This suggests that schools are guided by high-stakes examinations in terms of choosing what to read for students in instruction and time allocation for the instruction. Time in reading classes is mainly used for reading, discussion and instruction about the “must-read”. Teachers may even assign students reading tasks and arrange them to read a specific book in a specific period of time. As SchL1 said, in this way, they “feed” students the required books. Reading of the other books included in the *Recommended bibliography* is also encouraged at schools, as school-level interviewees reported; however, it is up to students’ personal interest.

For some concepts of PISA that are also considered valuable by school-level educational practitioners, for example, applying knowledge in real-life contexts, some interviewees (e.g. SchL6) suggested that those concepts probably could draw higher attention if they are incorporated into high-stakes examinations.

5.1.4.3 Parents' role in these effects

Parents are considered to play an important role in the influence of national curriculum and high-stakes examinations on school implementation of PISA-motivated initiatives, especially those promoting reading. It is perceived that family atmosphere influences whether students could have broad reading (RL, Sch 5-R). RL commented that "In terms of parents, I think they would not encourage students to read for pleasure" by further explaining as below:

> Because they wanted them to do better and achieve higher in mathematics, physics, and chemistry, which would be certainly hindered by spending time on pleasure reading. … Now *Zhongkao* and *Gaokao* have begun to assess reading of some books (of classics). You see, in book stores, parents become willing to buy the books that are required by the new curriculum, *Zhongkao*, and *Gaokao*. They buy the whole set (of those books) {laughs}. (RL, turn20)
>
> They may not read by themselves, but they are willing to spend money on that for their children {laughs}, believing that reading those books is beneficial to children and could help improve their achievement. (RL, turn69)

These two quotes suggest that for the initiatives which are in line with national curriculum and high-stakes examinations, for example, increasing classics reading, parents will encourage students to get involved in these kind of initiatives for achieving high in high-stakes examinations.

Making use of parents' desire for their children's success in high-stakes examinations, parents, to some extent, are involved in the initiatives promoting reading. As stated by RL, the reading project has suggested parents to support students to read. At school level, according to Sch5-R, reading is underlined on parents' meetings, and parents are also invited to participate in school events such as students' reporting meetings on reading.

5.1.5 The local reflection on embracing PISA

From the perspectives of policymakers and those deeply getting involved in local PISA-motivated initiatives, the limited assessment capacity which could restrain the wider and better use of PISA at the local area is reported. As reported by the local educational official, applying the PISA coding approach which gives different codes (i.e. score points) to different types of students' responses to routine examinations is considered to be "more helpful in terms of improving educational quality and data mining", as this approach can capture the detailed information about students' thoughts. Yet, he further reflected that they lack the application of this approach in routine examinations and analysing the corresponding data, since this kind of coding approach is complex. Regarding interpreting students' performance with questionnaire data, it is suggested by him that experts for helping with data interpretation are needed. He stated:

> Connecting the (cognitive test) tool with questionnaires, and investigating factors (influencing students' performance) by using the test together with questionnaires, these are worth learning from for us. However, for our academic-track education, we are quite short of expertise in this area. Regarding how to conduct comprehensive analyses by using test results and questionnaires, and then come up with solutions (to address issues), we still feel quite confused. (Local educational official, turn30)

This suggests that in the local context the expertise in analysing and interpreting PISA data is limited, which restrains local educational policymakers and practitioners from making full use of PISA data to identify factors associated with students' performance and take evidence-based initiatives to improve practices accordingly. Similar point of view is also stated by interviewees deeply involved in the local PISA-motivated initiatives, who felt the lack of professional support in the initiatives. For example, ML reflected that they lack knowledge of theories, yet are rich in knowledge of practices. Mobai, one of the key member of the PISA-motivated mathematics project, stated:

> For the PISA project led by ML, we never have had someone giving training for us. We do it by ourselves. For the upper-secondary education part (in this project), there are total five members, including four teachers and ML. The UBID was just compiled by the five of us. No big name came to approve of what is good about the way we do. (Mobai, turn61)

At school level, on the one hand, it is widely perceived that the proposed PISA concepts are welcomed (e.g. SchL2, SchL3, SchL6, Zhushan, Mobai, Mingyi, Sch5-R). They recognise the value of PISA in terms of refreshing their understandings of teaching and improving their teaching practices. On the other hand, there are voices pointing out that PISA assessment framework is not yet comprehensive, and cautions are needed when interpreting PISA results and translating PISA concepts into teaching practices. For example, although SchL3 recognised that some of PISA concepts are embraced in School 3, which have been beneficial for improving teaching at the school, he commented that:

> (Regarding PISA), in terms of comprehensive assessment for students, including comparing students' achievements, … some of the results are beyond our understanding. … We would say, its assessment indicators which are developed by the West may need to be improved, or need to be interpreted from different perspectives, although they've been rigorous. … I've seen the statistics (in the PISA results report). Some of them are consistent (with our understanding), while the others are quite different (from our understanding). Therefore, I think, with regard to **the interpretation** of PISA indicators, in the future, people may have different views due to regional differences or different time lengths of exposure to PISA concepts. (SchL3, turn43)

This indicates that for some local educational practitioners, some of PISA results are unexpected according to their knowledge about their students' learning. This makes them perceive that although PISA assessment framework is rigorous, it is not yet comprehensive in terms of taking into consideration the factors such as regional or cultural differences in developing assessment indicators. Hence, from their perspective, interpreting PISA results by also considering factors besides of statistics is suggested.

Reflection on the comprehensiveness of PISA content coverage in terms of its mathematics assessment framework is mentioned by mathematics teacher interviewees as well. Mingyi said:

> PISA highlights modelling, but abstract knowledge should not be neglected. … What PISA reveals is partial, not comprehensive. (Mingyi, turn64)
>
> I think traditions are still needed to be kept. In our country, high degree of abstraction and rigorous reasoning used to be the traditional content (in mathematical education). I think we still need to keep them. You should not completely turn to the direction focusing on mathematical modelling. (Mingyi, turn65)

The above two quotes suggest that teachers like Mingyi would not consider that PISA assessment framework could comprehensively portray students' performance. Mingyi perceives that some content that is not highlighted in PISA still needs to be valued in teaching practices, although she appreciates some of PISA concepts and has been initiatively translating them into her teaching practices.

5.2 The perceived impact of the use of PISA on mathematics learning

In this section, I will present the analysis results for addressing RQ 2.

RQ 2. What are the perceptions of Fangshan educational policymakers and practitioners of the impact of the use of PISA in educational policies and practices on mathematics learning?

As introduced before (see Section 4.1), RQ 2 includes the following three sub-questions:

2a. What are their perceptions of the impact of the use of PISA in educational policies and practices on the process of mathematics teaching and learning in schools?
2b. What are their perceptions of the impact of the use of PISA in educational policies and practices on mathematics learning outcomes?
2c. What is the perceived role of national education contextual factors in the impact on mathematics learning?

I will first present the answer to the sub-question 2a (see Section 5.2.1), followed by the answer to the sub-question 2b (see Section 5.2.2). And then I will present the findings about the role of naional education context in the impact of using PISA data in the initiatives for improving students' mathematic learning, which is to address sub-question 2c (see Section 5.2.3).

5.2.1 Perceived impact of the use of PISA on the process of mathematics teaching and learning

In this subsection I will present the perceptions of Fangshan educational policymakers and practitioners about the impact of PISA and PISA-motivated initiatives on the process of mathematics teaching and learning. "Process" of teaching and learning is a broad concept which could include a number of components (e.g. Hughes 1993, cited in Bailey, 1996; Scheerens and Bosker, 1997). As revealed from interview data, the process of mathematics teaching and learning in terms of teaching practices, teaching quality, and content of learning have been influenced.

5.2.1.1 Teachers' refreshed understanding of teaching has brought changes to teaching practices

Along with the implementation of PISA-motivated mathematics initiatives, interviewees widely perceived that, first of all, mathematics teachers' understanding of teaching had been updated, and they considered that refreshed understanding would inevitably bring changes to teaching practices. At the local level, for example, the local educational official said:

> First of all, teachers have had to some extent recognition of the concept 'mathematics literacy'. They have been gradually accepting it, and starting to pay attention to it in teaching. We are translating that concept through initiatives such as local demonstration classes. (Local educational official, turn24)

This informs us that during the process of mathematics teaching and learning, it is teachers and their teaching that are first impacted by PISA concepts. PISA concepts are widely delivered to teachers through the initiatives such as local demonstration classes which are a part of teacher education activities and are available for teachers across Fangshan. In this way, teachers in Fangshan have been influenced progressively. This perception is consistent with that of most of the school-level interviewees, especially those who have awareness of the promoted PISA concepts (e.g. SchL3, SchL6, Zhushan, Mingyi, Mobai). For example, SchL6 stated:

> In terms of mathematics, these initiatives have a coverage of not just (students of) one class but also (students in) a grade. Teachers' understanding of teaching are definitely updated. (SchL6, turn44)
>
> They would not always claim that they've learned from PISA. However, they may have accepted and internalised PISA concepts, and have gradually applied them to teaching. (SchL6, turn48)

5.2.1.1.1 Teaching has been transformed to highlight student orientation and cognitive activation

With the progressive influence on teachers' understanding of teaching, teaching practices have been gradually changed. According to the classroom observation conducted by the PISA-motivated mathematics project, ML perceived that, through the training provided by the mathematics project, "teachers have had great changes". He considered that teachers' change was "the biggest effect" brought by the mathematics project. Teaching practices transforming from adopting traditional teacher-centred approach to student-oriented approach, and highlighting cognitive activation are the changes commonly reported by interviewees.

As stated by ML, when teaching students new knowledge, instead of delivering new concepts directly to students as teachers used to do, teachers' role in the classroom has commonly changed to be a "facilitator".

> Now how do our teachers instruct in a new concept? First they design a context, and then they design stimulus, design student activities, and design problems. For teachers, we've proposed that the class is students' class, not yours. You cannot dominate it. Your role is a facilitator. Therefore, through these ways teachers guide students to come up with the new concept by students themselves. (ML, turn45)

As this quote shows, under the influence of the mathematics project, teachers design contextualised problems and classroom activities to facilitate the acquisition of knowledge. Besides the changes in the way of delivering learning content, focus of teachers also has changed. ML said:

> Previously, teachers focused on their talk, while rarely considered how students listen to them. … Now, you see, teachers' main concern is not their own feeling, but students' feeling. When designing a lesson, they used to think about how to give a wonderful lecture. … Now they think about how to design it to make the content accessible to more of their students. (ML, turn46)

This indicates that teachers have had higher awareness of students' needs, and design instruction meeting the needs of students of different competency levels. Most interviewees perceived that students are more actively engaged in class. ML described that students used to "put two arms on the desk and just listen”, while now they have turned to be “the busy ones” engaged in classroom activities such as discussion and presentation. SchL3 stated that available activities for engaging students have been “obviously increased”. In this way, teachers do not spoon-feed the instructional content to students, instead, through various classroom activities students actively practice mathematical thinking and acquire knowledge by themselves with the assistance of their teachers.

These perceptions of teachers’ changes are consistent with what mathematics teachers described about their teaching practices (e. g. Linyun, Mobai, Mingyi). For example, Linyun described her experience as below:

> Previously, when I was lecturing, if I ask a question, they just need to give a response. Now, if I’d like them to give an answer to a question, I would design a question in advance to allow them to discuss, to communicate, and then they present their thinking. Usually they can present fairly well, so I just need to do a summary. While in the previously teaching practices, usually, I introduced a piece of knowledge and then asked “right or wrong” followed by students’ answer “right”. That’s it. (Linyun, turn29)

As this quote shows, students become the dominant role in class. Along with the transformation of teaching practices, cognitive activation for developing students’ mathematics thinking has been highlighted in mathematics classes. Just like Mingyi stated below:

> For the problems which could be linked to real world, I usually prepare presentation slides. I also ask students to do that. For example, to deliver the instruction in the application of trigonometric functions, I asked students to find resources. They then gave a presentation within five minutes. That presentation was very good. They also further proposed more questions (in relation to contextual factors). (Mingyi, turn5)

Similar to Mingyi, Mobai asked her students in 10^{th} grade to work in group to prepare lesson plans and also to give accessible instruction, while she simply gave students the lesson topic in advance. We could see that students are more engaged in mathematics classes and are given more cognition-challenged tasks and problems to solve. Rather than lecturing students by directly presenting the procedures of solving a problem in details, teachers encourage students to apply their knowledge in contextualised problems, find their solutions by themselves, and also explain their solutions.

ML commented that "these changes indeed have much to do with the efforts made by our mathematics project." However, he also acknowledged that these are the "joint-effect" of the PISA-motivated mathematics project, other projects, teachers' own efforts, and the implementation of the new curriculum.

5.2.1.2 Improved teaching practices influence students' learning content

5.2.1.2.1 Highlighting applied mathematics

The changed teaching practices which was to some degree impacted by PISA also have brought the changes to students' exposure to learning content in terms of learning time and content coverage. As I presented in Section 5.1.2.2.2, PISA-motivated mathematics project proposes instruction in applied mathematics in light of their perceptions of PISA featuring applying knowledge to solving real-life problems and their perceptions of student's weakness in solving applied mathematics problems. As stated by ML, the time for students' exposure to the content of applied mathematics has been assured with the rearrangement of instructional content as required by the PISA-motivated mathematics project.

Like ML who has reported that teachers first employ a context when preparing a lesson (see Section 5.2.1.1.1), teacher interviewees' responses also reveal that teachers have been integrating this proposed concept into their teaching practices. For example, Zhushan reported that, to some extent, influenced by PISA, he "designs problems embedded in real-life contexts as much as possible" and "engages students in activities as much as possible which allow them to apply mathematics knowledge to real life". For Mobai who had adopted the steps of designing lesson plans suggested by the PISA-motivated mathematics project for several years, she employed contexts in problems to stimulate students' interest.

5.2.1.2.2 Highlighting mathematics-related reading

Along with more exposure to applied mathematics, opportunities for students to read mathematics-related texts are more available in class. ML and mathematics teacher Yunlian noted that compared with pure mathematics, applied mathematics problems usually contain loads of information and therefore have relatively high demands of students' reading ability for problem comprehension. Most teacher interviewees consistently stated that reading has been an important part in their students' mathematics learning. For example, Mingyi intentionally develops students' ability of mathematical reading by employing contextualised problems, she said:

> Like the reading comprehension in PISA. To solve the applied problems I give to students, students need to retrieve useful information from the problems, that is, reading comprehension. (Mingyi, turn16)

She further reported that her students are asked to read mathematics textbook in class or after class to develop the ability of identifying key points. Similarly, in Mobai's classes, students are also provided time to read mathematics textbooks.

Engaging students in mathematical reading is not just to help students in comprehending applied problems, but is also considered by local educational practitioners as a way to promote students' mathematics learning interest. As Zhushan reported, his students are encouraged to get involved in the activity of designing *Shouchao bao* (i.e. hand-copied paper), in which students have to search and read mathematics-related texts. He said:

> For example, … (*Shouchao bao* includes the topic about) ancient mathematics problems, while our textbooks include few of this kind of problem. So they would search (related resources) on line. Students have actually collected quite a lot, such as resources about Chickens and Rabbits, mathematics history, and stories of mathematicians. Reading those would be gradually beneficial for promoting their interest in learning mathematics. (Zhushan, turn43)

5.2.1.3 The impact on teaching is influenced by teachers' involvement in PISA and PISA-motivated mathematics initiatives

Although the delivery of some PISA concepts has wide coverage in Fangshan local area, it has different degree of impact on different teachers. The degree of teachers' involvement in the mathematics project and whether they have had direct exposure to PISA influence the impact of PISA and the PISA-motivated mathematics initiatives on their understanding of teaching and their teaching practices.

Wenting, who was involved in designing one lesson of UBID materials, considered that the concepts in UBID "certainly has influence" on her teaching; however, she also stated that "I haven't yet been able to deal with the relationship between the 'unit' and the specific lesson very well. I still have problems in this regards." This suggests that her involvement in the initiative is not sufficient for her to integrate the proposed concepts well into her own teaching.

By contrast, teachers who had deeper involvement than their peers in PISA-related work would have relatively clear perception of the impact of PISA or PISA-motivated initiatives on their teaching practices. Zhushan reported that the UBID developed by the mathematics project has made him have the awareness of preparing mathematics lessons from a broad view of the unit and the wider mathematics knowledge system. He perceived that his involvement in the mathematics project has an impact on the focus of his teaching. As reported by him, enlightened by PISA concepts, now in his teaching practices he pays more attention to bringing real-life contexts into mathematics problems to develop students' mathematical thinking and promote students' interest in learning mathematics.

For teachers who are not only key members of the local PISA-motivated projects but also have had direct exposure to PISA, for example, participation in PISA coding and attended related training on introducing PISA or PISA approaches, they have even more explicit recognition of the impact of PISA on their understanding and teaching practices. According to Mingyi, one of these teachers, the training on PISA coding was "eye-opening" for her. She further stated:

> Previously, we thought, for some kinds of responses (of students), we would not give marks. But it is not like that in PISA coding system, in which as long as pupils have displayed some thoughts, you should give them a sense of accomplishment. … I think that is pretty good. That's why I said that I'd like to update my understanding (of teaching). (Mingyi, turn63)

From these perceptions, we see that compared with the progressive impact of the PISA-motivated initiatives on teaching practices of general teachers, PISA has more direct impact on those teachers who were directly exposed to it. Since there are only a small number of these teachers, the wide scope of PISA's relatively direct impacts is questioned. Mingyi said:

> We attended the training on introducing PISA before our participation in PISA coding. It's really eye-opening. … I think I am one of the lucky ones who were involved in it. However, many teachers don't know this. … I can implement these ideas because I attended the training luckily. Yet, many teachers didn't attend it. Through our teaching team led by me, or through my lessons, some teachers might be impacted. However, the scope of that impact is limited after all. So I always say that although PISA is good stuff, few front-line teachers have exposure to it. … Without wide concern, I think it's hard to assert it's impact. (Mingyi, turn63)

Although these teachers could bring PISA's impacts to their colleagues around, as Mingyi stated above, this direct influence was perceived to be limited. SchL 6 explained the limited direct impacts of PISA on teachers from the perspective of PISA's nature.

> PISA's concepts are not necessarily known by many teachers. It is not yet popular for teachers. Because PISA is **the kind of** tests which have no stakes for students to be enrolled in a higher level of education. Hence, many teachers may not know a lot about it. They would think it does not have much to do with them. They may just know that there is a test called PISA. Regarding the degree of understanding, those who have participated in related work would attach importance to it, while those who have no this experience would not think they have to get to know what PISA is, or get any idea from it. (SchL 6, turn48)

Unlike *Zhongkao* or *Gaokao*, two of the most concerned examinations in secondary education in China, PISA as an international assessment at least is not high-stakes for students, teachers or schools. This quote suggests that if there is no external motivation, teachers would not have the initiative to actively approach to this assessment.

5.2.2 Perceived impact of the use of PISA on students' mathematics learning outcomes

Since PISA is not a high-stakes assessment for students, students themselves probably would not consciously embrace PISA concepts and prepare themselves to perform better on PISA-style problems. However, interviewees perceived that the updates of teachers' understanding of teaching would certainly bring changes to their practical teaching so as to be helpful for students' learning. Students are influenced by PISA indirectly, unconsciously, and progressively. In the following, I will present Fangshan educational policymakers' and practitioners' views of the impact on students' non-cognitive as well as cognitive outcomes of mathematics learning.

5.2.2.1 Perceived impact on students' mathematics non-cognitive learning outcomes

Interviewees at school level widely perceived that through the changes to the process of teaching and learning, students have become more positive in attitudes towards learning mathematics. In the following I will present these perceived impact on students' non-cognitive outcomes in terms of their self-beliefs and interest in mathematics which are most reported by interviewees.

5.2.2.1.1 Perceived impact on students' self-beliefs in learning mathematics

Through student-oriented teaching practices and cognitive activation in class, it is perceived that students have become more confident in presenting their understanding and thoughts of mathematics, and more confident in solving and also explaining their solutions to mathematical problems. Linyun stated that compared with students' previous performance in class, they now "dare to speak out" and "dare to question". This change is also explicitly reported by SchL2 and Mingyi.

SchL2 said "During the presentation, students could confidently write down formulas, concepts or examples on the blackboard, and give explanations, sometimes they even do not need a script". She perceived that students' enhanced confidence in mathematics "at least has a certain relationship" with the school's initiatives stimulated by the local PISA-motivated mathematics project.

Another example given by Mingyi could also well evidence the perceived impact of PISA brought to students' confidence in solving mathematics problems. Like she said, in the cases that the student presenter fails to solve some sub-question of a mathematics problem, other student would volunteer, saying "let me help you to solve this sub-question. I'll present it." Mingyi stated that "they have such self-confidence" is one of the changes brought by her changes in teaching practices due to her involvement in PISA-motivated initiatives.

5.2.2.1.2 Perceived impact on students' interest in learning mathematics

Increased interest is another aspect showing students' change in non-cognitive mathematics learning. At school 1, preparing lesson plans employs the student-oriented approach which caters for the needs of students of different ability levels to make students have a sense of accomplishment in class, as reported by SchL1. SchL1 considered that this change in teaching practices made students' learning interest towards mathematics increased. She said "Through various activities, students' interest in learning has been stimulated. They are willing to learn (mathematics)". In Mingyi's classes, PISA concepts such as designing problems with real-life contexts, developing students' mathematical thinking and communication have been integrated into her teaching practices, as she stated. In the below two quotes, she described her students' increased interest of learning mathematics brought by this kind of improved teaching practices.

> They think it is **playing** with mathematics. When a class begins, my students would say "Teacher Mingyi, shall we play with mathematics for a while?" Awww, playing with mathematics. They never ask 'what are we gonna do in class?' They say "let's play with mathematics. We'd like to discuss in this way, what do you think?" (Mingyi, turn50)
>
> They are happy to learn. Mathematics could be very boring in terms of the aspects like loads of calculations. But through this continuous development I said previously, my students become loving learning mathematics. They consider learning mathematics as playing with mathematics. They think mathematics sometimes is very magical. (Mingyi, turn52)

In these two quotes, the emphasised tone on "playing" and repeated use of this word suggests that the teacher was impressed by students' enjoyment of learning mathematics.

It is widely perceived by teacher interviewees that promoting students' non-cognitive outcomes such as self-beliefs and interest in leaning mathematics is crucial for students' mathematics performance.

5.2.2.2 Perceived impact on students' mathematics cognitive learning outcomes

The impact of PISA or PISA-motivated initiatives may not easily be quantified, as commented by some interviewees (e.g. RL, SchL6). But interviewees perceived that students' cognitive learning outcomes including their achievements in examinations and cognitive performance in class could serve as the evidence indicating the impact.

5.2.2.2.1 Perceived impact of reading initiatives on students' mathematics cognitive learning outcomes

The initiatives conducted by the PISA-motivated reading project is widely perceived to be helpful for improving students' mathematics achievements. At Fangshan local level, the trend of students' performance in PISA was considered by interviewees as an explicit index evidencing this impact.

> Intuitively, seen from PISA results, after we launched this (reading) project, reading performance had a great improvement in PISA 2012 compared with that in PISA 2009. Moreover, performance in the other two subjects also had improvement. That means reading could also facilitate students' development in other subjects. PISA-style reading highlights information collection and processing from various texts. Therefore, it is actually helpful for (learning in) other subjects as well. Comparing (the performance in) PISA 2012 with (that in) PISA 2009, the improvement is very obvious. This should be the most persuasive evidence. (Local educational official, turn8)

The quote above indicates that compared with that in PISA 2009 China Trial, students had higher mathematics performance in PISA 2012 China Trial which was conducted after the launch of the reading project. This makes this local official believe that PISA-style reading which has been promoting by the reading project has facilitated students' development in mathematics, since this kind of reading features information retrieval which is also crucial for solving mathematical problems. The "obvious" "improvement" from PISA 2009 to PISA 2012 is considered as "the most persuasive evidence" indicating the reading project' impact on students' mathematics performance.

RL interpreted students' performance trends across PISA cycles from the perspective of the progress of Fangshan PISA-motivated reading initiatives. She said:

> This reading textbook started to be used in 2011, July of 2011. In 2012 it had been used in some lower-secondary schools. Anyway, it has been adopted in lower-secondary schools since 2011. So, you see the performance in PISA 2012 China Trial, it rather has had effects. Moving to PISA 2015, you see in 2013 the reading course had been widely available in whole Fangshan District. For the performance in PISA 2015, I think this course has contribution to it. (RL, turn67)

This suggests that the progress of the implementation of reading initiatives affects their impact on students' achievements in PISA. The reading project was launched in 2011, and promoting PISA-style reading had been initiated and implemented in some schools for a period of time before the test dates of PISA 2012 China Trial. Therefore, the reading initiatives are considered helpful, to some extent, for the improvement of performance in PISA 2012 China Trial. Since 2013, the reading course has been commonly added into school curriculum, as required by Fangshan local policy. The wide implementation of the reading initiatives in schools makes interviewees perceive that it is reasonable to believe the reading project's impact on the performance in the follow-up PISA 2015. RL said "I dare not claim full credit for our reading project, but it definitely helped" the improvement of performance in PISA.

Interviewees at school level also reported that reading abilities have been important for students' mathematics performance. Regarding the impact of the reading initiatives implemented at schools on students' mathematical learning, Sch5-R said:

> The reading materials in the PISA-style reading textbook contain loads of information. They include a lot of redundant information. Hence, children must be extremely careful when they read. I think this intensive reading is helpful for developing children' good reading habits and the habits of solving problems. (Sch5-R, turn33)
>
> I think it certainly has impact on students' learning in other subjects. The analysis ability involved in reading and some reading habits such as making notes during reading is certainly helpful for analysing and solving problems of other subjects. I think it definitely has impact. (Sch5-R, turn73)

These quotes show that the use of the reading textbook which is developed aligning with PISA reading was perceived to be helpful for students to solve problems in the subjects besides of reading itself. According to the curriculum (see Section 2.2), we could assume that the "other subjects" should include mathematics. She considered that using this reading textbook develops students' habits and abilities of intensive reading, which are important in solving problems of other subjects. Even some mathematics teachers reported that promoting reading accounts for Fangshan performance in PISA to some extent. For example, Yunlian said that “probably it is because years ago Fangshan had been promoting reading literacy, and teachers’ conception has been advanced. I think those rather have some effects.” These perceptions make interviewees believe that promoting reading benefited students’ mathematics performance.

Besides the evidence from PISA, students’ performance in *Zhongkao* and *Gaokao* in recent years were also illustrated as evidence showing the reading project’s impact brought to students’ learning in mathematics and wider field. The local educational official stated “Of course in our own domestic examinations, students’ all aspects of abilities also have improved”. In China, *Zhongkao* and *Gaokao* are generally considered as the most important examinations for students in secondary education, since students usually have to reach the given scores in *Zhongkao* or *Gaokao* for being enrolled in an upper-secondary school or a university (see Section 2.3). The local official reported that “in recent years, the rates of admission to universities and upper-secondary schools are increasing”, to demonstrate the improvement of students’ performance in *Zhongkao* and *Gaokao*.

5.2.2.2.2 Perceived impact of mathematics initiatives on students’ mathematics cognitive learning outcomes

Regarding the impact of the PISA-motivated mathematics project which was launched in 2014 on students’ mathematics performance, the local educational official stated “it rather has”. He further explained this perception by using PISA again as evidence.

> Comparing 2012 with 2009, we had a great improvement. Since we have already reached a particular level, the PISA 2015 result compared with 2012 at least indicates that we have stood stably on that level. In terms of this, we could say we still have made some achievement, because you have already been at the high level. Maintaining the relatively high level at least suggests that Fangshan’s achievement is not a flash in the pan. Comparing the results of PISA 2015 with that of PISA 2012 China Trial, it is evidenced that Fangshan education is developing steadily and has maintained being on a high level. (Local educational official, turn24)

As the quote shows, this local educational official considered that the PISA-motivated mathematics project “rather has” positive impact on students’ mathematics cognitive learning outcomes on the basis of his interpretation of Fangshan students’ mathematics performance trends across PISA cycles (see the performance trends in Table 5.2 in Section 5.3.1). According to this official, Fangshan's "great improvement" from PISA 2009 China Trial to PISA 2012 China Trial made Fangshan reach higher echelons on PISA international scale, and in PISA 2015 Fangshan held that achievement. With regard to whether students' mathematics performance made improvement from PISA 2012 China Trial to PISA 2015, this official interprets it by comparing its performance to those of other PISA participants, rather than directly comparing its own scores in these two cycles.

ML also held the similar view. He also justified his perception of the PISA-motivated mathematics project's effect on students' performance by comparing Fangshan with other PISA participants. He used PISA international rankings as a reference.

> In 2015, generally, the scores (of all participating systems) were lower than those in the previous cycle. On the basis of this, I think Fangshan's mathematics performance was quite good. (ML, turn63)
>
> By compared to all participating countries, that is the international rankings, and the integrated performance of B-S-J-G (China) with Fansghan' achievement, I found that Fangshan' achievement was quite good, indeed. Through this, we could see the changes of our students. We could see from multiple perspectives that our students are changing. (ML, turn64)

In this quote, before stating his perception of students' mathematics performance in PISA, ML gave the perceived context that, in PISA 2015, the performance of a number of participating education systems decreased. He interpreted Fangshan's mathematics performance in this context. Fangshan's participation in PISA 2015 was not in the form as an adjudicated region, which means its results would not be displayed in the PISA international league tables. However, this quote indicates that Fangshan inferred its possible international place by comparing its score with those on the international league table of PISA 2015 mathematics result. Based on this, ML considered that Fangshan students' mathematics performance "keeps improving".

As I presented before that the mathematics project is motivated by PISA to address the issues in teaching and learning (see Section 5.1.2.2.2). ML considered that those issues have been "**basically** resolved" through three and half years' research and practice conducted by the mathematics project.

According to the perceptions of interviewees, the impact of the PISA-motivated initiatives on students' cognitive learning outcomes is not only evidenced by PISA, but also by domestic examinations and students' cognitive performance in classes. After more than three years of working on improving students' ability of mathematics modelling, ML perceived that the effect of the mathematics initiatives in this respect has been observed. He said:

> Through the emphasis on this, teachers have highlighted it in teaching, and students have paid attention to it as well. We do see the effect. Previously, as long as students were examined with applied problems, teachers would worry about students' performance. Now they don't worry about that. Basically, **most** students could perform quite well in responding to applied problems. (ML, turn34)

This quote indicates that under the influence of the PISA-motivated mathematics project, the emphasis on the instruction of solving applied mathematics problems has improved students' ability in this respect. The enhanced tone on "most" suggests that overall students have been no longer weak in solving this kind of mathematics problems in domestic examinations.

Communication with mathematical language is another issue focused by the mathematics project. As I presented in Section 5.1.2.2.2, mathematical language is one of the seven categories of mathematics literacy proposed by the project. Interviewees perceived that the related mathematics initiatives have had impact on students' ability of mathematical communication as evidenced by domestic examinations.

> Our teachers have been emphasising precise communication with standard mathematical language. … The issue of communication with mathematical language is most typical in solving analytic geometry problems. For example, to judge the relationship between a line and a plane, we need three conditions to come to a conclusion, while previously our students would just write down two conditions. If you ask students, sometimes they would be argumentative, asking you in turn, like 'this line is originally outside of the plane, do I still need to state it?' … Students have changes in this respect, and the changes are rather big. Our (students') average score points on that problem (in *Gaokao*) used to be really low, just over 1 (out of 14), sometimes over 2.

> Now on average our students could attain about half of the full score, that is, about 7. (ML, turn58)

This quote gives an example that with the changes in teaching practices, students' ability of applying standard mathematical language has enhanced. As ML reported, in *Gaokao*, students' overall attainment in solving the analytic geometry problem which requires this ability has had obvious improvement according to the score points awarded.

Students' improved ability of explaining with mathematical language in examinations is consistent with their cognitive performance observed by mathematics teachers and school leaders in class. For instance, Mingyi reported that she found her students now could use mathematical language in communication well after she integrated the proposed concepts such as highlighting mathematics communication into her teaching practices. With the school-level efforts stimulated by the local-level mathematics project, SchL2 perceived that students' enhanced ability of mathematical communication is one of the impressive changes students have had in class. She said "With regard to changes, first, it is communicating with mathematical language. Now students could use relatively professional terminologies." She reflected that this could not happen in previous traditional classes, in which the teacher dominated the class. She considered that this change "has at least a certain relationship" with the school's efforts on developing students' mathematical thinking and providing students more opportunities of developing mathematical communication ability.

The extent of the impact of PISA and PISA-motivated initiatives on students' cognitive learning outcomes might vary across classrooms, even in the school cases (e.g. School 2) that all teachers in a teaching team have received the proposed ideas. SchL2 stated that the school's initiative targets students at all the grades, "however, due to the difference of teachers' understanding and their implementation in classrooms, the effects observed in different grades are different."

5.2.3 National education context moderates the effect of PISA and PISA-motivated initiatives on mathematics learning

Although most of the interviewees believed that PISA's impact is closely related to the changes of the process of teaching and learning, they also perceived that PISA and the related initiatives are not the only factors that brought these perceived changes in Fangshan education. For example, SchL3 said “These changes of understanding of teaching on the one hand is influenced by national policies, on the other hand, they are definitely affected by PISA”. This indicates that factors of national context also play an important role. According to interviewees, we could identify that these factors are mainly from national curriculum and high-stakes examinations such as *Zhongkao* and Gaokao. For instance, ML stated:

> Now **all** mathematics teachers have had big changes. Through the lead of our project in addition to the implementation of the new curriculum, the changes of teachers are considerable, indeed considerable. The curriculum has been implemented in Beijing for 10 years. It started to be used in Beijing in 2007. (ML, turn48)

This indicates that besides the local PISA-motivated initiatives, the implementation of the new curriculum is also perceived to be influential for the changes of teaching practices.

The reforms of high-stakes examinations also moderate the impact of PISA and PISA-motivated initiatives. In recent reforms, domestic high-stakes examinations for students in secondary education tend to include more texts in mathematics problems, which was widely perceived and stated by interviewees. The local educational official said:

> For example, in the exam papers of Beijing *Zhongkao* and *Gaokao*, reading demands in mathematical and scientific problems have been considerably increasing. Previously, mathematics and science problems demand relatively small amount of reading, and problem stems are usually concise. If you’ve used all the information given in one problem, you are basically solving this problem in the right way. However, now problems include a lot of redundant information. They firstly assess students’ ability of information retrieval, and also assess their ability of processing information. That is, after retrieving the information, what specific methods shall be used to process it for solving the problem. Now both *Zhongkao* and *Gaokao* in Beijing are changing. (Local education official, turn9)

This quote shows the perceived reform trend of *Zhongkao* and *Gaokao* in terms of requiring the reading ability of information retrieval. Because these examinations have this trend while the local PISA-motivated reading project has been working on promoting reading, the local educational official attributed Fangshan's initiatives of promoting reading to students' improvement in *Zhongkao* and *Gaokao*. He said:

> Just because we have been working on emphasising reading. Due to the exposure to PISA, the initiatives are made one step ahead of other areas. (Local educational official, turn10)

SchL1 gave another example indicating the demands of reading ability in mathematics examinations.

> Since 2015, the Academic Achievement Test has required a lot of reading. There even was one mathematics problem including more than 200 words. At that time, this was quite sensational. Everyone immediately felt that reading is very important. (SchL1, turn69)

"More than 200 words" included in one problem in this high-stakes test clearly shows the growing importance of reading in mathematical problem solving, and aroused the attention of people in the field of mathematical education. This suggests that besides the PISA-motivated reading project, even mathematical teachers have been paying attention to developing students' reading ability in mathematical teaching. Like SchL1 said, "I think (high-stakes) examinations still act as a director. How they assess (students) determines how teachers teach." The requirements of high-stakes examinations are still the dominant factors influencing teaching at school level.

In addition, some interviewees perceived that domestic examinations also have had other trends such as highlighting the employment of real-life contexts in problems. For example, when explaining the change in his teaching which tends to employ real-life contexts, Zhushan said:

> Previously we just focused on solving mathematics problems. Students can attain a high score as long as they are good at solving problems. But, now, on the one hand, under the context of (the influence of) PISA which employs a lot of applied problems, on the other hand, examinations in Beijing also have such a trend. Actually, the change in our teaching must be caused by the direction of examinations, by the change of examinations. (Zhushan, turn56)

From this we see that Zhushan perceived that this change in his teaching was brought by the influence of PISA as well as domestic examinations, both of which feature real-life contexts. As perceived by Zhushan, although both PISA and domestic examinations are influencing his teaching, the latter plays the decisive part in the influence. This convergence between PISA and domestic examinations was considered as a factor accounting for students' improved performance in the subjects such as mathematics in domestic examinations. The local educational official stated:

> Domestic examinations have been gradually influenced by this kind of international assessments and have valued reading and information retrieval highly. They are changing. Therefore we have benefited from the exposure to PISA. (Local educational official, turn8)

5.3 Students' performance trends and factors associated with performance

In the last section, I have presented local educational policymakers' and practitioners' perceptions of PISA's impact on the process of mathematics teaching and learning, as well as on students' mathematics performance. In this section, I will present what PISA data tell about the trends of students' performance in mathematics over the years (Section 5.3.1), and the factors having effects on students' mathematics performance (Section 5.3.2).

5.3.1 Students' performance trends in PISA

As introduced previously (see Section 3.1 and Section 4.6.2.1), for each domain in PISA, the international scale was set with mean=500 and SD=100 for the average of OECD countries in the cycle when the domain was first assessed as the majority domain (OECD, 2014a). Students' ability estimates of each participating country or region are transformed to this scale. Across the three cycles that Fangshan participated, the observed trends of students' mean performance are shown in Table 5.2.

Table 5.2 Students' mean performance trends in PISA

PISA domains	Mean performance (SD)			Average 3-year trend
	2009	2012	2015	
Mathematics	536 (88)	580 (85)	566 (81)	15
Reading	472 (78)	539 (68)	534 (81)	31
Science	504 (74)	550 (76)	551 (79)	24

Seen from Table 5.2, in PISA 2009 China Trial, Fangshan students achieved relatively low, especially in reading. Moving to PISA 2012 China Trial, students' performance in all these three domains had dramatically increased. When it came to PISA 2015, performance in mathematics and reading declined slightly. As to the average 3-year trend from PISA 2009 to PISA 2015, Fangshan students' performance generally improved.

As literature suggests (see Section 3.4.1), changes in demographic composition of student population would influence the estimate of students' performance trends. Hence, I reweighted Fangshan student data of PISA 2009 China Trial and PISA 2012 China Trial to align with the demographic composition of the cohort in PISA 2015 in terms of age in month, gender, and educational level. With the reweighted databases, I adjusted Fangshan students performance. The results are shown in Table 5.3.

Table 5.3 Students' mean performance trends in PISA, after adjusting for demographic composition

PISA domains	Mean performance (SD)			Average 3-year trend
	2009 (adjusted)	2012 (adjusted)	2015	
Mathematics	542 (87)	561 (89)	566 (81)	12
Reading	481 (77)	525 (71)	534 (81)	26
Science	510 (71)	535 (80)	551 (79)	21

This table shows that, after data are reweighted according to the changes in student population composition, the average 3-year trend of mean performance in each domain seems to be a little lower than the results without the adjustment. However, compared to the observed trends, the adjusted trends clearly show consistent improvement of Fangshan students' performance over PISA cycles.

PISA defines different levels of students' proficiency in mathematics, reading, and science. Students who failed to attain level 2 (baseline) are so-called low-achievers, while students who attained level 5 or even level 6 are so-called high-achievers. For mathematics scale, the cut point for level 2 is 420.1, and the cut point for level 5 is 607.0 (OECD, 2014a). Fangshan students' mathematics performance trends in the percentages of low-achievers and high-achievers are displayed below in Figure 5.1.

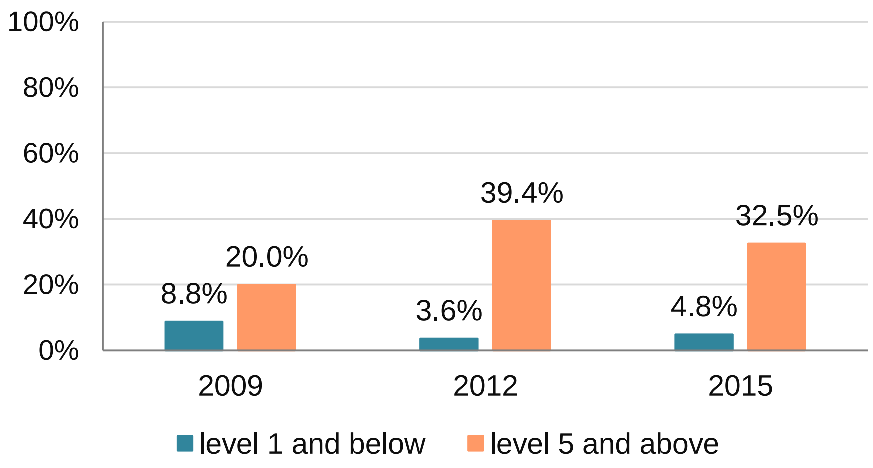


Figure 5.1 Trends in the percentages of mathematics low-achievers and high-achievers

Figure 5.1 shows that from PISA 2009 China Trial to PISA 2012 China Trial, low-achievers had about 5 per cent decrease, meanwhile, high-achievers had 20 per cent increase. In PISA 2015, the percentage of low-achievers increased a little compared to that in PISA 2012; high-achievers did not continue to increase. These observed trends seem to show a mixed picture about the changes of the percentages of students with different proficiency levels. However, as the Figure 5.2 shows below, the adjusted trends calculated with the reweighted databases clearly show that the percentage of low-achievers decreased, while the percentage of high-achievers increased over PISA cycles.

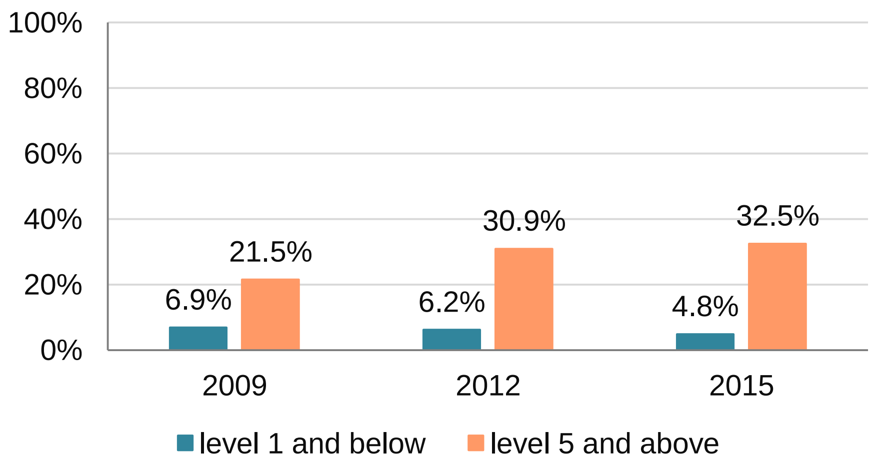


Figure 5.2 Adjusted trends in the percentages of mathematics low-achievers and high-achievers

5.3.2 Multilevel analysis of Fangshan PISA data

To explore factors that can explain variance in students' performance, MLM analyses were conducted (see Section 4.6.2.2). In this section, I will display the results of MLM analyses, by first presenting the null model result, following demonstrating the results of MLM models (Model 1 to Model 5) which include explanatory variables.

As the name indicates, in the null model, there are no explanatory variables. Only the dependent variable, here it is PVs of mathematics performance, and the identification variables of levels (i.e. student ID, school ID) and student final weights are included. Table 5.4 shows the result of the null model below.

Table 5.4 Analysis of the null model

Null model			
Fixed-effects	**Coefficient (SE)**	**t**	**P>t**
Intercept	559.126 (11.954)	46.77	<0.001
Variance components	**Estimate (SE)**	**[95% Conf. Interval]**	
Between-school variance	2898.770 (828.620)	[1654.810, 5077.844]	
Within-school variance	4610.622 (419.619)	[3841.289, 5534.036]	
ICC			
38.6%			

As shown in Table 5.4, $p<0.001$, suggesting the significance of the null model. Dividing between-school variance by the sum of variance (2898.77/(2898.77+4610.622)), the interclass correlation coefficient (ICC) which represents the proportion of variance explained by schools is obtained. According to this analysis result, 38.6% of variance of Fangshan students' mathematics performance in PISA 2012 China Trial could be accounted for by schools, indicating that Fangshan students' mathematical performance in PISA 2012 China Trial varied among schools. This magnitude of ICC suggests that it is necessary to take into consideration the multilevel structure of data in analyses. The null model was later used as the base for interpreting the results of models which include explanatory variables.

Based on the null model, I continued to examine fixed effects (i.e. fixed slopes) of variables of interest. Model 1 was built by entering the input explanatory variables to investigate the effects of students' background and school background. Centred_ESCS, centred_AGE, male (gender) and highersec (educational level) were added at level 1 (i.e. student level), and SCH_ESCS (mean ESCS within schools) was added at level 2 (i.e. school level). Model 2 and Model 3 which include processes explanatory variables or non-cognitive outcome explanatory variables were then fitted respectively.

In Model 4, students' cognitive outcome in reading (i.e. PVs of reading performance) was further added on the basis of Model 3. Following these, Model 5 involving the fixed effects of all the explanatory variables of interest was fitted. Although processes variables and non-cognitive outcomes variables were imputed separately, Model 5 was used to capture the overall picture of factors associated with students' mathematics performance. Analysis results of these models are shown in Table 5.5.

Table 5.5 Analysis of multilevel models with explanatory variables

Model	Model 1		Model 2		Model 3		Model 4	
Fixed effects	**Coefficient (SE)**	***p***	**Coefficient (SE)**	***p***	**Coefficient (SE)**	***p***	**Coefficient (SE)**	***p***
Intercept	520.036 (9.997)	<0.001	521.329 (12.184)	<0.001	518.349 (9.582)	<0.001	56.091 (22.626)	<0.05
Level 1								
centred_ESCS	-2.479 (4.400)	0.573	-.225 (4.387)	0.959	-3.692 (3.662)	0.315	-4.015 (2.319)	0.098
centred_AGE	-19.945 (10.288)	0.054	-20.954 (10.571)	0.051	-15.727 (8.765)	0.074	-5.574 (6.814)	0.416
male	7.957 (6.009)	0.187	18.096 (6.736)	<0.05	-1.276 (6.434)	0.844	26.654 (3.771)	<0.001
highersec	68.776 (12.204)	<0.001	58.400 (11.388)	<0.001	62.288 (11.597)	<0.001	23.357 (6.609)	<0.01
EXAPPLM			-6.792 (4.963)	0.196				
EXPUREM			8.026 (3.896)	0.051				
FAMCONC			3.036 (3.857)	0.443				
TCHBEHFA			-2.986 (7.832)	0.712				
TCHBEHSO			-7.300 (4.488)	0.122				
TCHBEHTD			-9.027 (5.586)	0.127				
TEACHSUP			11.070 (4.536)	<0.05				
ANCCOGACT			14.861 (7.214)	0.055				
DISCLIMA			4.486 (3.536)	0.218				
ANCMTSUP			.612 (4.622)	0.896				
ANCCLSMAN			.138 (4.088)	0.973				
ANCINTMAT					3.659 (9.387)	0.708	1.230 (5.050)	0.814
ANCINSTMOT					-9.846 (13.071)	0.484	-4.264 (6.331)	0.528
MATHEFF					16.770 (3.291)	<0.001	6.763 (2.357)	<0.01
ANCSCMAT					15.315 (5.694)	<0.01	6.821 (4.028)	0.102
ANXMAT					-10.892 (3.064)	<0.01	-6.902 (2.089)	<0.01
Reading							0.905 (0.041)	<0.001
Level 2								
SCH_ESCS	67.887 (31.056)	<0.05	62.626 (27.592)	<0.05	56.459 (25.706)	<0.05	21.358 (13.241)	0.108
Variance components	**Estimate (SE)**	**[95% Conf. Interval]**	**Estimate (SE)**	**[95% Conf. Interval]**	**Estimate (SE)**	**[95% Conf. Interval]**	**Estimate (SE)**	**[95% Conf. Interval]**
var(_cons)	1062.263 (323.814)	[584.145, 1931.718]	840.236 (258.092)	[459.856, 1535.256]	778.209 (216.187)	[451.413, 1341.587]	206.756 (67.971)	[107.374, 398.123]
var(Residual)	4493.485 (445.764)	[3686.21, 5477.553]	4057.98 (376.927)	[3373.265, 4881.681]	3673.227 (354.244)	[3008.377, 4485.009]	1577.723 (118.633)	[1355.6, 1836.243]
% of between-school variance explained	63.35		71.01		73.15		92.87	
% of within-school variance explained	2.54		11.99		20.33		65.78	

Table 5.5 Analysis of multilevel models with explanatory variables (continued)

Model	Model 5	
Fixed effects	**Coefficient (SE)**	***p***
Intercept	60.837 (25.580)	<0.05
Level 1		
centred_ESCS	-3.519 (2.583)	0.194
centred_AGE	-4.946 (7.091)	0.490
male	26.090 (4.508)	<0.001
highersec	22.521 (7.157)	<0.01
EXAPPLM	-.813 (2.994)	0.793
EXPUREM	1.057 (3.752)	0.786
FAMCONC	.045 (2.020)	0.983
TCHBEHFA	-.233 (3.431)	0.947
TCHBEHSO	1.526 (2.219)	0.497
TCHBEHTD	-3.512 (3.210)	0.283
TEACHSUP	1.227 (3.376)	0.721
ANCCOGACT	-2.474 (3.318)	0.459
DISCLIMA	-.096 (2.537)	0.970
ANCMTSUP	-.302 (3.178)	0.925
ANCCLSMAN	1.353 (3.166)	0.671
ANCINTMAT	1.013 (4.786)	0.837
ANCINSTMOT	-3.864 (6.163)	0.555
MATHEFF	7.149 (2.328)	<0.01
ANCSCMAT	7.579 (4.152)	0.076
ANXMAT	-6.830 (2.263)	<0.01
Reading	.902 (.045)	<0.001
Level 2		
SCH_ESCS	20.749 (12.927)	0.109
Variance components	**Estimate (SE)**	**[95% Conf. Interval]**
var(_cons)	191.733 (75.839)	[84.648, 434.284]
var(Residual)	1547.786 (115.491)	[1330.483, 1800.58]
% of between-school variance explained	93.39	
% of within-school variance explained	66.43	

As shown in Table 5.5 above, in Model 1, ESCS, AGE and male are not significantly associated with students' mathematics performance, while highersec (p<0.001) and SCH_ESCS (p<0.05) have significant relationships with students' mathematics performance. This suggests that for students from the same background in terms of age, gender, ESCS, and school average ESCS, those who were at upper-secondary schools tend to achieve 69 points (0.69 SD of PISA international scale) higher than lower-secondary school students. Students at the school of higher school ESCS were more likely to perform better in mathematics. According to the variance components compared with those in the null model, Model 1 accounts for 63.35% of between-school variance and 2.54% of within-school variance.

In Model 2 which additionally includes processes variables, among the process variables of students' perceived content exposure, students' experience with applied mathematics tasks at school had negative yet non-significant association with their mathematics performance (β_{EXAPPLM}=-6.792, p=0.196). In contrast, the association between students' experience with pure mathematics tasks at school was positive (β_{EXPUREM}=8.026, p=0.051). The adjusted variable of students' familiarity with mathematics concepts (FAMCONC) had no significant effect (β_{FAMCONC}=3.036, p=0.443). All the three variables about teaching practices (i.e. TCHBEHFA, TCHBEHSO, TCHBEHTD) had no significant effect on mathematics performance. Among the variables about teaching quality, teacher support had a significant and positive effect on mathematics performance (β_{TEACHSUP}=11.070, p<0.05). Cognitive activation in mathematics classes was positively associated with mathematics performance with a p value around the threshold 0.05 ($\beta_{\text{ANCCOGACT}}$=14.861, p=0.055). With processes variables including opportunity to learn specific mathematics content, teaching practices, and teaching quality controlled, students' gender (i.e. input variable male) turned to have significant effects (p<0.05), suggesting that boys performed better than girls in mathematics. Compared to Model 1, Model 2 explains 7.66% more of between-school variance and 9.45% more of within-school variance.

Moving to Model 3 which includes non-cognitive outcomes variables, both intrinsic and instrumental motivation in learning mathematics had no significant effect on students' mathematics performance ($\beta_{ANCINTMAT}$=3.659, p=0.708; $\beta_{ANCINSTMOT}$=-9.846, p<0.484). In contrast, students' mathematics self-efficacy (MATHEFF) (β_{MATHEFF}=16.770, p<0.001) and anchored mathematics self-concept (ANCSCMAT) (β_{ANCSCMAT}=15.315, p<0.01) had significant positive effects on mathematics performance, while students' mathematics anxiety (ANXMAT) (β_{ANXMAT}=-10.892, p<0.01) had significant yet negative effect on mathematics performance. Students who had higher beliefs in their abilities in solving certain mathematics tasks or their overall mathematics attributes tended to achieve higher in mathematics. For those who often found mathematics classes difficult and worried about their grades in mathematics, they tended to have lower mathematics performance. In this model, students' gender did not show significant effect on their mathematics performance. Comparing with Model 1, Model 3 explains 9.80% more of between-school variance and 17.79% more of within-school variance.

In Model 4 which further controls for students' cognitive outcome in terms of reading performance, mathematics self-efficacy and anxiety consistently had significant effects on mathematics performance; however, the effect of self-concept turned to be insignificant at 0.05 level (p=0.102). In this model, reading performance (p<0.001) was significantly associated with mathematics performance. With input variables and students' non-cognitive outcomes variables controlled, students who achieved 100 score points (1 SD of international scale) higher in reading tended to achieve 90.5 score points (0.905 SD of international scale) higher in mathematics. With reading performance added, significant gender effect (β_{male}=26.654, p<0.001) was revealed, and its effect size in this model is even larger than that identified in Model 2. It is also notable that the effect of school ESCS turned to be non-significant for mathematics performance after reading performance was added (p=0.108). With reading performance added, Model 4 explains much more variance than Model 3. Considering the changes of the effects of gender, anchored self-concept, and school ESCS brought by additionally involving reading performance as an explanatory variable, the interaction effects of these three variables with reading performance were examined. However, none of them were significant, and therefore not included in the final analyses results as displayed in Table 5.5.

In Model 5, consistent with the results shown in Model 4, input variable male (p<0.001) and highersec (p<0.01), outcomes variables MATHEFF (p<0.01), ANXMAT (p<0.01), and reading performance (p<0.001) were significantly associated with students' mathematic performance. However, none of the processes variables had significant effect on mathematics performance. To investigate possible reasons that influence the effects of processes variables on mathematics performance, multicollinearity diagnostics was conducted for all the explanatory variables involved in Model 5. The variance inflation factors (VIF) for these variables in each of the five datasets range from 1.07 to 3.61, indicating that there is no multicollinearity among these variables. This comprehensive model explains 93.39% of between-school variance and 66.43% of within-school variance.

5.4 Summary

This chapter has presented the findings of qualitative data analysis of how PISA has been used by Fangshan educational policymakers and practitioners in the local context (see Section 5.1), and their perceptions about the impact of the use of PISA in educational policies and practices on mathematics learning (see Section 5.2). It then presented the results of quantitative analyses of Fangshan PISA data, displaying Fangshan students' mathematics performance trends in PISA over time, and the factors accounting for their performance (see Section 5.3).

As revealed by interview data, Fangshan has reviewed its local educational quality using the international scale provided by PISA, and have gained insights from the international assessment framework of PISA to reflect on and improve its local education. Motivated by PISA, Fangshan has successively launched a number of policy initiatives for improving the quality of local education, with the three projects respectively focusing on reading, mathematics, and science as the main actions. Nevertheless, the motives of initiating these three projects are not the same. The reading project was the response to the local performance in PISA 2009 China Trial, while the following mathematics project and science project were established as proactive measures of local education policymakers to gain insights from PISA concepts to inform local education practices. The targets of the three projects are also different. The reading project, which focuses on increasing students' exposure to reading, directly target students, through initiatives such as making the specific reading textbook and reading course available for students, and encouraging broad reading. By contrast, the mathematics project and the science project aim to improve students' learning through improving teaching practices, and therefore greatly rely on involving teachers in shaping and implementing the relevant initiatives. The initiatives of the three projects have incorporated some concepts of PISA. However, they do not restrain their focus contents to those featured in PISA (see Table 5.1). Rather, some other considerations are included in shaping these policy initiatives to meet the domestic needs for developing students in this local area. When it comes to putting these policy initiatives into practice at schools, they are further adapted and recontextualised to fit the specific school settings during the process of school implementation which involve various school-level factors and even the influence of parents.

According to the perceptions of interviewees, students' mathematics learning has been gradually benefited from the PISA-motivated reading and mathematics initiatives, since the process of mathematics teaching and learning has been improved, especially for the teachers who have deep involvement in the mathematics project. It is considered that the extent to which PISA or the local PISA-motivated initiatives have impact on students' mathematics learning is hard to quantify. However, evidence from the changes of students' performance in PISA and domestic sources over time was used by interviewees to support their perceptions that the reading initiatives which promote students' exposure to reading, and the mathematics initiatives which improve mathematics teaching practices have contributed to the improvement of students' mathematics learning outcomes. Yet, the impact of PISA and PISA-motivated initiatives on students' learning in the local area is mediated by the convergence and divergence between the factors in national education context, such as reforms of national assessment and curriculum, and the concepts of PISA promoted by the local initaitives.

The quantitative data analysis results suggest a mixed picture of the observed trends of students' performance over the three cycles Fangshan has participated in. However, the adjusted trends accounting for the changes in demographic composition of student population reveal that Fangshan students' performance has continuously improved. Through multilevel analysis, the factors regarding teaching and learning process and learning outcomes, which have significant effects on students' mathematics performance, have been identified. In the next chapter, I will further discuss the key findings from both of qualitative and quantitative data strands and relate these to the existing literature.

Chapter 6 Discussion

My research which focuses on a Chinese local context, that is, Fangshan District of Beijing, analysed the impact of PISA on students' learning. Mixed methods were used. For the qualitative strand, 16 interview participants were recruited. Face-to-face interviews were conducted individually with these interviewees to investigate the use of PISA in the local education context, school enactment of local PISA-motivated initiatives, local educational policymakers' and practitioners' perceptions of PISA's impact on the process of teaching and learning, and on student learning outcomes. For the quantitative strand, Fangshan PISA data were employed to identify the trends of Fangshan students' performance in PISA across PISA cycles, and explore factors associated with student learning outcomes as measured by PISA. Quantitative data analyses results are used to triangulate and expand the understanding of qualitative findings in terms of interviewees' perceived development of student learning outcomes brought by PISA's impact, and their perceived factors contributing to the development.

This Discussion chapter consists of six sections. The first three sections discuss the key findings of my research with the lenses of the conceptual framework (see Section 3.6). Following these sections, Section 6.4 highlights the contributions of this research to the knowledge, methodology and theories. Section 6.5 discusses the limitations of this research. And then, Section 6.6 summarises the key points discussed in this chapter.

6.1 The use of PISA in the local context

To address the first research question, this section discusses the use of PISA data such as the local results in PISA and PISA assessment framework in Fangshan local education context. The findings drawn from the interview data clearly show that PISA has been actively used as an important tool informing the educational policies and practices in the focused local area, that is, Fangshan District of Beijing. According to literature, despite that PISA has its limitations (see Section 3.5), policymakers generally perceive that PISA has the capacity of accurately describing the performance of education systems (Carvalho, 2014). My research also found that for the educational policymakers in the local area, PISA is considered as an international benchmark which has gained worldwide recognition, and could reveal the strengths and weaknesses in the local educational quality. In addition, they appreciate the assessment framework of PISA. These are consistent with theories about washback effect that participants' perception of the assessment influences the occurrence of washback effect (Hughes, 1993, cited in Bailey, 1996). Believing the claim made by the OECD that PISA aims to evaluate "how well the students master key subjects in order to be prepared for real-life situations in the adult world" (OECD, 2019c), the local educational policymakers perceived that the abilities assessed in PISA, for example, practical reading, applied mathematics problem solving, and scientific enquiry, are important for equipping students for their future life in international development. The perceived credibility of PISA measures and the appreciation of PISA assessment framework pave the way for actively using and translating PISA data in the local educational policies and practices.

6.1.1 Using PISA as a new perspective to review local educational quality and teaching practices

According to Breakspear (2012), at national level, PISA is considered as an important indicator of education system performance in many PISA-participating education systems. Similarly, the local educational policymakers in Fangshan also see and use PISA as an important measure for evaluating the local educational performance. They considered that PISA has an assessment framework that does not conflict with but has a "different perspective" with the extant schooling system. As stated by the OECD (2019c), although PISA is not based on a specific curriculum, it employs "content that can be found in curricula across the world". This explains the local educational policymakers' perception of PISA as relevant to national curriculum and domestic assessments in the extant schooling system. It also indicates that in terms of focused knowledge and skills, to some extent, PISA assessment framework is different from national curriculum and domestic assessments. With the different perspective of PISA, the local educational policymakers recognised some issues in teaching and learning that otherwise would not be captured by domestic assessment resources. For example, from PISA 2009 China Trial, Fangshan identified students' weakness in reading ability, especially practical reading, one of the focused abilities in PISA yet not typically assessed by domestic examinations for secondary education. As I introduced before (see Chapter 2), not only in

Fangshan, but also in wider area in mainland China, reading, which is not specified in national curriculum as a subject, was not traditionally set as an independent course at secondary schools. High-stakes examinations in secondary education, *Zhongkao* and *Gaokao* in particular, did not include a test specifically assessing students' ability in this regard either. In this context, PISA results reveal this issue which previously was not realised and paid attention to by local educational policymakers and practitioners.

Although in PISA 2009 China Trial, as perceived by local-level interviewees (e.g. the local educational official, RL), Fangshan had relatively poor performance compared with the OECD average, especially in reading, my research findings do not show that Fangshan experienced "PISA shock". It seems that Fangshan's performance in its first participation in PISA did not led to scandalisation (Steiner-Khamsi, 2003), the extreme type of policy responses in education systems such as Germany which did poorly against high expectations (e.g. Ertl, 2006; Grek, 2009; Neumann et al., 2010; Kelly and Kotthoff, 2017; Niemann et al., 2017). As mentioned in the Literature Review (see Section 3.2.1), to some extent, the intensity of policy responses to PISA is in relation to the gap between the public perception of education performance and that actually received from the assessment results (Martens and Niemann, 2010). As introduced earlier (see Section 4.5.1.1), Fangshan which is located in a relatively rural area in Beijing, historically did not perform very well in education compared with peer districts in this capital city (Tao and Wang, 2014). It is reasonable to assume that the local performance in PISA 2009 China trial, the cycle that Fangshan was involved in PISA for the first time, might not be quite out of the expectation of the local educational policymakers. Perhaps because of this, hot debates on "PISA shock" were not raised in the local educational discourse.

In many cases, the different perspective of PISA is used by the local educational policymakers as an external and additional measure complementing domestic assessment resources, rather than the only single measure for reviewing the quality of local education. Although the local educational policymakers recognised PISA's credibility as an international benchmark, national assessment system still has its important position in evaluating local educational quality. For example, the local PISA results about students' weakness in solving mathematics problems which underlie non-continuous texts (e.g. tables, charts, etc.) was cross-verified by local educational policymakers with domestic high-stakes examinations. Another example is about gender difference in mathematics performance. Students' mathematics performance in PISA 2012 China Trial reversed the local educational policymakers' stereotype of gender difference that boys would not fall behind girls in mathematics on average. This evidence obtained from PISA results was also further cross-verified with students' performance in *Gaokao* mathematics examination in recent years (e.g. 2016), before convincing ML who then suggested that, rather than girls, boys are now the disadvantaged group in learning mathematics, and need to be paid attention to in mathematics teaching.

The specific perspective of PISA also allows the local educational policymakers to further reflect on the teaching practices in the local area, and the possible factors that account for students' performance in PISA and beyond. By comparing the content of PISA assessment framework with that focused in the local teaching practices, the local educational policymakers have reflected on their conception of reading, mathematics, and science in schooling. On this basis, they came up to the conclusion that students in the local area had not sufficient or appropriate exposure to some of the content that is featured in PISA, that is, practical reading, applied mathematics problems, and scientific enquiry methods. As the findings show (see Section 5.1.1.2), PISA has therefore expanded their understanding of what is reading, mathematics, and school science, and complemented their understanding of what and how to teach in these subjects. Brochu (2014) suggests that in the use of PISA, PISA indicators should be integrated into policies and practices, rather than replace extant ones in domestic context. Consistent with this point of view, the local educational policymakers did not belittle the extant instructional content and teaching practices; the concepts learning from PISA are considered as a supplement worthy of adding into teaching practices for developing students' abilities.

Based on the use of PISA in terms of reviewing the local educational quality and teaching practices, PISA has mobilised a number of initiatives for improving teaching and learning in this local area. In the following subsection, I will discuss how PISA is used in this respect.

6.1.2 Using PISA as an impetus to improve local educational quality

The use of PISA in the local policymaking suggests that PISA has direct impact on the local educational policy in terms of motivating and inspiring initiatives of improving local educational quality. The findings clearly show that in Fangshan, mobilised by PISA, educational policymakers in this local area have taken a number of actions, among which the three projects respectively focusing on reading, mathematics, and science are the main initiatives. Moreover, in shaping these initiatives, to various extents and in different forms, PISA concepts perceived by the local educational policymakers and practitioners from PISA assessment framework and items are made as a reference and integrated. These findings provide solid evidence that some PISA concepts have been used in local-level policies and have been gradually permeating into teaching practices.

To be specific, in response to students' weakness in reading as identified in PISA 2009 China Trial, the reading project was launched in 2011 to increase students' exposure to reading. Under the influence of this project, adding reading into local curriculum as a course for 8^{th} grade students has been required by local educational policy. The reading project had specifically developed a reading textbook and a reading examination, both of which follow PISA reading style, to support and motivate the teaching and learning of practical reading in the course. It also had developed *Recommended bibliography* to expand the range of books from literature to wider types recommended to students at secondary education (see Section 5.1.2.1). The experience of the reading project further convinced the local educational policymakers that some PISA concepts are really helpful for improving teaching, the mathematics project and the science project were then therefore launched successively to explore the way of integrating PISA concepts into teaching to improve learning.

Despite the fact that the mathematics and science projects were not motivated by the local results in PISA, the local area's involvement in PISA opened the opportunity for the local educational policymakers and practitioners to go into the classroom to systematically investigate and reflect on teaching and learning. The mathematics project had developed UBID materials for mathematics teachers as the means to address the issues identified through the investigation. In these materials, some PISA concepts, for example, highlighting applied mathematics problems and embedding problems in real-life contexts, are integrated. Inspired by PISA mathematics literacy to some extent, the mathematics project introduced the term "literacy" by proposing seven categories of "mathematics literacy" (see Section 5.1.2.2.2) in guiding local teaching practices. This term has been used in the instructional design materials and local teacher education activities. It is important to note that, instead of directly employing the concept mathematics literacy and the classification of its subareas as defined in PISA, the mathematics project proposed the seven categories of mathematics literacy according to the project leader and key members' own understanding of elements of mathematics literacy for developing students in the local area. They are the products of the project leader and key members' learning from resources including PISA and domestic research in mathematics. According to PISA 2012 mathematics framework, PISA mathematics uses a set of fundamental mathematics capacities which are condensed from the research of Niss and colleagues (Niss, 2003; Niss and Højgaard, 2011). The fundamental capacities are *communication*; *mathematising*; *representation*; *reasoning and argument*; *devising strategies for solving problems; using symbolic, formal and technical language and operations;* and *using mathematical tools* (OECD, 2013a)*.* Although the proposed seven categories of mathematics literacy do not exactly correspond to the subareas of cognitive process specified in PISA mathematics literacy, they are partly consistent with the fundamental mathematics capacities underpinning PISA mathematics items. Along with the delivery of the UBID materials and the local teacher education sessions, the mathematics project has been promoting these concepts to mathematics teachers. The science project aims to translate PISA science framework into school science teaching by integrating the competencies defined in PISA science framework with national science curriculum in teaching practices. To achieve this, rather than make changes to curriculum standards as some national education systems have done (Breakspear, 2012; Hopfenbeck et al., 2018), this local area was working on changing and improving teachers' teaching methods and strategies to develop students' PISA-like science competencies.

Although PISA assessment framework has been used as a reference and provided insights into the local PISA-motivated policy initiatives, the content of these initiatives are by no means only sticking to what is featured in PISA. For example, as I mentioned above, moving beyond PISA reading framework, the reading project also has proposed broad reading by developing the *Recommended bibliography* for secondary education students, which includes the list of books on literature and a number of other fields. For mathematics initiatives, the proposed PISA concepts are embedded in the UBID which is developed as the teaching materials for mathematics teachers within the local area to also address the issue in teaching that a number of teachers prepared lessons without a holistic view of mathematics knowledge system. Consistent with the use of PISA in reviewing quality of local education that PISA is perceived as additional evidence supplementing domestic sources, the use of PISA in informing local policies and practices is also considered as a part complementing extant content.

Regarding the way that PISA knowledge is used in policymaking, Pons (2012) categorises it into two types: knowledge for policy, denoting using PISA knowledge to legitimise pre-existing policy ideas; and knowledge for learning, referring to deepening the analysis of PISA data for furthering the understanding of PISA results. Ho (2016) suggests that in the East Asian areas such as Shanghai-China, Hong Kong-China, Singapore, and South Korea which have achieved success in PISA, rather than have direct impact on shaping educational policies, PISA is usually used to legitimise pre-existing policy ideas. Not only in East Asian, but also in wider areas, generally, PISA knowledge for policy is more often used, while PISA knowledge for learning is often missing (Pons, 2012). However, the findings of this research that the local PISA-motivated initiatives explicitly drawn concepts from PISA strongly suggest that, rather than “political rhetoric” (Baird et al. 2016), PISA has an important role in informing and inspiring the local policy initiatives.

Literature shows that at national level, one of the major aspects of policy responses to PISA is to develop educational standards and assessment system to evaluate and monitor education system performance, such as in Germany and Danmark (Ertl, 2006; Dolin and Krogh, 2010; Martens and Niemann, 2010; Ringarp, 2016). From the findings of my research we see the difference in the local area in terms of the policy focus. The research data do not show that the local PISA-motivated policy initiatives include introducing a local-level assessment system. Rather, this local area tends to be more focusing on improving local curriculum and teaching methods. It might be partially because that in mainland China there have been established assessment system for secondary education for years, in which high-stakes examinations such as *Zhongkao* and *Gaokao* are traditionally used to evaluate the local or regional educational quality (see Section 2.3). Although a local-wide reading examination targeting students at 8th grade in

Fangshan has been initiated, it is one of the measures to motivate the school-enactment of the local initiatives promoting practical reading, rather than monitor educational development.

It is suggested that PISA provides direction with regard to “best practices” to educational policymakers based on PISA results (Niemann et al., 2017). Similarly, my research reveals that the local educational policymakers in Fangshan also considered that they have got from PISA the direction of making educational improvement. However, as my research findings indicate, the local educational policymakers found the “direction” basically from PISA itself as an assessment tool, for example, PISA framework and items rather than the characteristics and practices of education systems which achieved high in PISA. Researchers argue that PISA has been increasing the phenomenon of “policy borrowing”, since a number of education systems seek best practices from high-performing systems (Sellar and Lingard, 2013a). It is considered as an unintended consequence of assessment use (Johansson, 2016), and has been consistently cautioned in the use of PISA (Oates, 2011; Auld and Morris, 2016). My research does not suggest that at the local context in Fangshan, borrowing policies from elsewhere plays a role in the formation of the local policy initiatives. Rather, besides the knowledge obtained from PISA data, evidence from teaching practices was collected for shaping the PISA-motivated policy initiatives.

Actually, closely linking schools and teaching practices, which is different from that of policy responses to PISA at national level in many other education systems, is the main characteristic of the local PISA-motivated initiatives. The policy-level formation and school-level enactment of these initiatives involve front-line teachers in various ways. For the local educational policymakers, teachers are considered as the ones who know well about teaching and learning, and can be the carrier delivering the proposed ideas inspired by PISA to classroom teaching, so as to influence students’ learning. Involving teachers in these initiatives allows the initiatives to draw rich knowledge and experience about teaching practices from teachers; to inform and educate teachers with the proposed concepts; and also to motivate teachers to bring the proposed concepts into their classroom teaching along with their changing teaching practices.

As discussed in the above, to some extent, the use of PISA in Fangshan seems be in line with what Lockheed and Wagemaker (2013) propose that ILSAs, including PISA, are generally used as “thermometers” and “whips”.

However, the findings of my research further suggest that simply using these metaphors may not be appropriate enough to accurately depict the use of PISA in the focused local context, and may even cause misunderstanding. In terms of the use as “thermometers”, as evidenced in my research and a number of studies focusing on a range of contexts (e.g. Breakspear, 2012; Ho, 2016), PISA is commonly used for reviewing the performance of education systems in international scope. It is true that the OECD average performance in PISA, compared to the normal interval of temperature, is typically used as the benchmark in discussing and comparing results (OECD, 2010; 2014b; 2016a). However, approaching to the top on the international scale of PISA is desired by participating education systems (e.g. Gorur and Wu, 2015). Regarding the use of “whips”, although PISA has motivated a number of policy initiatives in this local area, the findings of this research do not provide evidence suggesting that the initiation of these policies and actions, especially those in relation to mathematics and science, is basically due to the pressure brought by PISA or peer pressure brought by the involvement in PISA. Possibly, this is because the way Fangshan has been involved in PISA is different from that for many of the participating national systems. Fangshan’ local results are not officially published on the league tables by the OECD or China’s national education departments. Therefore, Fangshan’s performance in PISA may not put its local policymakers under as much pressure as that for national policymakers. Although Fangshan’s relatively poor performance in PISA 2009 China Trial motivated the local educational policymakers to immediately take actions to address the issues in reading teaching and learning, the following policy initiatives in relation to mathematics and science were launched due to the appreciation of PISA assessment framework rather than the stimulation of the local mathematics and science performance in PISA. The local educational policymakers, to some extent, seemed to make policies voluntarily and intentionally parallel with some of the concepts in PISA.

The research findings indicate that the use of PISA in informing the local PISA-motivated initiatives is a complex process that many contextual factors can be involved in negotiating the translation of PISA data in the formation of policy initiatives. When it comes to school enactment of these initiatives, even more factors and participants are involved in the negotiation, and therefore subsequently mediate the use of PISA and its impact on students’ learning. In the next section, I will discuss these factors and how they work in the mechanism of PISA's impact on students' learning.

6.2 The mechanism of PISA's impact on students' learning

As introduced before, since PISA is not a high-stakes assessment for individual students, it may not have direct washback on students, that is, students may not intentionally prepare themselves to achieve higher in PISA. However, PISA may indirectly impact students' learning through local educational PISA-motivated policies and teachers' teaching. In the focused context of my research, Fangshan, as shown in interview data, this impact is interacted with and moderated by various factors in different layers of ecological systems (Bronfenbrenner, 1979) around students' learning. Based on the conceptual framework which combines the theories about washback effect (e.g. Alderson and Wall, 1993; Hughes, 1993, cited in Bailey, 1996) and ecological systems theory (Bronfenbrenner, 1979), I created the following Figure 6.1 to conceptualise the findings regarding the mechanism of PISA's impact on students' learning (see the development process of this figure in Appendix A.15). This figure displays various contextual factors in different ecological layers, which influence and mediate the translation of PISA concepts into teaching practices so as to impact students' learning.

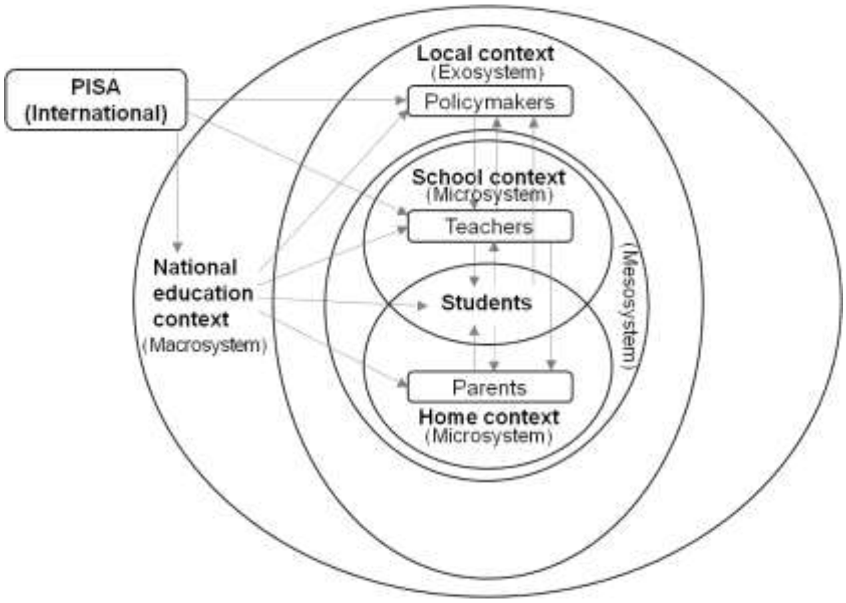


Figure 6.1 The mechanism of PISA's impact on students' learning

In the following, I will discuss how PISA's impact on students' learning occurs in the local area with the influence of the contextual factors in each layer of the ecological systems.

6.2.1 Factors in local context

As presented in Chapter 5, the local PISA-motivated policy initiatives which drawn on PISA concepts to various degrees are closely linked to schools, indicating PISA's indirect impact on students' learning through the school enactment of these initiatives. The findings suggest that these initiatives are the main pathway that PISA has impact on students' learning in the local area, considering the attributes and coverage of these initiatives. However, this impact is not a simple top-down process. Engel (2015) has suggested that PISA's global educational governance is a complex process which is not necessary to be exaggerated, considering the negotiation of domestic educational policymakers. In shaping the local policy initiatives, the local educational policymakers' perceptions about students' performance in PISA and the issues in students' learning identified from PISA and other sources were taken into consideration, even though students were not explicitly present in the local educational policymaking context as active participants. For example, due to students' weak performance in reading non-continuous texts (e.g. tables, graphs), that is, practical reading, one of the main initiatives of the PISA-motivated reading project is to develop students' ability of reading this format of texts (see Section 5.1.2.1). These findings have resonance in the ecological system theory (Bronfenbrenner, 1979) that students' learning and contextual factors in exosystem, that is, local PISA-motivated initiatives, have reciprocal relations.

The findings on PISA's impact on the local educational policy initiatives suggest that, in the exosystem, PISA impacts students' learning mainly in two ways. In one way, through the local PISA-motivated initiatives, some PISA concepts (e.g. practical reading, applied mathematics problems, scientific enquiry) are intended to be translated and incorporated into teaching and learning practices. In the other way, PISA, as an impetus of local educational reforms, provides the local educational policymakers an opportunity to take systematic measures, either targeting student learning content (e.g. adding practical reading into local curriculum) or through improving teaching, to improve students' learning. The findings reveal that the initiatives do not simply aim to prepare students to achieve higher in PISA, which is considered as the negative aspect of washback (Cheng et al., 2004). They are launched not simply for preparing students for performing better on PISA-style tasks by highlighting PISA concepts in teaching and learning, which is of concern of researchers (e.g. Bieber and Martens, 2011; Fulge et al., 2016) arguing the consequences of PISA in terms of bringing policy convergence. Rather, as interview data show, these initiatives pay attention to addressing not only the issues identified in PISA but also those identified in domestic examinations. These findings suggest that PISA's impact on students' learning is negotiated by the particular interests of local educational policymakers (Pons, 2012; Rautalin et al., 2019).

However, the local institutional capacity of assessment in terms of analysing and interpreting PISA data restrains the range and depth of the use of PISA in the local context. This also moderates PISA's impact on students' learning. According to literature (e.g. Breakspear, 2012; Lockheed, 2013), at national level, participation in PISA has built and developed national capacity of assessment in some participating education systems. Nevertheless, at the local area, it seems that resources in relation to this kind of capacity are not yet widely accessible and sufficient for the local consumers of PISA data. For example, the local PISA data consumers perceived the gender difference in favour of boys in PISA mathematics performance, simply based on the comparison of mean scores which are available from the local results report (see Section 5.1.1.1). However, further analysis of PISA data uncovered that mean scores obscure some important factors accounting for the gender difference (see Section 5.3.2). As reflected by the local educational official (see Section 5.1.5), although local educational policymakers have the desire to deepen the use of "PISA knowledge for learning" (Pons, 2012), the lack of expertise in analysing, interpreting and translating PISA data causes confusion when they try seeking insights from PISA for informing solutions to local educational problems. As suggested by Pons (2012), domestic expertise is one of the factors influencing the translation of PISA knowledge. This has echoes in my research. Due to limited technical and analytical capacity of assessment, the local educational policymakers greatly made use of local expertise and rich knowledge in teaching practices for shaping the initiatives. Since the evidence is drawn not only from PISA but also local teaching practices, the application of solutions is situated in the specific local context, which is not decontextualised as argued by some researchers (Gillis et al., 2016).

6.2.2 Factors in school context

At the microsystem level, when the local PISA-motivated initiatives are put into practice in school settings through school-level implementation, the impact of these initiatives on students' learning is further contextualised in the practices in schools' specific settings. Braun et al. (2011) have proposed that the process of policy enactment involves recontextualisation with situated contexts in schools and classrooms, since the environments where policies are enacted may be different in resources, leadership, and challenges in teaching and learning. As my research findings (see Section 5.1.3) reveal, subject to available resources such as staff and budget in schools and schools' own policies, the school implementation of the local PISA-motivated initiatives varies across schools. In the school settings, these initiatives are adapted to some extent to fit the specific situations. For example, to encourage students' reading, reading teachers are specifically assigned in some schools (e.g. School 2); while in other schools (e.g. School 1), the reading course is led by a Chinese teacher due to limited teaching staff. As suggested by Rautalin et al. (2019), schools have their own interpretation of the local PISA-motivated policy initiatives, and adapt these policies according to their own particular interests in practice. For example, some schools (e.g. School 3) set the reading course for students in 7^{th} grade which is the first year of lower-secondary education, while the corresponding local policy actually targets students in 8^{th} grade. These findings are consistent with literature (Braun et al., 2010; Braun et al., 2011) that, rather than the "implementation" of PISA-motivated initiatives, this is "policy enactment", in which policies are translated and interpreted by various policy actors such as school leaders and teachers in school settings.

The findings suggest that, among the factors at school level, teachers play a crucial role as a key agent bringing PISA's impact to students' learning in classrooms, which has resonance in Spillane (1999). Spillane (1999, p.144) proposes that teachers are "the final policy brokers". For the local PISA-motivated initiatives, especially mathematics and science initiatives which aim to improve learning through improving teaching, the degree of their impact on students' learning substantially relies on the changes in teaching. In this case, as suggested by Cohen and Ball (1990), teachers are targets as well as agents of changes. Consistent with McLaughlin (1990), the current research found that this process greatly depends on teachers' capacity and will of changing their teaching practices in response to the policies and initiatives. To some extent, teachers' capacity and will in this respect are built based on their understanding and perceptions of the PISA-motivated initiatives, which are subject to the degree of teachers' involvement in these initiatives. As interview data show (see Section 5.1.3 and Section 5.2.1.3), for the mathematics teachers who were key members of the PISA-motivated mathematics project and even participated in PISA coding (i.e. scoring) (e.g. Mobai, Mingyi), they have relatively deep understanding of PISA concepts, and have been intentionally translating the concepts meaningful for them into their own teaching practices. PISA's impact on students' learning may be more explicit in these teachers' classrooms. For those (e.g. Wenting, Yunlian) who were only marginally involved in a small part of these initiatives or only exposed to PISA concepts in an unconscious way (e.g. attending a demonstration lesson led by the mathematics project), their teaching and their students' learning may be impacted by PISA progressively and unconsciously. This evidences the suggestion proposed by Carvalho (2014) that policy-brokers' proximity-distance regarding PISA framework in terms of their theoretical and methodological knowledge can influence the degree of PISA's impact. It is also consistent with washback hypotheses proposed by Alderson and Wall (1993) that a test may have washback effects for some teachers and some learners but not others.

Although the local PISA-motivated initiatives are adapted to some extent in the school settings, it seems that school-level practitioners tend to endeavour to enact the initiatives which have high stakes attached. As interview data show (see Section 5.1.3), in some schools (e.g. School 4), the instruction time for practical reading as required by the local policy is reduced and even not allocated in the reading course as required by the local policy, there is still time specifically used for instruction in this kind of reading, since students have to attend the reading examination. This is consistent with Maguire et al. (2015) that policy types and power are in relation to policy enactment, and would make the process of enacting policy in schools unstable. From the perspective of the mechanism of washback, this is also in line with one of the hypotheses proposed by Alderson and Wall (1993, p.120-121), that is, "tests that have important consequences will have washback". With the motivation of obtaining good performance in the reading examination, it seems that practical reading has been added into the content of teaching and learning in lower secondary schools, even though in some schools the policy-intended time for teaching this type of reading is compressed.

6.2.3 Factors in interconnection between school context and home context

PISA's impact on students' learning is also influenced by the interconnection between school settings and home settings in which parents play a role. This is consistent with the ecological systems theory that in mesosystems, that is, transitions between different micro systems (i.e. school, home) matter for students' learning (Bronfenbrenner, 1979). It has been recognised that this respect is worth investigation yet underresearched in the studies on washback effect (Bailey, 1999). As previously presented in the conceptual framework (see Section 3.6), neither Hughes' washback model (1993, cited in Bailey, 1996, p.262) nor the washback hypotheses proposed by Alderson and Wall (1993) include parents as one type of participants in washback. Although my research does not particularly look at parents' perceptions about PISA and PISA's impact on students' learning, inspired by the conceptual framework which employs ecological systems theory (Bronfenbrenner, 1979) as one of the theoretical lenses, factors in the mesosystem are identified from data analysis. Interviews with local-level and school-level interviewees (RL, Schl5-R) confirm that the interactions between parents and schools exert influence on students' learning during the process of school enactment of PISA-motivated initiatives. Specifically, parents' perceptions of PISA-motivated initiatives mediate PISA' impact through school-level enactment of these initiatives. It seems that for the initiatives enacted at schools which are of high value for parents, for example, whole-book reading, students are more likely encouraged and supported by their parents to engage in these initiatives at schools or even similar activities at home. In this case, parents' role would enhance PISA' impact on students' learning. Otherwise, it may impede this impact. By making use of this point, as the interview data show (see Section 5.1.4.3), school enactment of some of the PISA-motivated initiatives (especially reading initiatives) have been intentionally involving parents and informing parents about the meaning and value of these initiatives for students' development. Meanwhile, parents' perceptions of these initiatives are also related to the factors in the macrosystem, that is, national education context, which I will discuss below.

6.2.4 Factors in national education context

PISA's impact on students' learning is also mediated by national education contextual factors, especially national high-stakes examinations and national curriculum, to which high value is attached by all the participants situated in various contextual layers in PISA's impact in the local area, as shown in Figure 6.1.

According to the ecological systems theory (Bronfenbrenner, 1979), all the factors in the lower-order systems as discussed above are embedded in the macrosystem, that is, national education system, and would have some consistencies which exist in the system. Bachman and Palmer (1996) suggest that the goals and values of the educational system would mediate the washback of a test. Consistent with these theories, my research findings reveal that, interviewees, either at the local administrative level or school level, are consistently concerned with the content and requirements specified in the syllabus of domestic high-stakes examinations and national curriculum. The interconnections between national education context and the lower-order systems mediate PISA's impact on students' learning in terms of influencing the formation as well as the school enactment of the local PISA-motivated initiatives.

As interview data show (see Section 5.1.4), specifically, at local level, the launch and formation of the local PISA-motivated initiatives have taken into consideration developing students' knowledge and skills which are required by domestic high-stakes examinations and national curriculum. Moving to school settings, meeting requirements of these examinations and curriculum is prioritised in developing students. The reforms in national education context (see Section 2.2 and 2.3) and the literature on the use and impact of PISA at national level in China (Wang and Tong, 2015; Zhao, 2017) indicate that, to some extent, national high-stakes examinations and curriculum tend to be convergent with some PISA concepts. According to interviewees, these changes have also been recognised by school-level practitioners (see Section 5.2.3). Amongst the local PISA-motivated initiatives or the concepts underpinned, those perceived to be in line with the trend in national education context are of priority for school-level practitioners, and are therefore enacted relatively well. Confirming Carvalho (2014), this suggests that domestic on-going reforms on education assessment system and national curriculum can influence the impact of PISA. According to the washback hypotheses (Alderson and Wall, 1993), tests having important consequences will have washback effect on teaching practice, while those having no significant consequences will have no such effect. This is reflected in the influence of national contextual factors on PISA's impact on student learning. Since PISA is not a high-stakes assessment for teachers, as perceived by interviewees (e.g. SchL6), teachers may not necessarily notice PISA or actively translate its concepts in teaching practices (see Section 5.2.1.3). However, teachers consistently pay high attention to the concepts embedded in national high-stakes examinations and the curriculum content covered by high-stakes examinations (see Section 2.3 and Section 5.1.4). For those initiatives or proposed concepts, for example, increasing students' exposure to whole-book reading and contextualising problems, the factors in the national education context facilitate school enactment with regard to these aspects. However, for those not explicitly in relation to high-stakes examinations and national curriculum, they would not attract equal attention during the process of school enactment. These suggest that PISA-motivated initiatives are selected and filtered in school enactment under the mediation effect of national education context. PISA's impact on students' learning is therefore enhanced or diluted to some extent, depending on the strength of their echo in national education context. This is also reflected in the mesosystem. The local policymakers even make use of the washback effect of national high-stakes examinations on parents by encouraging schools to involve parents in school enactment of some of the initiatives. As interviewees (e.g. RL, SchL5-R) reported, under the influence of the national education context, parents have tended to perceive that encouraging their children to read is helpful for children's development and their performance in high-stakes examinations. In turn, as an important role in the transitions between school settings and home settings for students, parents facilitate school enactment of the corresponding initiatives with this perception and attitude.

Although it seems that national assessments and curriculum have been manifesting some convergence with PISA concepts, which enhance PISA's impact on students' learning in the local context in related respects, it is also important to note that domestic knowledge or cultural traditions, for example, promoting whole-book reading, especially reading of classics, also have strong power influencing local educational policymaking and the priority in teaching practices. This is consistent with the suggestion of Michel (2017) that domestic traditions could limit the degree of policy convergence brought by PISA.

6.3 PISA’s impact on student’s mathematics learning outcomes

In this section, I will discuss in detail PISA’s impact on the outcomes of students’ mathematics learning by pooling qualitative as well as quantitative findings. According to Hughes’ washback model, the actions taken by participants of test washback will influence the process of students’ learning at schools which will then influence students’ learning outcomes (Hughes, 1993, cited in Bailey, 1996). For mathematics, the subject which is particularly focused in this research, as the qualitative findings (see Section 5.2) of my research show, it is perceived that, to some extent, PISA and PISA-motivated initiatives have been changing learning content and teaching practices which in turn affect what and how students learn at schools, and contribute to the improvement of students’ learning outcomes. The local educational policymakers used PISA results and domestic resources (e.g. students’ performance in *Zhongkao* and *Gaokao*, classroom observation) as evidence justifying their perceptions. In terms of some aspects, their perceptions have alignment with PISA data, as the quantitative findings suggest. In the following, I will first synthesise the findings of the two strands, and discuss PISA’s impact on the students’ mathematics learning outcomes based on the synthesised evidence.

6.3.1 Synthesising qualitative findings about PISA’s impact on mathematics learning with quantitative findings

To understand and interpret PISA’s practical impact on students’ mathematics learning outcomes, qualitative findings about the perceived impact of PISA on students’ mathematics learning (see Section 5.2) and quantitative findings about students’ performance trends and explanatory factors (see Section 5.3) were synthesised. In the synthesis process, the technique of PIP (Johnson et al., 2019) was employed. It allows for identifying the consistency and contradiction between what people perceived and what PISA data tell, and also for expanding the understanding of local educational policymakers’ and practitioners’ interpretation of local results in PISA (Johnson et al., 2019). Under this technique, categories identified from each of these two strands are listed, matched, checked, based on which the synthesised themes are then built. Qualitative data extracts or quantitative data results are also listed for demonstrating the qualitative or quantitative categories. If there are no matching categories, the boxes are left as “---”. Table 6.1 below displays the process of synthesising the focused part of findings as presented in Section 5.2 and Section 5.3. The following discussion on the synthesised findings are based on the themes built with the PIP technique.

Table 6.1 Pillar integration process

QUAL data	QUAL categories	Pillar building themes	QUAN categories	QUAN data
"Intuitively, seen from PISA results, after we launched this (reading) project, … Moreover, performance in the other two subjects also had improvement." (Local educational official) (see Section 5.2.2.2.1) "(The mathematics project) rather has impact. … the PISA 2015 result compared to 2012 at least indicates that we have stood stably on that level. … Maintaining the relatively high level at least suggests that Fangshan's achievement is not a flash in the pan." (Local educational official) (see Section 5.2.2.2.2)	Perceived impact on students' mathematics cognitive learning outcomes in PISA	The improvement of mathematics performance which suggests the impact of PISA-motivated initiatives	Mathematics performance trend (PISA 2009 China Trial, PISA 2012 China Trial, PISA 2015)	Trend in mathematics mean score: 536-580-566 (see Table 5.2 in Section 5.3.1) Adjusted trend in mathematics mean score: 542-561-566 (see Table 5.3 in Section 5.3.1) Trends in percentages of mathematics high-achievers and low-achievers (see Figure 5.1 in Section 5.3.1) Adjusted trends in percentages of mathematics high-achievers and low-achievers (see Figure 5.2 in Section 5.3.1)
"Of course in our own domestic examinations, students' all aspects of abilities also have improved." (Local educational official) (see Section 5.2.2.2.1)	Perceived impact on students' mathematics cognitive learning outcomes in domestic high-stakes examinations		---	---

Table 6.1 Pillar integration process (continued)

QUAL data	QUAL categories	Pillar building themes	QUAN categories	QUAN data
"… after we launched this (reading) project, … PISA-style reading highlights information collection and processing from various texts. Therefore, it is actually helpful for (learning in) other subjects as well. Comparing (the performance in) PISA 2012 with (that in) PISA 2009, the improvement is very obvious. This should be the most persuasive evidence" (Local educational official) (see Section 5.2.2.2.1)	Increasing students' exposure to reading	Factors associated with mathematics performance: cognitive outcomes-reading	Reading performance has significant effect on mathematics performance	$\beta_{Reading}=0.905$, $p<0.001$ (see Model 4 in Table 5.5 in Section 5.3.2)
"Probably it is because years ago Fangshan had been promoting reading literacy, and (mathematics) teachers' conception has been advanced." (Yunlian) (see Section 5.2.2.2.1)	Highlighting mathematics-related reading			
"engages students in activities as much as possible which allow them to apply mathematics knowledge to real life" (Zhushan) (see Section 5.2.1.2.1)	Highlighting applied mathematics	Factors associated with mathematics performance: OTL-exposure to applied mathematics problems	Students' exposure to applied mathematics had no significant effect on mathematics performance	$\beta_{EXAPPLM}=-6.792$, $p=0.196$ (see Model 2 in Table 5.5 in Section 5.3.2)
"When designing a lesson, they used to think about how to give a wonderful lecture. … Now they think about how to design it to allow the content accessible to more of their students. " (ML) (see Section 5.2.1.1.1)	Teaching has been transformed to highlight student orientation and cognitive activation	Factors associated with mathematics performance: teaching practices and quality (student-oriented approach and cognitive activation)	Student-orientated mathematics teaching approach had no significant effect on mathematics performance	$\beta_{TCHBEHSO}=-7.300$, $p=0.122$ (see Model 2 in Table 5.5 in Section 5.3.2)
"For example, to deliver the instruction in the application of trigonometric functions, I asked students to find resources. They then gave a presentation within five minutes." (Mingyi) (see Section 5.2.1.1.1)	Teaching has been transformed to highlight student orientation and cognitive activation		Cognitive activation in mathematics teaching had positive effect on mathematics performance at the border line of 0.05 significance level	$\beta_{ANCCOGACT}=14.861$, $p=0.055$ (see Model 2 in Table 5.5 in Section 5.3.2)

Table 6.1 Pillar integration process (continued)

QUAL data	QUAL categories	Pillar building themes	QUAN categories	QUAN data
---	---		Teacher support had significant effect on mathematics performance	$\beta_{TEACHSUP}=11.070$, $p<0.05$ (see Model 2 in Table 5.5 in Section 5.3.2)
"During the presentation, students could confidently write down formulas, concepts or examples on the blackboard and give explanations, sometimes they even do not need a script", "at least has a certain relationship (with the school's initiatives stimulated by the local PISA-motivated mathematics project)" (SchL2) (see Section 5.2.2.1.1)	Perceived impact on students' self-beliefs in learning mathematics	Factors associated with mathematics performance: non-cognitive outcomes (self-beliefs, interest)	Self-efficacy, self-concept, anxiety in learning mathematics had significant effects on mathematics performance	$\beta_{MATHEFF}=16.770$, $p<0.001$ $\beta_{ANCSCMAT}=15.315$, $p<0.01$ $\beta_{ANXMAT}=-10.892$, $p<0.01$ (see Model 3 in Table 5.5 in Section 5.3.2)
"Through various activities, students' interest in learning has been stimulated. They are willing to learn (mathematics)" (SchL1) (see Section 5.2.2.1.2)	Perceived impact on students' interest in learning mathematics		Interest in learning mathematics had no significant effect on mathematics performance	$\beta_{ANCINTMAT}=3.659$, $p=0.708$ (see Model 3 in Table 5.5 in Section 5.3.2)
"Through the lead of our project in addition to the implementation of the new curriculum, the changes of teachers are considerable, indeed considerable." (ML) (see Section 5.2.3) "I think (high-stakes) examinations still act as a director. How they assess (students) determines how teachers teach." (SchL1) (see Section 5.2.3)	National education context moderates the effect of PISA and PISA-motivated initiatives on mathematics learning		---	---

6.3.2 The local mathematics performance has been improved

The synthesised findings from qualitative and quantitative data indicate that, to some extent, with the influence of PISA-motivated initiatives, Fangshan students' mathematics performance has been improved. The increased performance trends across PISA cycles are consistently considered by the local interviewees (i.e. the local official, RL, ML) as the explicit evidence suggesting the PISA-motivated initiatives' practical impact on students' mathematics performance. As reported by them, in terms of scores, Fangshan students' performance, including that in mathematics, has made great improvement from PISA 2009 China Trial to PISA 2012 China Trial. By comparing Fangshan's performance with that of other participants in PISA 2015, they believed that Fangshan' performance remains high in PISA 2015.

Quantitative analysis of PISA data confirm these perceptions, but also reveal the limitations with regard to their understanding and interpretation of local PISA results. As the analysis on observed trends in mean performance shows (see Table 5.2 in Section 5.3.1), from PISA 2009 China Trial to PISA 2012 China Trial, Fangshan students' mathematics performance had a considerable improvement (536 versus 580). Consistent with what ML stated, despite that PISA 2015 data show that the score of Fangshan academic school students dropped to 566, the overall mathematics performance of Fangshan is still much higher than the international benchmark, that is, average mean score of OECD countries (490) in that cycle (OECD, 2014b). Perhaps because of the mixed picture of the observed trends in performance, the local educational policymakers justified the effects of the local PISA-motivated initiatives in an arguable way by comparing its local performance with other systems rather than its own trends in the mathematics scores over the three PISA cycles. Researchers found that the changes in student population composition in age, gender, and educational level over cycles can be influential for interpreting trends in performance (Aloisi and Tymms, 2017; Melkonian and Areepattamannil, 2018). Echoing this, my quantitative analysis adjusting for these changes over the three cycles of PISA do find the difference between the adjusted scores and the observed scores. The adjusted trends clearly suggest the continuous improvement of students' mathematics performance in Fangshan from PISA 2009 China Trial to PISA 2015 (542 – 561 – 566) (see Table 5.3 in Section 5.3.1). The adjusted trends in the percentages of mathematics high-achievers and low-achievers further indicate the performance improvement (see Figure 5.2 in Section 5.3.1). However, the qualitative findings (see Section 5.2.2.2) based on interview data do not provide evidence suggesting that the local educational policymakers had realised the importance of the changes in student population demographic characteristics (e.g. age, gender, educational level) for interpreting the trends of students' performance. One reason explaining their unawareness of this point would be the limited assessment capacity in analysing and interpreting data in the local area, which has been reflected by the local educational official, as I discussed in Section 6.2.1. The unawareness of adjusted trends also suggests that in delivering PISA results to national and local contexts, some important information for appropriately interpreting the results are not sufficiently circulated or are even omitted. Although in PISA international results reports, students' performance adjusted for demographic changes (e.g. age, gender) in target student population has started to be reported since PISA 2012 (OECD, 2014b; 2016a), the importance of interpreting performance trends with the consideration of the demographic changes is far from sufficiently underlined in the reports. It is not uncommon to be neglected by general users of PISA data, considering the content about adjusted trends is only a small piece embedded in hundreds of pages of the report. As to Fangshan's local results reports, according to my knowledge, it was composed by the PISA national team, mainly describing the key results of the local PISA data, while content about adjusted trends is not included.

6.3.3 The factors contributing to the improvement of mathematics performance

Interviewees perceived that, to some extent, both PISA-motivated reading initiatives and mathematics initiatives are linked to the improvement of students' mathematics performance. They considered that the initiatives conducted by these projects had been gradually influencing the process of teaching and learning by changing teaching practices and learning content, which subsequently influences the non-cognitive (e.g. self-beliefs, interest) and cognitive learning outcomes (i.e. mathematics achievement). Those identified perceptions of the local education policymakers and practitioners are largely supported by what I found from PISA data. In the following, I will discuss the factors embedded in the PISA-motivated initiatives that interviewees perceived to have contribution to the improvement of students' mathematics performance, and to what extent they are supported by PISA data.

6.3.3.1 Increasing students' exposure to reading

The interview data (see Section 5.2.2.2.1) show that local educational policymakers believed that the local initiatives promoting reading has had contribution to the improvement of mathematics performance. The reading project, in which PISA-style reading is promoted, is the first and only PISA-motivated project in Fangshan launched between the administration of PISA 2009 China Trial and PISA 2012 China Trial. The "very obvious" "improvement" shown not only in reading but also in mathematics and science from PISA 2009 to PISA 2012 is considered by the local educational official as "the most persuasive evidence" suggesting the positive effect of promoting students' reading ability, which is the main initiative of the PISA-motivated reading project. Although it is admitted by interviewees (e.g. local education official, RL) that it is hard to quantify the contribution of the reading project for the improvement of mathematics performance, they believed that the reading project "definitely" facilitated that improvement. Besides the perceived influence of the reading project, mathematics teachers have gradually had higher awareness of the importance of reading for mathematics problem solving. As I discussed before (see Section 2.3, 3.3.1, 6.2.4), inspired by PISA, the ongoing reforms of domestic high-stakes examinations tend to emphasise the employment of contextualised mathematics problems. The convergence between these reforms and the local initiatives in this respect have strengthened its influence on teaching practices. As school-level interviewees (e.g. SchL1, Yunlian) perceived, under this influence, reading has been gradually valued by mathematics teachers in their teaching practices, since teachers perceive that reading ability can help for comprehending mathematics problems, especially those with relatively high reading demands. Some schools even have had their own policies promoting mathematics-related reading within their schools (e.g. School 2). With these initiatives, it is reasonable to assume that students in Fangshan tend to have more exposure to the PISA-style texts of problems, and have more opportunities to practice and develop their ability of comprehending this type of texts.

Interviewees' perceptions regarding the influence of initiatives promoting students' reading ability on mathematics performance are confirmed by the quantitative results of multilevel modelling with PISA data, which show significant association between students' reading performance and mathematics performance ($\beta_{reading}$=0.905, p<0.001). The consistent findings in these two strands of data analyses also coordinate with a number of studies.

PISA mathematics tasks which typically employ word problems are known as having relatively high reading demands (Wu, 2010). For solving this type of mathematics problems, researchers suggest that problem comprehension is a crucial process (Mayer, 1986; Boonen et al., 2013; Boonen et al., 2016; Fuchs et al., 2016), in which students often make errors (Lewis and Mayer, 1987). The relatively long text (Mullis et al., 2013; Walkington et al., 2017) and linguistic-semantic factors underlining the text (van der Schoot et al., 2009; LeFevre et al., 2010) may hinder students' understanding of problems. Although the result of my research based on the analysis of PISA data still cannot quantify the extent to which the local PISA-motivated reading initiatives have contributed to the improvement of mathematics performance over the three PISA cycles, it does show that reading ability explains some variation in mathematics performance. By jointly looking at both the qualitative and quantitative results, it seems plausible to suggest that the local reading initiatives have had some impact on students' mathematics achievement. Moreover, this impact has been facilitated and strengthened with the influence of the ongoing reforms in domestic high-stakes examinations which have begun to have increased reading demands.

6.3.3.2 Increasing students' exposure to applied mathematics problems

Another change in the content of student mathematics learning at schools, as suggested by qualitative findings, is increasing students' exposure to applied mathematics problems embedded in real-life contexts. According to interview data, interviewees reported that under the influence of PISA-motivated mathematics project, mathematics teachers had started to have the awareness of developing students' ability of applying mathematics knowledge to solving real-life problems in their teaching practices. They perceived that these changes have been gradually increasing students' self-beliefs and interest in learning mathematics, which then contribute to improving their mathematics performance. These perceptions concur with Schmidt et al., (2001) and Schmidt and Maier (2009) who suggest that the OTL, that is, students' content exposure, is related to students' learning at schools. However, these perceptions do not have echoes in quantitative results. Multilevel analysis results show that students' experience with applied mathematics was not significantly associated with mathematics performance ($\beta_{EXAPPLM}$=-6.792, p=0.196). The inconsistent results between qualitative data and quantitative data may suggest more complex rather than linear relationships between the exposure to applied mathematics and the performance in mathematics. The international results reports of PISA have found curvilinear relationships between them across education systems that mathematics performance increases along with more exposure to applied mathematics and then after a certain point it decreases (OECD, 2014b). As Schmidt and Burroughs (2015) suggest, regarding the employment of applied mathematics in teaching practices, “more is not necessarily better” (p.31). Actually, this point has been noted by some school-level practitioners (e.g. Mingyi) that, when embracing PISA concepts and translating them in practices, it is still important to keep some traditionally-focused contents in teaching (see Section 5.1.5).

6.3.3.3 Increasing use of student-oriented teaching approach and cognitive activation

Changes in teaching practices and instructional quality in terms of increasing the employment of student-oriented teaching approach and cognitive activation activities in classrooms were also perceived by local educational policymakers as factors which contribute to improving students’ mathematics performance. At school level, these changes were also commonly and consistently reported by mathematics teacher interviewees. Moving to quantitative findings, however, student-oriented teaching had no significant effect on mathematics performance with a negative coefficient ($\beta_{TCHBEHSO}$=-7.300, p=0.122), while cognitive activation was positively associated with mathematics performance at the border line of 0.05 significance level ($\beta_{ANCCOGACT}$=14.861, p=0.055).

Regarding the effects of student-oriented teaching approach and cognitive activation on mathematics performance, literature (e.g. Caro et al., 2016; Yi and Lee, 2017; Demir, 2018), as reviewed previously, have shown inconsistent results (see Section 3.4.2.2). In terms of student orientation, Demir (2018) found negative effect of student-oriented teaching practices on students’ PISA mathematics performance in Turkey, while Caro et al (2016) found that the relationships between this factor and mathematics performance are inconsistent across education systems. The inconsistent relationships as revealed by the findings of these previous studies and the merged results of my research support Cuban’s (1983) view that it is hard to evaluate the effect of a specific teaching approach on students’ performance. They also indicate the need of critical interpretation of the effectiveness of student-oriented teaching practices. Its effectiveness can be differential for students of different characteristics and learning strategies and in different contexts (Caro et al., 2016). Since PISA has limitations in portraying actual teaching practices in classroom (Feniger and Lefstein, 2014), it would be problematic to deny interviewees' perception about the effect of this kind of teaching practices simply with PISA data-based evidence. However, to claim that the use of student-oriented teaching approach improves students' performance, which was perceived by local educational policymakers, further investigation on how this approach is practiced in specific classroom settings may be needed before making that conclusion.

As to cognitive activation, the synthesised findings of my research are in line with Yi and Lee's (2017) and Demir's (2018) studies which suggest its positive effects on mathematics performance in some other contexts. However, Caro et al. (2016) argue that their relationships are curvilinear. Researchers argue that it is problematic to use a single indicator for the complex construct of cognitive activation (as PISA does) to predict student performance, since cognitive activation can be different in different types of lessons (e.g. introduction lessons versus practice lessons), and is subject to the content of lessons and how it is implemented (Praetorius et al., 2014). Another point of concern is that, as stated by the OECD (2010), PISA assesses cumulative student learning outcomes; however, the measures of cognitive activation employed in PISA student questionnaire ask "Thinking about the mathematics teacher that taught your last mathematics class: How often does each of the following happen" (OECD, 2013a, p.238). Hence, considering the complexity of this construct and the limitations in PISA data, further investigation is still needed to explore the impact of cognitive activation on student performance.

6.3.3.4 Enhancing self-beliefs and interest

Along with the changes in the process of teaching and learning at schools, interviewees perceived that non-cognitive learning outcomes in terms of self-beliefs and interest in mathematics learning had been increasing, which were considered by interviewees to be crucial for students' mathematics performance. The importance of these factors for mathematics performance have been discussed in a range of studies (e.g. Hembree, 1990; Bandura, 1992; Pajares and Miller, 1994; Bandura, 1997; Ryan and Deci, 2009; Lee and Stankov, 2018). In terms of the effect of self-beliefs, the qualitative findings of my research in this respect are confirmed by the quantitative analysis on the relationships between these non-cognitive outcomes factors and mathematics performance. As shown in the multilevel modelling results, both students' self-efficacy and self-concept in mathematics learning had positive association with mathematics performance, while anxiety about mathematics was negatively related with performance in this domain. The quantitative findings are in line with the findings based on PISA data in other contexts (Lee and Stankov, 2013; OECD, 2013d; Şahin and Yıldırım, 2016; Cheung et al., 2018; Lee and Stankov, 2018). Although for East Asian areas such as Shanghai-China, Hong Kong-China, Macao-China, Chinese Taipei, Japan and Korean, the contradictory results of students' self-concept and their mathematics performance are reported in some studies (OECD, 2004; Ho, 2009; OECD, 2013d), researchers argue that it might be explained by the "Big-Fish-Little-Pond effect" (Marsh and Parker, 1984; Marsh et al., 2008) or response styles on Likert-scales which are used for measuring self-concept (Cheung et al, 2018). The multilevel modelling in my research, employing the anchored indicator of self-concept which adjusted for this potential influence on students' responses, reveals the positive effect of self-concept on mathematics performance. The consistency between qualitative and quantitative findings in terms of the effects of self-beliefs suggests that PISA-motivated initiatives may have had impact on students' mathematics performance through the increased self-beliefs in mathematics learning, if these initiatives do have increased students' self-beliefs as perceived by interviewees. For interest in learning mathematics, the analysis of Fangshan PISA data tells a story different from qualitative finding. Multilevel modelling analysis shows that it had no significant relationship with mathematics performance. This is aligned with that observed in other East Asian areas such as Shanghai-China, Hong Kong-China, Chinese Taipei and Singapore (OECD, 2013d). Unlike self-concept, interest in learning mathematics does not have a corresponding anchored indicator in PISA database. Therefore, it is not clear whether this result can be explained to some extent by response-styles. The inconsistency between qualitative and quantitative findings in this respect indicates that further investigation on the relationship between learning interest and mathematics performance may be needed.

In summary, the findings synthesised from both qualitative and quantitative data strands with the PIP approach provides me different lenses to understand the topic of interest. On the one hand, with deep discussion with interviewees, some important themes are identified from interview data, which are not captured by the analyses of Fangshan PISA data. For example, the effects of national education contextual factors on students' mathematics learning has been commonly and consistently reported by interviewees, while this important theme is even not covered in PISA assessment framework. On the other hand, employing PISA databases to further analyse the local performance, trends, and factors associated with the performance allows me to obtain clearer understanding about the accuracy and appropriateness of local educational policymakers and practitioners' interpretations about PISA results. In terms of the interpretation of performance trends in PISA, in addition to interviewees' reported perception, deep analyses on Fangshan PISA data conducted in my research provide evidence of the limited capacity of analysing and interpreting PISA data in the local area. Synthesising the findings of these two data strands therefore further the understanding in this respect. The synthesised findings also provide extended evidence suggesting PISA's impact on students' learning. Because PISA is a cross-sectional study, one cannot claim causality simply from the association between factors and mathematics performance identified from quantitative analysis based on PISA databases. The factors statistically correlated with mathematics performance do not necessarily affect mathematics performance causally, since they might be proxies of some other constructs, or the causality relationships might be reversed. However, the interview data which show consistency with these quantitative findings provide some confidence to argue that, to some extent, students' improvement in mathematics performance, as measured by PISA, is attributed to the local PISA-motivated initiatives with the influence of the underlined factors.

6.4 Contributions of this research

My research shows an detailed account of an empirical investigation on PISA's impact on students' learning, in which how that impact occurs has been analysed and evaluated. Through this research, contributions have been made in terms of adding to the body of knowledge of understanding PISA's impact on students' learning, providing insights into methodology for investigating the impact of influential ILSAs on educational practices, and extending related theories. The contributions to these aspects are discussed further in the following three subsections.

6.4.1 Contributions to knowledge

My research contributes to the body of knowledge of the impact of PISA in domestic contexts. Although discussion on this topic, particularly the impact of PISA on national policies, has been widely documented in a range of studies, my research extends the discussion by looking further into PISA's practical impact on students' learning in a local context, which is rarely addressed in extant literature (see Section 6.1). Moving from the theoretical debate and argumentation on PISA's impact in terms of its global governance power on national education policies (e.g. Grek, 2009; Meyer and Benavot, 2013; Sellar and Lingard, 2014; Sjøberg, 2015), and theoretical assumption of mediation effect of contextual factors on PISA's impact (e.g. Carvalho, 2014; Carvalho and Costa, 2015; Tikly, 2015), the empirical investigation conducted in my research provides substantial evidence of how the impact actually occurs in the local context. The research findings depict how PISA data have been used and translated at policy level, and reveal how the use and translation of PISA data influence students' learning through policy enactment in schools (see Section 6.2). As found in my research, to some degree, PISA has been influencing students' learning. However, not like that claimed by some researchers (e.g. Grek, 2009; Meyer and Benavot, 2013), its impact, which is subject to the mediation effects of various domestic factors and their interactions, appears to be not strong enough for governing the local education. My research also reveals the paradox that there is a strong desire and substantial initiatives of using PISA for informing local education policies and practices, while the local capacity of expertise in analysing, interpreting, and translating the data of complex assessment programmes (e.g. PISA) is limited. This is an important issue, considering that PISA would have increasing influence in local contexts along with the expanding participation of local areas and schools in its extension programme, PISA-based Test for Schools (Rutkowski, 2015; Lewis et al., 2016; Lewis, 2017). However, it is rarely discussed in literature. For this situation, my research suggests that working on technical and professional support for local PISA data consumers to appropriately interpret and translate PISA data into practices seems to be urgently needed.

6.4.2 Contributions to methodology

My research also has contributions to the methodology for investigating the impact of PISA or other influential ILSAs on educational practices in domestic contexts. It offers an innovative methodological approach which synthesised the theories about washback effect (Anderson and Walker, 1993; Hughes, 1993, cited in Bailey, 1996) and the ecological systems theory (Bronfenbrenner, 1979) as the theoretical underpinnings for investigating the impact of PISA. The theories about washback effect inform my mixed-methods research design, and the employment of the ecological systems theory supports the rationale to build the considerations of dynamic contextual factors into my research. The hybrid of these theories underpins well the analysis of what PISA's impact is and how it occurs in practice with the consideration that various policy agents at different contextual levels may participate in and mediate the impact. Hence, it further provides the analytical framework for analysing data and interpreting findings.

One could argue that the employment of these theories in research is not uncommon, since in the fields such as language testing, developmental psychology, and education, there have been a range of studies using either washback effect or the ecological systems theory to underpin research design and discussion. However, to my knowledge, bringing together these two strands of theories is rarely seen in extant studies on the use and consequences of PISA or other influential ILSAs. It is found that the term "washback effect" (e.g. Baird et al., 2011; Aloisi and Tymms, 2017) or "the ecological systems theory" (e.g. Tikly, 2015) has been referred to in some studies discussing the impact of PISA or ILSAs. However, it appears that none of them jointly use these theories to frame and underpin their research, and describe in detail how these theories inform their investigation and discussion. The conceptual framework employed in my research could provide insights into the design of future empirical research on examining the impact of PISA or other ILSAs on local and school policies and practices.

6.4.3 Contributions to theories

Based on the framework and my research findings, the mechanism of PISA's impact on students' learning in the local context, as conceptualised in my research (see Figure 6.1), further developed the theories about washback effect and the ecological systems theory. Adding to the trichotomy washback model proposed by Hughes (1993, cited in Bailey, 1996, p.262), the conceptualised mechanism of PISA's impact on students' learning suggests that besides the perceptions and actions of washback participants which subsequently affect student learning outcomes, the interconnections amongst these participants and their agencies also play an important role influencing PISA's impact on student learning. Moreover, with the evidence of parents' influence in the interconnections between school settings and home settings, it adds to Hughes' washback model that parents can also be a type of participants involved in washback effect. Furthermore, complementing to the theories about washback, my research findings about PISA's impact mediated by national high-stakes examinations indicate that, depending on the characteristics and significance of tests perceived by washback participants, the washback effect of one test can be strengthened or diluted by that of another test. In addition to these contributions, the conceptualised findings of my research also extends the ecological systems theory. As introduced before (see Section 3.6.2), in the ecological systems theory (Bronfenbrenner,1979), the outmost contextual layer for human development is macrosystem. It is defined as consistencies exist in lower-order systems in terms of common culture, belief or ideology within a given society or social group (Bronfenbrenner,1979). The empirical evidence about PISA's impact on students' learning as identified in my research further indicates that the macrosytem can be expanded to the global scale, considering the "glocalisation" (Robertson, 1995) of the world. This theoretical contribution could shed light on the theoretical framework to be used for future empirical research on investigating the impact of PISA or other influential ILSA in local contexts in other education systems.

6.5 Limitations of the research

Despite the contributions of this research as outlined above, it is admitted that this research also has some limitations.

6.5.1 Limitations in generalisation and transferability of findings

This research employed interviews to investigate the use of PISA and the perceived impact of PISA on students' learning in Fangshan local context. Although the interviews provide in-depth understanding in this respect, based on which the overall picture of how PISA's impact occurs in this local area has been conceptualised, it is important to note that the findings of this research by no means suggest that, in Fangshan, PISA's impact on students' learning in each school and each classroom works exactly the same. To identify the extent to which PISA's impact on students' learning in a specific school setting and classroom setting, the conceptualised mechanism could be used as the framework for investigation, yet, as discussed before, the particular micro settings (e.g. schools) are crucial factors still meriting consideration. Fangshan was focused in my research. It would be problematic to generalise the research findings for understanding PISA's impact in other local contexts in China, or directly transfer the findings to local contexts in other countries. Fangshan is a special case, since it is the only local area that had been continuously involved in PISA for three cycles. It is reasonable to assume that because of this experience, Fangshan had much more intensive exposure to PISA concepts than many other local areas. Moreover, in different local areas, the agents who might participate in the process of translating PISA data in local contexts and their activities would also vary. Therefore, the impact of PISA may not be necessarily manifested in other local contexts in China or other countries the same as that identified in Fangshan. However, the mechanism of PISA's impact on students' learning, as revealed in this research, indicates some common factors that are worthy of consideration in investigating and understanding PISA's impact in other contexts.

6.5.2 Limitations in making causal inference

The data employed in my research also restrain my ability to make causal inference based on research findings, since longitudinal data for examining the perceived improvement of student performance which is attributed to the local PISA-motivated initiatives are not available. As the local PISA-motivated initiatives were launched before the start of my PhD, it was not feasible to undertake a longitudinal study design in which one could employ pre-test and post-test to respectively collect data of students' performance before and after the enactment of these initiatives to analyse the effect of these initiatives on students' learning. Domestic longitudinal assessment data that could be used to achieve this aim are also not available. Therefore I chose to use PISA data to identify factors associated with students' performance, and then matched and synthesised them with the perceived attributes of the local PISA-motivated initiatives to triangulate interviewees' perception of the impact of PISA and related initiatives. The systhesised results from qualitative and quantitative data provide some evidence supporting that PISA has been gradually influencing students' learning in the local area. However, to make a strong causation claim, further evidence is still needed.

6.6 Summary

This chapter has dicussed the key findings of this research (see Section 6.1, 6.2, and 6.3). PISA has been explicitly used as a supplementary tool informing local education policymaking in Fangshan, with the local education policymakers' recognition of the credibility of PISA measures and framework. Although "PISA shock" did not occur in Fangshan, PISA has directly motivated a number of local policy initiatives. The reading project is the response to the local results in PISA to address the students' weaknesses identified, while the mathemactics project and the science project are mobilised by the local education policymakers' appreciation of some of PISA concepts and their voluntary intentions of translating them into teaching to improve learning. The explicit links between PISA and these initiatives suggest that PISA is not used in this local area as "political rhetoric" which is the case in many national-level uses as documented in the literature. It has had substantial impact on the local education policymaking. Many of the initiatives appear to align with some concepts of PISA (e.g. practical reading, applied mathematics, contextualising problems). However, additional considerations of addressing issues in teaching and learning identified from domestic recources and meeting the practical needs of teaching and learning are taken in shaping these initiatives. These initiatives worked closely with schools and teachers in seeking answers to improving teaching and learning based on the knowledge obtained from PISA and local teaching practices, rather than borrowing policies from elsewhere.

Through the school implementation of these initiatives, PISA indirectly impacts students' learning, with various contextual factors involved in negotiating how the impact occurs. Informed by the conceptual framework, this chapter conceptualised and discussed the mechanism of this impact. At local contextual level, PISA impacts students' learning through informing and inspiring the local PISA-motivated initiatives which aim to address the issues identified in PISA and those identified in domestic resources to improve teaching and learning. PISA concepts and data are an important part of resources shaping many of these initiatives. However, local limited technical and analytic capacity of assessment constrains the width and depth of the local use of PISA in informing policymaking. The variation of school enactment of the initiatives suggests their impact is negotiated by school-level agents' interpretation and translation, subject to school resources, policies, interests and needs. Teachers are the key agent for the initiatives focusing on improving learning through improving teaching practices. Their capacity and will to change their teaching practices in response to the policies initiatives, to some extent, are subject to the degree of their involvement in these initiatives. Hence, the intensity of the impact is different across teachers and students. PISA's impact on students' learning is also influenced by the interconnections between parents and schools, with parents' engagement in the school implementation of some of the PISA-motivated initiatives. All these local-level, school-level, and school-home interconnection factors are influenced by national education context during the process of shaping and enacting the local PISA-motivated initiatives. Although recent reforms of national assessment and curriculum show some convergence with PISA concepts, cultural-specific contents and domestic traditions are still kept. Because of the significance of national high-stakes examinations and national curriculum, which is not the case for PISA, the initiatives which have convergence with national high-stakes and curriculum are enacted better in schools, and tend to receive more support from parents. PISA's impact on students' learning is therefore enhanced or diluted, subject to the strength of the initiatives' resonance with the factors in national education context.

The synthesis of qualitative and quantitative findings convinces stakeholders that, to some extent, PISA-motivated initiatives have improved Fangshan students' mathematics performance. It also further reveals the challenges in using and interpreting PISA data faced by local PISA data consumers due to limited capacity in assessment. The consistent evidence obtained from both data strands suggests that the local initiatives increasing students' exposure to reading are an important factor contributing to the improvement of students' performance in PISA mathematics assessment which has relatively high reading demands. The inconsistent results as to the effect of initiatives increasing students' opportunity to learn applied mathematics problems supports the caution on overuse of this type of problems in teaching. The synthesised results about the effectiveness of student-oriented teaching approach and cognitive activation indicate that these teaching practices' contribution to students' performance is arguable, and needs to be further investigated with considerations of characteristics of specific students and classes. However, students' self-beliefs in learning mathematics, enhanced through these teaching practices, foster the improvement of mathematics performance.

Drawing on these key findings, this chapter has highlighted the research contributions to knowledge, methodology, and theories (see Section 6.4). Firstly, it adds to the knowledge about the impact of PISA on students' learning in local context, and identified the factors in different contextual layers which negotiate the range and depth of the impact. It also uncovers the urgent needs of providing local PISA data consumers with technical and professional support for appropriate and efficient use of PISA in local context. Secondly, the mixed-methods research design and contextual considerations of this research, which are underpinned by the conceptual framework weaving theories about washback and ecological systems theory, provide methodological insights into the design of future empirical investigation into the impact of PISA or other influential ILSAs in local context. Thirdly, this research also adds to the theories about washback effect regarding the participants involved in students' learning, and extends the connotation of "macrosystems" as defined in the ecological systems theory. Since this research only focused on one local context and student-level longitudinal data are not available, limitations in generalisation and transferability of the research findings, and in making causal inference have also been pointed out in this chapter (see Section 6.5).

Chapter 7 Conclusion

This research has investigated the use of PISA and the overall picture of the impact of PISA on students' learning in the local area, Fangshan District of Beijing. It has revealed the mechanism of how the impact occurs with the negotiation of different contextual factors. Particularly focusing on the impact on mathematics learning, this research also investigated the perceived impact on the process of mathematics teaching and learning in schools and the outcomes of mathematics learning by analysing the data of perspectives of both local educational policymakers and practitioners. Those perceptions were further triangulated and understood based on the analyses of Fangshan PISA data. From the key findings as discussed in Chapter 6 above, it can be concluded that PISA has been used by the local educational policymakers to motivate and inform policy initiatives for improving local education. As to the mechanism of PISA's impact on students' learning, it involves the mediation effects of various agents and their interconnections in different contextual layers. Through school enactment of the local PISA-motivated initiatives in addition to recent reforms in national high-stakes examinations and national curriculum which have some convergence with PISA, to some extent, PISA has been gradually influencing some aspects of students' learning. However, whether all of these initiatives contributed to the students' performance improvement is still arguable and needs to be further investigated.

In the following, I will conclude with research implications and recommendation for future research, and finally make methodological as well as personal reflection on conducting this research.

7.1 Research implications

In this section, I discuss my research implications for researchers, assessment developers and educational policymakers respectively in Section 7.1.1, Section 7.1.2 and Section 7.1.3.

7.1.1 Implications for researchers

This research indicates that with the global reach and expanding participation of PISA, PISA has been promoting the breaking of barriers between international discourse on education and local educational practices. As to discussing the extent of the impact that PISA has brought to local education, this research highlights the importance of taking into consideration contextual factors that would moderate the impact of PISA in domestic contexts. The research findings indicate that although PISA has been influencing domestic education, due to culture inertia, PISA's power of governing domestic education systems and narrowing the diversity in the definitions of education across domestic education systems seems not to be as strong as some suggest (e.g. Riley and Torrance, 2003; Grek, 2009; Labaree, 2014). As revealed by the findings, it is admitted that some of PISA concepts have been incorporated into domestic education systems, which perhaps makes education systems tend to be convergent to some extent in some aspects. However, those appreciated concepts are generally translated and adapted to fit the needs of the specific domestic educational environment. That process is also subject to the capacity, interests, and willingness of various agents, such as educational policymakers and practitioners. Most importantly, besides those concepts learned from PISA, there are still some culturally specific elements that are kept, valued, and probably also prioritised in domestic education systems. Therefore, this research indicates that it is far from sufficient to evaluate and judge the consequences of an assessment such as PISA simply based on policy documents or speeches. It underlines the necessity of conducting empirical investigation on the assessment consequences in a specific context, and recruiting various participants who are involved in the mechanism of PISA's impact.

7.1.2 Implications for assessment developers

Indicated by interviewees' reflection on translating PISA data into policies and practices (see Section 5.1.5), this research confirms the importance of seeking answers to facilitating better use of PISA (or other influential ILSAs) data and thereby facilitating positive impact of this assessment. Assessment developers which release assessment data files and results reports could make an effort in this respect. There is no doubt that, like many other assessments, PISA also has limitations (see Section 3.5), and it is important to critically scrutinise and evaluate those limitations, a process which may advance the improvement of PISA assessment instruments. Clearly, informing PISA data consumers of those limitations and other issues (e.g. possible changes in the demongraphic composition of student population) is also crucial for data interpretation. As my research findings indicate, although some of the local policymakers and practitioners recognised the limitations in terms of content coverage, it appeared that they basically did not have clear ideas about the technical limitations or issues in the data that need to bear in mind for interpreting data. The expanding impact of PISA in domestic contexts and even in local areas seems to urge the need of widely informing PISA data consumers in local areas of these limitations and issues in an accessible way. For example, the OECD could provide the instruction on PISA databases and the manual (or online training sessions) on secondary data analyses which highlight those limitations and key issues and are written in relatively plain language accessible to general PISA data users.

7.1.3 Implications for education policymakers

This research has some implications for national and local education policymakers as well. As discussed previously (see Section 3.5.5), PISA data have limitations in identifying causality relationships, which restrains the depths of using PISA data for informing educational policies and practices. To overcome these limitations, for national education policymakers in relevant authorities or government departments which are in charge of national-level administration of PISA, it is worth considering collecting additional data through extending PISA international assessment instruments and sampling design. In addition, Fangshan's continuous participation in PISA, which allows for tracking its local performance trends and to some extent evaluating the effects of the local educational initiatives, indicates that either for a region or a country, continuous participation in an assessment with samples consistently representing the same population could make more uses of assessment data possible. This merits education policymakers' consideration in discussing and determining the way of participation. As for local education policymakers who would like to make attempts to use and translate PISA data for informing local educational policies and practices, insights can be gained from the use of PISA in Fangshan which confirms the importance of engaging teachers (see Section 5.1.2). Besides, deep analyses of PISA data are also important for appropriate data interpretation, which serves for informing policymaking. Although these may bring challenges to local education policymakers, as revealed by my research findings (see Section 5.1.5), they are still necessary before translating PISA data into educational policies and practices. To achieve that, local education policymakers could seek external expertise in analysing and interpreting PISA data, and technical support from PISA national team or the OECD.

7.2 Future research

Since this research only focused on one specific local context, it is meaningful to replicate this research and apply the conceptualised mechanism as the basis to investigate PISA's impact in other local contexts which also have had high engagement with PISA. The growing body of local areas participating in PISA-based Test for Schools (Lewis et al., 2016; Lewis, 2017) suggests that in next few years PISA may have direct impact in more and more local areas, which underlines the importance of conducting such research in the future. Moreover, further research on comparing the mechanism in different local contexts within China or across different education systems can be conducted to develop the conceptualised mechanism to see how PISA's impact occurs and varies across different contexts, which may shed light on the practices of learning from other systems.

My research findings show consistent evidence from both qualitative and quantitative data that reading performance is important for students' mathematics performance. In another study conducted by Ding and Homer (2020), it is found that specific cognitive processes of reading are associated with mathematics performance. To claim whether they have causation relationship, and how one affects the other, further research seeking additional and stronger evidence is still needed. Along with the reform in domestic assessment system that high-stakes examinations on mathematics tend to have higher reading demands than they used to be, reading had been considered by educational practitioners as a sought-after ability in mathematics teaching and learning. Yet, in which way that reading ability influences mathematics performance, and in which way developing students' reading ability can advance their mathematics learning is not very clear. Hence, it might be meaningful to further examine the relationship between reading and mathematics performance through empirical classroom intervention and evaluate the impact of the intervention that highlights reading (or specific cognitive processes of reading) in mathematics classes.

This research focused on the understanding and perspectives of local educational policymakers and practitioners; therefore, it did not additionally collect data from national-level stakeholders, such as those engaged in recent reforms in national assessment system and curriculum. As conceptualised in the mechanism of PISA's impact, national education contextual factors, especially national reforms of high-stakes examinations and curriculum, play a crucial role in the negotiation of PISA's impact on students' learning. Further research could also consider the perspectives of national-level participants involved in PISA's impact on national education reforms. For example, through interviews with these people, future research could obtain rich data for deeper understanding of how PISA data are used and translated in informing national-level policymaking so as to depict a more comprehensive picture about interconnections between national contextual factors and local contextual factors that negotiate and moderate the impact of PISA on students' learning.

7.3 Methodological reflection

In this section, methodological reflection on the timing for conducting interviews, background of interviewees, the design of assessment instrument, and the display of the process of synthesising findings is presented.

7.3.1 Timing for conducting interviews

To investigate school leaders' perceptions of students' performance change and the impact of PISA-related initiatives at school level, ideally, the potential school leader participants should be those who had worked in their current schools for several years, knew teaching and students of their schools very well, and knew something about school enactment of PISA-related initiatives (and PISA). However, during the recruitment of school leaders, I realised that the population of these potential participants was small. I learned from the local educational official and one of the school contacts that, due to the national policy of teacher rotation between schools, school leaders might be transferred to another school every three or five years. I did not clearly note this issue at the stage of research design. In this case, the timing of commencing fieldwork for collecting interview data from this group of participants is important. For example, in School 5, the school leader who could contribute to the understanding of the enactment of local PISA-motivated initiatives in this school was unavailable because he/she had moved to another school. Hence, when recruiting participants, I was flexible to some extent if necessary. As to School 5, instead of recruiting a school leader who had few idea in this respect, I interviewed one reading teacher (i.e. Sch5-R) who could contribute informative data. Future research doing similar investigation can take the timing issue into consideration, and when necessary, may allow flexibility to some extent in recruiting interviewees.

7.3.2 Background of interviewees

The knowledge that school leaders contributed to answering my research questions was in relation to their expertise, that is, their subject disciplines, although it was also related to their division of responsibilities in schools. School leaders' expertise would to some extent affect their involvement in and understanding about the PISA-motivated initiatives undertaken in their schools. Therefore, it would affect their perceptions of these initiatives. For example, SchL3 and SchL4, their subject discipline is Chinese, so they talked more about reading-related initiatives as opposed to mathematics and science initiatives in their schools. When I tried to probe more about the school enactment of the initiatives related to mathematics, they sometimes even directly suggested me to explore that in the interviews with the mathematics teachers. This indicates that researchers can have some pre-interview expectation about interviewees' contribution to answering RQs according to interviewees' expertise background. When it is feasible, taking into account the background of interviewees in interviewee recruitment is helpful for collecting rich data.

7.3.3 Design of assessment instrument

As described in Section 4.6.2.2.2 that PISA 2012 database contains a number of missing values due to the rotation design of PISA 2012 student questionnaire that each student only received one of the three questionnaire forms, in my research, the issue of missing values was addressed with multiple imputation. Considering the complex procedure of addressing missing values in PISA 2012 student questionnaire, the rotation design of PISA 2012 student questionnaire seems not user-friendly for general users of PISA data, since in most cases one has to impute missing values before exploring something of interest with the questionnaire data. As found out by my research, local PISA data consumers such as local educational policymakers are struggling with limited capacity of processing and analysing PISA data. The rotation design employed in PISA 2012 student questionnaire, to some extent, may have hindered the potential utilisation of this questionnaire database. In terms of PISA, an influential assessment informing educational policymaking, the usability and convenience of its data for secondary analyses are needed to be critically discussed and considered in designing the assessment instrument.

7.3.4 Display of the process of synthesising findings

Displaying the process of synthesising findings in mixed methods research is important, since it helps readers to understand how researchers merge findings, and how researchers come up with the interpretation of the merged findings. Explicit display of the synthesis process could also enhance the trustworthiness of research. As introduced before (see Section 6.3.1), the PIP technique (Johnson et al., 2019) was employed in my research for explicitly displaying the process of synthesising findings identified from two data strands. However, before I identified this technique and determined to use it, I was struggling to find an appropriate way for displaying the synthesis process. Although there have been many reference books on mixed methods research, few of them methodologically introduce the possible procedures of the synthesis process and clearly describe how the process can be displayed. Finally, I decided to target the journals which specifically focus on mixed methods research, for example, *Journal of Mixed Methods Research*, and tried to seek the answer from the state-of-art studies published in such journals. Through searching these studies, I identified the PIP technique and recognised its appropriateness for being used in my research. This process of seeking the synthesis technique indicates that it can be a valuable attempt to seek methodological solutions from latest innovative studies when common reference books fail to give an answer. Still, the appropriateness of using new techniques in one’s own research needs to be critically considered.

7.4 Personal reflexive account

My research topic was inspired by my previous work experience in PISA National Centre in China. This experience provided me the convenience of obtaining Fangshan PISA data, and approaching to the research field site and interviewees. However, this experience and the previous role I had in national-level PISA administration also presented me some constraints in conducting interviews in the local area, even though I had clearly told interviewees that I had left PISA National Centre, and their data would not be shared with any individuals or institutions. For example, some of the interviewees may have been reserved to some extent in expressing freely their perceptions of some issues such as those about the local PISA results reports which were composed by PISA National Centre. To keep comfortable discussion during interviews, I therefore did not probe into details about their opinions and thoughts on those issues.

Through conducting this research I learned that the knowledge and experience that one has obtained from different disciplines and in different spaces can be connected, and provide one the lenses to view and understand the world. With this thought, I came up with the theoretical underpinnings of the research. My master's degree is in language testing, from which I have got knowledge about washback effect. With the understanding of the related theories, I considered that policymakers' responses to PISA look like the washback effect in language testing and also in domestic high-stakes examinations. Since PISA has had explicitly-claimed impact on educational policymaking, it might also have some impact on students' learning if the PISA-motivated policy initiatives are actually put into practice. The mechanism of washback which has been considerably studied in the fields of language testing and high-stakes examinations could be used as the framework to research into PISA's impact on students' learning. In addition, my experience as a teaching assistant in the module ”What is a Child? Child Development and Learning” at undergraduate level extended my knowledge about learning theories, one of which is ecological systems theory. I thought it fitted the need of my research in terms of accounting for factors in different layers of contexts in which PISA's impact on students' learning might occur. It can also be well connected and combined with washback effect which also suggests contextual considerations.

List of References

Adams, R. 2003. Response to “cautions on OECD’s recent educational survey (PISA)”. *Oxford Review of Education*. **29**(3), pp.377-389.

Adams, R. and Wu, M. 2002. *PISA 2000 technical report*. Paris: OECD.

Acock, A.C. 2014. *A gentle introduction to Stata*. 4th ed. Lakeway Drive: Stata Press.

Addey, C. 2017. Golden relics & historical standards: how the OECD is expanding global education governance through PISA for Development. *Critical Studies in Education*. **58**(3), pp.311-325.

Alderson, J.C. and Wall, D. 1993. Does washback exist? *Applied Linguistics*. **14**(2), pp.115-129.

Aloisi, C. and Tymms, P. 2017. PISA trends, social changes, and education reforms. *Educational Research and Evaluation*. **23**(5–6), pp.180-220.

Alpert, R. and Haber, R.N. 1960. Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology*. **61**(2), pp.207-215.

Ashcraft, M.H. and Krause, J.A. 2007. Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review*. **14**(2), pp.243-248.

Auld, E. and Morris, P. 2016. PISA, policy and persuasion: translating complex conditions into education “best practice”. *Comparative Education*. **52**(2), pp.202-229.

Aune, T.K., Ingvaldsen, R.P., Vestheim, O.P., Bjerkeset, O. and Dalen, T. 2018. Relative age effects and gender differences in the National Test of Numeracy: a population study of Norwegian children. *Frontiers in Psychology*. **9**, article no: 1091 [no pagination].

Bachman, L.F. and Palmer, A. S. 1996. *Language testing in practice*. Oxford: Oxford University Press.

Bailey, K.M. 1996. Working for washback: a review of the washback concept in language testing. *Language Testing*. **13**(3), pp.257-279.

Bailey, K. M. 1999. *TOEFL monograph series: washback in language testing*. Princeton, NJ: Educational Testing Service.

Baird, J-A., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T. L. and Daugherty, R. 2011. *Policy effects of PISA*. [Online]. Oxford: Oxford University Centre for Educational Assessment. [Accessed 10 November 2016]. Available from http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2011/10/Policy-Effects-of-PISA-OUCEA.pdf

Baird, J.-A., Johnson, S., Hopfenbeck, T.N., Isaacs, T., Sprague, T., Stobart, G. and Yu, G. 2016. On the supranational spell of PISA in policy. *Educational Research*. **58**(2), pp.121-138.

Baker, D.P. and Jones, D.P. 1993. Creating gender equality: cross-national gender stratification and mathematical performance. *Sociology of Education*. **66**(2), pp.91-103.

Bandura, A. 1977. *Social learning theory*. London: Prentice-Hall.

Bandura, A. 1997. *Self-efficacy: the exercise of control*. NY: W.H. Freeman ad Company.

Bandura, A. 1992. Exercise of personal agency through the self-efficacy mechanism. In: Schwarzer, R. ed. *Self-efficacy: thought control of action*. Washington, D. C.: Hemisphere, pp.3-28.

Barron, K.E. and Harackiewicz, J.M. 2001. Achievement goals and optimal motivation: testing multiple goal models. *Journal of Personality and Social Psychology*. **80**(5), pp.706-722.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M. and Tsai, Y.-M. 2010. Teachers’ mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*. **47**(1), pp.133-180.

Baumgartner, H. and Steenkamp, J.-B.E.M. 2001. Response styles in marketing research: a cross-national investigation. *Journal of Marketing Research*. **38**(2), 143-156.

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzales, E.J., Kelley, D.L. and Smith, T.A. 1996. *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Beijing Fansghan government. 2017. *Fangshan overview*. [Online]. [Accessed 1 December 2017]. Available from: http://www.bjfsh.gov.cn/zjfs/fsgk/index.htm

Beilock, S.L. and Maloney, E.A. 2015. Math anxiety: a factor in math achievement not to be ignored. *Policy Insights from the Behavioral and Brain Sciences*. **2**(1), pp.4-12.

Bell, J. 2005. *Doing your research project: a guide for fist-time researchers in education, health and social science*. 4th ed. Maidenhead: Open University Press.

Bieber, T. and Martens, K. 2011. The OECD PISA study as a soft power in education? Lessons from Switzerland and the US. *European Journal of Education*. **46**(1), pp.101-116.

Bonnet, G. 2002. Reflections in a critical eye: on the pitfalls of international assessment. *Assessment in Education: Principles, Policy & Practice*. **9**(3), pp.387-399.

Boonen, A.J.H., de Koning, B.B., Jolles, J. and van der Schoot, M. 2016. Word problem solving in contemporary math education: a plea for reading comprehension skills training. *Frontiers in Psychology*. **7**, article no: 191 [no pagination].

Boonen, A.J.H., van der Schoot, M., van Wesel, F., de Vries, M.H. and Jolles, J. 2013. What underlies successful word problem solving? A path analysis in sixth grade students. *Contemporary Educational Psychology*. **38**(3), pp.271-279.

Boylan M., Wolstenholme, C., Maxwell, B., Jay, T., Stevens, A. and Demack, S. 2016. *The longitudinal evaluation of the mathematics teacher exchange: China-England*. *Interim research report*. [Online]. London: Department for Education. [Accessed 19 September 2018]. Available from https://www.gov.uk/government/publications/evaluation-of-the-maths-teacher-exchange-china-and-england

Bracey, G.W. 2004. International comparisons: less than meets the eye? *Phi Delta Kappan*. **85**(6), pp.477-478.

Bradley, R.H. and Corwyn, R.F. 2002. Socioeconomic status and child development. *Annual Review of Psychology*. **53**, pp.371-399.

Brady, H.E. 1985. The perils of survey research: inter-personally incomparable responses. *Political Methodology*. **11**(3/4), pp.269-291.

Braun, A., Ball, S.J., Maguire, M. and Hoskins, K. 2011. Taking context seriously: towards explaining policy enactments in the secondary school. *Discourse: Studies in the Cultural Politics of Education*. **32**(4), pp.585-596.

Braun, A., Maguire, M. and Ball, S.J. 2010. Policy enactments in the UK secondary school: examining policy, practice and school positioning. *Journal of Education Policy*. **25**(4), pp.547-560.

Braun, V. and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*. **3**(2), pp.77-101.

Braun, V., Clarke, V. and Weate, P. 2016. Using thematic analysis in sport and exercise research. In: Smith, B. and Sparkes, A. C. eds. *Handbook of qualitative research in sport and exercise*. London: Routledge, pp.191-205.

Breakspear, S. 2012. *The policy impact of PISA: an exploration of the normative effects of international benchmarking in school system performance. OECD Education Working Papers*, No. 71. Paris: OECD Publishing.

Brochu, P. 2014. The influence of PISA on educational policy in Canada: take a deep breath. *Šolsko polje*. **25**(5/6), pp.73-86.

Bronfenbrenner, U. 1979. *The ecology of human development: experiments by nature and design*. Cambridge, Mass: Harvard University Press.

Brown, P., Halsey, A.H., Lauder, H. and Wells, A.S. 1997. The transformation of education and society: an introduction. In: Halsey, A.H., Lauder, H., Brown, P. and Wells, A.S. eds. *Education culture economy society*. Oxford: Oxford University Press, pp.1-44.

Brown, G., Micklewright, J., Schnepf, S.V. and Waldmann, R. 2007. International surveys of educational achievement: how robust are the findings? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. **170**(3), pp.623-646.

Bryman, A. 2016. *Social research methods*. 5th ed. Oxford: Oxford University Press.

Buck, G. 1988. Testing listening comprehension in Japanese university entrance examinations. *JALT Journal*. **10**(1), pp.15-42.

Buckingham, J. 2012. *Keeping PISA in perspective: why Australian education policy should not be driven by international test results*. St Leonards: the Centre for Independent Studies. [Accessed 7 May 2017]. Available from: http://www.cis.org.au/product/keeping-pisa-in-perspective-why-australian-education-policy-should-not-be-driven-by-international-test-results/

Campbell, J., Smith, D., Boulton-Lewis, G., Brownlee, J., Burnett,.P.C., Carrington, S. and Purdie, N. 2001. Students' perceptions of teaching and learning: the influence of students' approaches to learning and teachers' approaches to teaching. *Teachers and Teaching*. **7**(2), pp.173-187.

Cao, L. 2014. Chengshihua jiasu fazhan jieduan ruhe tongchou chengxiang fazhan – jiyu Beijing Fangshan chengshihua shijian de fenxi. *Mordern Economic Information*. 13, pp.471-472.

Carnoy, M., Khavenson, T., Loyalka, P., Schmidt, W.H. and Zakharov, A. 2016. Revisiting the relationship between international assessment outcomes and educational production evidence from a longitudinal PISA-TIMSS sample. *American Educational Research Journal*. **53**(4), pp.1054-1085.

Caro, D.H., Lenkeit, J. and Kyriakides, L. 2016. Teaching strategies and differential effectiveness across learning contexts: evidence from PISA 2012. *Studies in Educational Evaluation*. **49**, pp.30-41.

Caro, D.H., McDonald, J.T. and Willms, J.D. 2009. Socio-economic status and academic achievement trajectories from childhood to adolescence. *Canadian Journal of Education / Revue canadienne de l'ĂŠducation*. **32**(3), pp.558-590.

Carroll, J.B. 1963. A model of school learning. *Teachers College Record*. **64**(8), pp.723-733.

Carvalho, L.M. 2014. The attraction of mutual surveillance of performances: PISA as a knowledge policy instrument In: Fenwick, T., Mangez, E., and Ozga, J. eds. *World yearbook of education 2014. Governing knowledge: comparison, knowledge-based technologies and expertise in the regulation of education*. London: Routledge, pp.58-72.

Carvalho, L.M. and Costa, E. 2015. Seeing education with one's own eyes and through PISA lenses: considerations of the reception of PISA in European countries. *Discourse: Studies in the Cultural Politics of Education*. **36**(5), pp.638-646.

Chai, W. and Xiang, Q. 2015. Woguo jianli yiwu jiaoyu zhiliang jiance zhidu, jinnianqi kaizhan jiance gongzuo: zhongguo tese "PISA" jiekai miansha. *China Education Daily*. 16 April, p.1.

Chen, F. 2016. PISA impact on basic educational reform trend in UK: a case study of and comments on "England-Shanghai Mathematic Teacher Exchange Project". *Journal of Schooling Studies*. **13**(5), pp.107-112+90.

Chen, C., Lee, S. and Stevenson, H.W. 1995. Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*. **6**(3), pp.170-175.

Cheng, L., Sun, Y. and Ma, J. 2015. Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*. **48**(4), pp.436-470.

Cheng, L., Watanabe, Y. and Curtis, A. eds. 2004. *Washback in language testing: research contexts and methods*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cheung, K., Mak, S. and Sit, P. 2018. Resolving the attitude–achievement paradox based on anchoring vignettes: evidences from the PISA 2012 mathematics study. *Asia Pacific Education Review*. **19**(3), pp.389-399.

Chung, J. 2010. Finland, PISA, and the implications of international achievement studies on education policy. In: Wiseman, A.W. ed. *The impact of international achievement studies on national education policymaking (International perspectives on education and society, Vol. 13)*. Bingley: Emerald Group Publishing Limited, pp.267-294.

Chung, J. 2016. The (mis)use of the Finnish teacher education model: "policy-based evidence-making"? *Educational Research*. **58**(2), pp.207-219.

Clarke, V. and Braun, V. 2017. Thematic analysis. *The Journal of Positive Psychology*. **12**(3), pp.297-298.

Cobley, S., McKenna, J., Baker, J. and Wattie, N. 2009. How pervasive are relative age effects in secondary school education? *Journal of Educational Psychology*. **101**(2), pp.520-528.

Cohen, D. K. and Ball, D. L. 1990. Policy and practice: an overview. *Educational Evaluation and Policy Analysis*. **12**(3), pp. 233-239.

Cong, G. 2014. Wash back effects of new College Entrance Examination chemistry papers: taking test papers of Jiangsu province as examples. *Examinations Research*. 2, pp.19-23+18.

Creswell, J.W. 2007. *Qualitative inquiry and research design: choosing among five approaches*. 2nd ed. Los Angeles: Sage.

Creswell, J.W. and Clark, V.L.P. 2011. *Designing and conducting mixed methods research*. 2nd ed. Thousand Oaks: Sage.

Cuban, L. 1983. How did teachers teach, 1890–1980. *Theory Into Practice*. **22**(3), pp.159-165.

Dai, J. 2007. Cong kaoshi dao pingjia: jiaoyu gaige de shidai renwu. *China Higher Education*. **Z2**, pp.21-23.

Damiani, V. 2016. Large-scale assessments and educational policies in Italy. *Research Papers in Education*. **31**(5), pp.529-541.

Deci, E. L. and Ryan, R.M. 1985. *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Demir, S.B. 2018. The effect of teaching quality and teaching practices on PISA 2012 mathematics achievement of Turkish students. *International Journal of Assessment Tools in Education*. **5**(4), pp.645-658.

Ding, H. and Homer, M. 2020. [Forthcoming] Interpreting mathematics performance in PISA: taking account of reading performance. *International Journal of Educational Research*

Dobbins, M., and Martens, K. 2010. A contrasting case: the USA and its weak response to internationalisation processes in education policy. In: Martens, K., Alexander Kenneth, N., Windzio, M. and Weymann, A. eds. *Transformation of education policy*. Basingstoke: Palgrave, pp.179-195.

Dolin, J. and Krogh, L.B. 2010. The relevance and consequences of PISA science in a Danish context. *International Journal of Science and Mathematics Education*. **8**(3), pp.565-592.

Dong, H. 2015. Woguo jiaoyu shiye maishang xin taijie. *People's Daily*. 13 October, p.12.

Duru-Bellat, M. and Suchaut, B. 2005. Organisation and context, efficiency and equity of educational systems: what PISA tells us. *European Educational Research Journal*. **4**(3), pp.181-194.

Eccles, J.S. 1994. Understanding women's educational and occupational choices: applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly*. **18**(4), pp.585-609.

Education Supervision Office of Shanghai Municipal Government, 2013. Shanghai tuijin yiwu jiaoyu junheng fazhan de zhuyao jingyan yu chengxiao. [Online]. [Accessed 5 October 2018]. Available from: http://www.moe.gov.cn/s78/A11/s3077/moe_2421/201307/t20130709_154010.html

Egelund, N. 2008. The value of international comparative studies of achievement - a Danish perspective. *Assessment in Education: Principles, Policy & Practice*. **15**(3), pp.245-251.

Eivers, E. 2010. PISA: issues in implementation and interpretation. *The Irish Journal of Education / Iris Eireannach an Oideachais*. **38**, pp.94-118.

Else-Quest, N.M., Hyde, J.S. and Linn, M.C. 2010. Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological Bulletin*. **136**(1), pp.103-127.

Engel, L.C. 2015. Steering the national: exploring the education policy uses of PISA in Spain. *European Education*. **47**(2), pp.100-116.

Engel, L.C. and Frizzell, M.O. 2015. Competitive comparison and PISA bragging rights: sub-national uses of the OECD's PISA in Canada and the USA. *Discourse: Studies in the Cultural Politics of Education*. **36**(5), pp.665-682.

Engel, L.C. and Rutkowski, D. 2014. Global influences on national definitions of quality education: examples from Spain and Italy. *Policy Futures in Education*. **12**(6), pp.769-783.

Ertl, H. 2006. Educational standards and the changing discourse on education: the reception and consequences of the PISA study in Germany. *Oxford Review of Education*. **32**(5), pp.619-634.

Fan, G. and Sun, W. 2018. The historical and logical analysis of the education system and mechanism reform in the past 40 years of reform and opening up in China. *Educational Research (China)*. 7, pp.15-23+48.

Fangshan District Bureau of Statistics and Fangshan Survey Team of National Bureau of Statistics. 2016. *Statistical yearbook 2016 of Fangshan District of Beijing*. Beijing: Fangshan District Bureau of Statistics.

Feniger, Y. and Lefstein, A. 2014. How not to reason with PISA data: an ironic investigation, *Journal of Education Policy*. **29**(6), pp.845-855.

Ferla, J., Valcke, M. and Cai, Y. 2009. Academic self-efficacy and academic self-concept: reconsidering structural relationships. *Learning and Individual Differences*. **19**(4), pp.499-505.

Forestier, K., Adamson, B., Han, C. and Morris, P. 2016. Referencing and borrowing from other systems: the Hong Kong education reforms. *Educational Research*. **58**(2), pp.149-165.

Freitas, P., Nunes, L.C., Reis, A.B., Seabra, C. and Ferro, A. 2016. Correcting for sample problems in PISA and the improvement in Portuguese students' performance. *Assessment in Education: Principles, Policy & Practice*. **23**(4), pp.456-472.

Fuchs, L.S., Gilbert, J.K., Powell, S.R., Cirino, P.T., Fuchs, D., Hamlett, C.L., Seethaler, P.M. and Tolar, T.D. 2016. The role of cognitive processes, foundational math skill, and calculation accuracy and fluency in word-problem solving versus pre-algebraic knowledge. *Developmental psychology*. **52**(12), pp.2085-2098.

Fuchs, T. and Wößmann, L. 2007. What accounts for international differences in student performance? A re-examination using PISA data. *Empirical Economic*s. **32**(2–3), pp.433-464.

Fulge, T., Bieber, T. and Martens, K. 2016. Rational intentions and unintended consequences: on the interplay between international and national actors in education policy. In: Mundy, K., Green, A., Lingard, B., and Verger, A. eds. *The handbook of global education policy*. Malden, MA: Wiley, pp. 453-469.

Gabriel, F., Signolet, J. and Westwell, M. 2018. A machine learning approach to investigating the effects of mathematics dispositions on mathematical literacy. *International Journal of Research & Method in Education*. **41**(3), pp.306-327.

Gao, S. and Zhang, X. 2015. Yi PISA wei qiji tisheng quyu yuedu suyang. *Teachers' Journal*. 10, pp.13-17.

Gao, Y. 2010. Gaokao yingyu chansheng de fumian fanbo xiaoying de biaoxian ji yinsu touxi. *Modern Education Science*. 2, pp.10-11+29.

Gebhardt, E. and Adams, R.J. 2007. The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement*. **8**(3), pp.305-322.

Gillis, S., Polesel, J. and Wu, M. 2016. PISA data: raising concerns with its use in policy settings. *The Australian Educational Researcher*. **43**(1), pp.131–146.

Gist, J.R. 1998. Decision making in public administration. In: Rabin, J., Hildreth, W.B. and Miller, G.J. eds. *Handbook of Public Administration*. 2nd ed. New York: Marcel Dekker, pp.265-292.

Goldstein, H. 2004. International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education: Principles, Policy & Practice*. **11**(3), pp.319-330.

Goldstein, H. 2014. *REALCOM-impute: multiple imputation using MLwiN, user guide*. [Online]. Bristol: University of Bristol. [Accessed 27 June 2018]. Available from: http://www.bristol.ac.uk/cmm/software/realcom/imputation.html

Goldstein, H. 2018. Measurement and evaluation issues with PISA. In: Volante, L. ed. *The PISA effect on global educational governance*. New York: Routledge, pp.49-58.

Goldstein, H. and Thomas, S.M. 2008. Reflections on the international comparative surveys debate. *Assessment in Education: Principles, Policy & Practice*. **15**(3), pp.215-222.

Gorur, R. 2011. ANT on the PISA trail: following the statistical pursuit of certainty. *Educational Philosophy and Theory*. **43**(s1), pp.76-93.

Gorur, R. 2016. Seeing like PISA: a cautionary tale about the performativity of international assessments. *European Educational Research Journal*. **15**(5), pp.598-616.

Gorur, R. and Wu, M. 2015. Leaning too far? PISA, policy and Australia's "top five" ambitions. *Discourse: Studies in the Cultural Politics of Education*. **36**(5), pp.647-664.

Gove, M. 2010. Pisa slip should put a rocket under our world-class ambitions and drive us to win the education space race. *Times Education Supplement*. [Online]. 17 December. [Accessed 22 September 2018]. Available from: https://www.gov.uk/government/speeches/education-secretary-writes-for-the-times-educational-supplement

Gove, M. 2011. *Education for economic success*. [Online]. 11 January. Education World Forum. [Accessed 22 September 2018]. Available from: https://www.gov.uk/government/speeches/michael-gove-to-the-education-world-forum

Greene, J.C. Caracelli, V.J. and Graham, W.F. 1989. Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*. **11**, pp.255-274.

Grek, S. 2009. Governing by numbers: the PISA 'effect' in Europe. *Journal of Education Policy*. **24**(1), pp.23-37.

Grisay, A. 2003. Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*. **20**(2), pp.225-240.

Grisay, A., de Jong, J.H.A.L., Gebhardt, E., Berezner, A. and Halleux-Monseur, B. 2007. Translation equivalence across PISA countries. *Journal of Applied Measurement*. **8**(3), pp.249-266.

Gu, M., Cai, Z. and Zhang, H. 2019. Educational reform and development in China: yesterday, today and tomorrow. *Chongqing Higher Education Research*. **7**(2), pp. 5-11.

Guan, D. 2017. Woguo zhongxuesheng shuxue chengji de xingbie chayi yanji: jiyu PISA 2009, 2012 and 2015*. Shuxue Tongbao.* **56** (9), pp.47-51.

Guba, E.G. 1981. Criteria for assessing the trustworthiness of naturalistic inquiries. *ECTJ*. **29**(2), p.75.

Guo, Z., Wu, Y. and Zhang, J. 2015. Jiyu PISA shijian de quyu jiaoyu zhiliang tisheng. *Teacher's Journal*. 10, pp.8-12.

Gustafsson, J.-E. 2008. Effects of international comparative studies on educational quality on the quality of educational research. *European Educational Research Journal*. **7**(1), pp.1-17.

Hackman, D.A., Farah, M.J. and Meaney, M.J. 2010. Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nature Reviews Neuroscience*. **11**(9), pp.651-659.

Hamp-Lyons, L. 1997. Washback, impact and validity: ethical concerns. *Language Testing*. **14**(3), pp.295-303.

Han, J. 2001. Gaokao neirong gaige tanxi. *Zhongxue Huaxue Jiaoxue Cankao*. 1-2, pp.5-12.

Hauser, R.M. 1994. Measuring socioeconomic status in studies of child development. *Child Development*. **65**(6), pp.1541-1545.

He, J., Buchholz, J. and Klieme, E. 2017. Effects of anchoring vignettes on comparability and predictive validity of student self-reports in 64 cultures. *Journal of Cross-Cultural Psychology*. **48**(3), pp.319-334.

Heckman, J.J. 2000. Policies to foster human capital. *Research in Economics*. **54**(1). pp.3-56.

Hembree, R. 1990. The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*. **21**(1), pp.33-46.

Heyneman, S. P. and Lee, B. 2014. The impact of international studies of academic achievement on policy and research. In: Rutkowski, L., von Davier, M. and Rutkowski, D. eds. *Handbook of international large-scale assessment: background, technical issues, and methods of data analysis*. Boca Raton: CRC Press, pp.117-140.

Ho, S.C. 2009. Characteristics of East Asian learners: what we learned from PISA. *Educational Research Journal*. **24**(2), p.327-348.

Ho, S.C. 2016. The use of large-scale assessment (PISA): insights for policy and practice in the case of Hong Kong. *Research Papers in Education*. **31**(5), pp.516-528.

Hoffman, B. 2010. “I think I can, but I’m afraid to try”: the role of self-efficacy beliefs and mathematics anxiety in mathematics problem-solving efficiency. *Learning and Individual Differences*. **20**(3), pp.276-283.

Hopfenbeck, T.N., Lenkeit, J., Masri, Y.E., Cantrell, K., Ryan, J. and Baird, J.-A. 2018. Lessons learned from PISA: a systematic review of peer-reviewed articles on the Programme for International Student Assessment. *Scandinavian Journal of Educational Research*. **62**(3), pp.333-353.

Hopkins, D. and King, G. 2010. Improving anchoring vignettes: designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly*. **74**(2), pp.201-222.

Hopko, D.R., Ashcraft, M.H., Gute, J., Ruggiero, K.J. and Lewis, C. 1998. Mathematics anxiety and working memory: support for the existence of a deficient inhibition mechanism. *Journal of Anxiety Disorders*. **12**(4), pp.343-355.

Hox, J.J. 2010. *Multilevel analysis* [Online]. 2nd ed. New York: Routledge. [Accessed 25 June 2018]. Available from: https://www.dawsonera.com/abstract/9780203852279

Hugener, I., Pauli, C., Reusser, K., Lipowsky, F., Rakoczy, K. and Klieme, E. 2009. Teaching patterns and learning quality in Swiss and German mathematics lessons. *Learning and Instruction*. **19**(1), pp.66-78.

Hughes, A. 1989. *Testing for language teachers*. Cambridge: Cambridge University Press.

Hughes, A. 2003. *Testing for language teachers*. 2nd ed. Cambridge: Cambridge University Press.

Husén, T. 1967. *International study of achievement in mathematics: a comparison of twelve countries*, Vol. 2. New York: Wiley.

Husén, T. 1979. An international research venture in retrospect: the IEA surveys. *Comparative Education Review*. **23**(3), pp. 371-385.

Husman, J. and Lens, W. 1999. The role of the future in student motivation. *Educational Psychologist*. **34**(2), pp.113-125.

Hyde, J.S. 2005. The gender similarities hypothesis. *American Psychologist*. **60**(6), pp.581-592.

IBM Corp. 2015. *IBM SPSS Statistics for Windows* (Version 23.0). [Software]. [Accessed 1 October 2016].

IEA. 2018. *Help manual for the IEA IDB Analyzer (Version 4.0)*. Hamburg, Germany.

Jerrim, J. 2015. Why do East Asian children perform so well in PISA? An investigation of Western-born children of East Asian descent. *Oxford Review of Education*. **41**(3), pp.310-333.

Ji, M. and Wang, M. 2012. Guanyu “Lvse zhibiao” xiangmu jianjie. *Modern Teaching*. Z2, pp.8-15.

Jiang, G. 2017. Implementation opinions: the programmatic document for China’s new round of Gaokao reform. *China Examinations*. 2, pp.1-4.

Jiang, Y. 2015. Zhejiang jiaogai “xuanze” poti. *China Education Daily*. 22 December, p.1.

Johansson, S. 2016. International large-scale assessments: what uses, what consequences? *Educational Research*. **58**(2), pp.139-148.

Johnson, R.B. and Christensen, L. 2017. *Educational research: quantitative, qualitative, and mixed approaches*. 6th ed. London: Sage.

Johnson, R.E., Grove, A.L. and Clarke, A. 2019. Pillar integration process: a joint display technique to integrate data in mixed methods research. *Journal of Mixed Methods Research*. **13**(3), pp.301-320.

Kamens, D.H. 2013. Globalization and the emergence of an audit culture: PISA and the search for ‘best practices’ and magic bullets. In: Meyer, H. and Benavot, A. eds. *PISA, power, and policy: the emergence of global educational governance*. Oxford: Symposium Books, pp.117-139.

Keller, M.M., Neumann, K. and Fischer, H.E. 2017. The impact of physics teachers’ pedagogical content knowledge and motivation on students’ achievement and interest. *Journal of Research in Science Teaching*. **54**(5), pp.586-614.

Kelly, P. and Kotthoff, H.-G. 2017. PISA, National and regional education policies and their effect on mathematics teaching in England and Germany. *Research in Comparative and International Education*. **12**(3), pp.318-332.

Kim, J-S. Anderson, C.J. and Keller, B. 2014. Multilevel analysis of assessment data. In: Rutkowski, L. von Davier, M. and Rutkowski, D. eds. *Handbook of international large-scale assessment*. Boca Raton: CRC Press, pp.389-424.

King, G., Murray, C.J.L., Salomon, J.A. and Tandon, A. 2004. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*. **98**(1), pp.191-207.

King, G. and Wand, J. 2007. Comparing incomparable survey responses: evaluating and selecting anchoring vignettes. *Political Analysis*. **15**(1), pp.46-66.

Kirsch, I., Lennon, M., von Davier, M., Gonzalez, E. and Yamamoto, K. 2013. On the growing importance of international large-scale assessments In: M. von Davier, E. Gonzalez, I. Kirsch and K. Yamamoto, eds. *The role of international large-scale assessments:*

perspectives from technology, economy, and educational research. [Online]. Dordrecht: Springer Netherlands, pp. 1-11. [Accessed 11 November 2016]. Available from: http://link.springer.com/chapter/10.1007/978-94-007-4629-9_1

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S. and Baumert, J. 2013. Teachers' content knowledge and pedagogical content knowledge: the role of structural differences in teacher education. *Journal of Teacher Education*. **64**(1), pp.90-106.

Klieme, E., Schümer, G. and Knoll, S. 2001. Mathematikunterricht in der Sekundarstufe I: "Aufgabenkultur" und Unterrichtsgestaltung. In: Klieme, E. and Baumert, J. eds. *TIMSS - Impulse für Schule und Unterricht*. Bonn: Bundesministerium für Bildung und Forschung, pp.43-57.

Knodel, P., Martens, K. and Niemann, D. 2013. PISA as an ideational roadmap for policy change: exploring Germany and England in a comparative perspective. *Globalisation, Societies and Education*. **11**(3), pp.421-441.

Knodel, P. and Walkenhorst, H. 2010. What's England got to do with it? British underestimation of international initiatives in education policy. In: Martens, K., Alexander Kenneth, N., Windzio, M. and Weymann, A. eds. *Transformation of education policy*. Basingstoke: Palgrave, pp.132-152.

Kreiner, S. and Christensen, K.B. 2014. Analyses of model fit and robustness. A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika*. **79**(2), pp.210-231.

Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. and Hachfeld, A. 2013. Professional competence of teachers: effects on instructional quality and student development. *Journal of Educational Psychology*. **105**(3), pp.805-820.

Labaree, D.F. 2014. Let's measure what no one teaches: PISA, NCLB, and the shrinking aims of education. *Teachers College Record*. **116**(9), pp.1-14.

Lafontaine, D., Baye, A., Vieluf, S. and Monseur, C. 2015. Equity in opportunity-to-learn and achievement in reading: a secondary analysis of PISA 2009 data. *Studies in Educational Evaluation*. **47**, pp.1-11.

Lee, V.E. and Burkam, D.T. 2002. *Inequality at the starting gate: social background differences in achievement as children begin school*. Washington, D.C.: Economic Policy Institute.

Lee, J. and Stankov, L. 2013. Higher-order structure of noncognitive constructs and prediction of PISA 2003 mathematics achievement. *Learning and Individual Differences*. **26**, pp.119-130.

Lee, J. and Stankov, L. 2018. Non-cognitive predictors of academic achievement: evidence from TIMSS and PISA. *Learning and Individual Differences*. **65**, pp.50-64.

LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B.L., Bisanz, J., Kamawar, D. and Penner-Wilger, M. 2010. Pathways to mathematics: longitudinal predictors of performance. *Child Development*. **81**(6), pp.1753-1767.

Lepper, M.R., Corpus, J.H. and Iyengar, S.S. 2005. Intrinsic and extrinsic motivational orientations in the classroom: age differences and academic correlates. *Journal of Educational Psychology*. **97**(2), pp.184-196.

Leung, F.K.S. 2014. What can and should we learn from international studies of mathematics achievement? *Mathematics Education Research Journal*. **26**(3), pp.579-605.

Lewis, A.B. and Mayer, R.E. 1987. Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology*. **79**(4), pp.363-371.

Lewis, S. 2017. Governing schooling through "what works": the OECD's PISA for Schools. *Journal of Education Policy*. **32**(3), pp.281-302.

Lewis, S., Sellar, S. and Lingard, B. 2016. PISA for schools: topological rationality and new spaces of the OECD's global educational governance. *Comparative Education Review*. **60**(1), pp.27-57.

Li, Y., Xu, F. Zhao, J. and Cheng, L. 2019. The ideas and pathways of Gaokao content reform based on the development of key competency. *Curriculum, Teaching and Method*. **39**(7), pp.76-83.

Lincoln, Y.S. and Guba, E.G. 1985. *Naturalistic inquiry*. Newbury: Sage.

Lingard, B. and Lewis, S. 2017. Placing PISA and PISA for schools in two federalisms, Australia and the USA. *Critical Studies in Education*. **58**(3), pp.266-279.

Lisbon European Council. 2000. Presidency *conclusions*. [Online]. [Accessed 11 March 2020]. Available from: https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/00100-r1.en0.htm

Liu, B. 2014. Reform of test content since the resumption of College Entrance Examination. *Theory and Practice of Education*. **34**(16), pp.19-22.

Liu, B. 2018. Wei xuesheng de zhongshen fazhan dianding jichu: Jiedu putong gaozhong kecheng fang'an he kecheng biaozhun (2017). *China Education Daily*. [Online]. 17 January. [Accessed 8 January 2020]. Available from: http://paper.jyb.cn/zgjyb/html/2018-01/17/node_2.htm

Liu, K. 2015. Dui PISA tisheng xuesheng kexue suyang de sikao. *Teachers' Journal*. 10, pp.32-33.

Liu, L.2012. *Zai quanguo jichu jiaoyu zhiliang jiance gongzuohuiyishang de jianghua*. [Online]. 20 March, Quanguo jichu jiaoyu zhiliang jiance gongzuo huiyi. [Accessed 5 October 2018]. Available from: http://www.moe.gov.cn/s78/A11/s6353/201203/t20120326_133113.html

Liu, O.L. and Wilson, M. 2009. Gender differences in large-scale math assessments: PISA trend 2000 and 2003. *Applied Measurement in Education*. **22**(2), pp.164-184.

Liu, W. 2000. Cong Gaokao de jiazhi quxiang tan muqian de Gaokao gaige. *Contemporary Education Sciences*. 5, pp.44-46.

Liu, X. and Zhang, W. 1986. Biaozhunhua kaoshi yu Gaokao gaige. *People's Education*. 4, pp.19-20.

Liu, Y. 2017. Zai 2017 nian quanguo jiaoshu yuren kaimo ji youxiu jiaoshi daibai zuotanhuishang de jianghua. *China Education Daily*. 7 September, p.1.

Lockheed, M. 2013. Causes and consequences of international assessment in developing countries. In: Meyer, H-D. and Benavot, A. eds. *PISA, Power, and policy: the emergence of global educational governance*. Oxford: Symposium Books, pp.163-183.

Lockheed, M. E. and Wagemaker, H. 2013. International large-scale assessments: thermometers, whips or useful policy tools? *Research in Comparative and International Education*. **8**(3), pp.296-306.

Luttenberger, S., Wimmer, S. and Paechter, M. 2018. Spotlight on math anxiety. *Psychology Research and Behavior Management*. **11**, pp.311-32.

Maddox, B. and Addey, C. 2016.The futures and promises of international educational assessment. 13 August. *Laboratory of International Assessment Studies*. [Online]. [Accessed 3 March 2017]. Available from: http://international-assessments.org/the-futures-and-promises-of-international-educational-assessment/

Maguire, M., Braun, A. and Ball, S. 2015. "Where you stand depends on where you sit": the social construction of policy enactments in the (English) secondary school. *Discourse: Studies in the Cultural Politics of Education*. **36**(4), pp.485-499.

Marks, G.N. 2006. Are between- and within-school differences in student performance largely due to socio-economic background? Evidence from 30 countries. *Educational Research*. **48**(1), pp.21-40.

Marks, G.N. 2008. Accounting for the gender gaps in student performance in reading and mathematics: evidence from 31 countries. *Oxford Review of Education*. **34**(1), pp.89-109.

Marsh, H.W. and Hau, K.-T. 2003. Big-fish--little-pond effect on academic self-concept: a cross-cultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*. **58**(5), pp.364-376.

Marsh, H.W., Hau, K.-T., Artelt, C., Baumert, J. and Peschar, J.L. 2006. OECD's brief self-report measure of educational psychology's most useful affective constructs: cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*. **6**(4), pp.311-360.

Marsh, H.W. and Parker, J.W. 1984. Determinants of student self-concept: is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*. **47**(1), pp.213-231.

Marsh, H.W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K.-T., O'Mara, A.J. and Craven, R.G. 2008. The big-fish-little-pond-effect stands up to critical scrutiny: implications for theory, methodology, and future research. *Educational Psychology Review*. **20**(3), pp.319-350.

Martens, K. and Niemann, D. 2010. *Governance by comparison: How rating and ranking impacts national policy-making in education*. *TranState working paper*, No.139. University of Bremen, Collaborative Research Center 597: Transformations of the State.

Martin, M.O. and Mullis, I.V.S. eds. 2013. *TIMSS and PIRLS 2011: relationships among reading, mathematics, and science achievement at the fourth grade—Implications for early learning*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and IEA.

Massey, A., Elliott, G. and Ross, E. 1996. Season of birth, sex and success in GCSE English, mathematics and science: some long-lasting effects from the early years? *Research Papers in Education*. **11**(2), pp.129-150.

Mayer, R.E. 1986. Mathematics. In: Dillon, R. F. and Sternberg, R. J. eds. *Cognition and instruction*. London: Academic Press INC, pp. 127-154.

Mayer, R.E. 2004. Should there be a three-strikes rule against pure discovery learning? *American Psychologist*. **59**(1), pp.14-19.

McConney, A. and Perry, L.B. 2010. Science and mathematics achievement in Australia: the role of school socioeconomic composition in educational equity and effectiveness. *International Journal of Science and Mathematics Education*. **8**(3), pp.429-452.

McLaughlin, M.W. 1990. The rand change agent study revisited: macro perspectives and micro realities. *Educational Researcher*. **19**(9), pp.11-16.

McQueen, J. and Mendelovits, J. 2003. PISA reading: cultural equivalence in a cross-cultural study. *Language Testing*. **20**(2), pp.208-224.

Melkonian, M. and Areepattamannil, S. 2018. The effect of absolute age-position on academic performance: a study of secondary students in the United Arab Emirates. *Educational Studies*. **44**(5), pp.551-563.

Meyer, H.-D. and Benavot, A. eds. 2013. *PISA, power, and policy: the emergence of global educational governance*. Oxford: Symposium Books Ltd.

Michel, A. 2017. The contribution of PISA to the convergence of education policies in Europe. *European Journal of Education*. **52**(2), pp.206-216.

MOE. 1999. *Zhongxiaoxue jiaoshi jixu jiaoyu guiding*. [Online]. Beijing: MOE. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/srcsite/A02/s5911/moe_621/199909/t19990913_180474.html

MOE. 2001a. *Jichu jiaoyu kecheng gaige gangyao (shixing)*. [Online]. Beijing: MOE. [Accessed 8 January 2020]. Available from: http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_309/200412/4672.html

MOE. 2001b. Zhongxiaoxue jiaoxue dagang weihe gaicheng kecheng biaozhun? [Press release]. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/jyb_xwfb/xw_zllssj/moe_183/tnull_1717.html

MOE. 2002. *Jiaoyubu guanyu jiji tuijin zhongxiaoxue pingjia yu kaoshi zhidu gaige de tongzhi*. [Online]. Beijing: MOE. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/srcsite/A26/s7054/200212/t20021227_166074.html.

MOE. 2005. *Jiaoyubu guanyu jinyibu tuijin yiwu jiaoyu junheng fazhan de ruogan yijian*. [Online]. Beijing: MOE. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/moe_1485/tnull_7695.html

MOE. 2008. *Jiaoyubu guanyu shenru tuijin he jinyibu wanshan Zhongkao gaige de yijian*. [Online]. Beijing: MOE. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/srcsite/A26/s7054/200804/t20080403_78505.html

MOE. 2012. Jiaoyubu jichu jiaoyu kecheng jiaocai zhuanjia gongzuo weiyuanhui jiu yinfa yiwu jiaoyu kecheng biaozhun (2011) da jizhewen. [Press release]. [Accessed 23 June 2018]. Available from: http://www.moe.edu.cn/jyb_xwfb/s271/201202/t20120207_130063.html

MOE. 2013a. *Jiaoyubu guanyu tuijin zhongxiaoxue jiaoyu zhiliang zonghe pingjia gaige de yijian*. Beijing: MOE. [Accessed 8 January 2020]. Available from: http://old.moe.gov.cn//publicfiles/business/htmlfiles/moe/s7054/201306/153185.html

MOE.2013b. Kaizhan yi xuesheng quanmian fazhan wei hexin de lvse pingjia: jiaoyubu youguan fuzeren jiu *Guanyu tuijin zhongxiaoxue jiaoyu zhiliang zonghe pingjia gaige de yijian* da jizhewen. [Press release]. [Accessed 5 October 2018]. Available from: http://www.moe.gov.cn/jyb_xwfb/xw_fbh/moe_2069/s7135/s7396/s7398/201306/t20130618_153226.html

MOE. 2015. *Woguo yiwu jiaoyu zhikiang jiance jinzhan qingkuang*. [Online]. Beijing: MOE. [Accessed 5 October 2018]. Available from: http://www.moe.gov.cn/jyb_xwfb/xw_fbh/moe_2069/xwfbh_2015n/xwfb_150415/150415_sfcl/201504/t20150415_187152.html.

MOE. 2016. *Jiaoyubu guanyu jinyibu tuijin gaozhong jieduan xuexiao zhaosheng zhidu gaige de zhidao yijian*. [Online]. Beijing: MOE. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/srcsite/A06/s3732/201609/t20160920_281610.html

MOE. 2018. *Jiaoyubu yinfa putong gaozhong kecheng fang'an he kecheng biaozhun (2017), luoshi lideshuren genben renwu*. [Online]. Beijing: MOE. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/201801/t20180116_324668.html

MOE. 2019. *Jiaoyubu guanyu jiaqiang chuzhong xueye shuiping kaoshi mingti gongzuo de yijian*. [Online]. Beijing: MOE. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/srcsite/A06/s3321/201911/t20191128_409951.html

MOE. 2020. Jiang lideshuren rongru kaoshi pingjia quanguocheng: Jiaoyubu kaoshizhongxin fuzeren jiu *Zhongguo Gaokao Pingjia Tixi* da jizhewen. [Press release]. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/jyb_xwfb/s271/202001/t20200107_414613.html

Morris, M. 2015. Comparative education, PISA, politics and educational reform: a cautionary note. *Compare: A Journal of Comparative and International Education*. **45**(3), pp.470-474.

Morse, J.M. 1991. Approaches to qualitative-quantitative methodological triangulation. *Nursing Research*, **40**, pp.120-123.

Mullis, I.V.S., Martin, M.O. and Foy, P. 2013. The impact of reading ability on TIMSS mathematics and science achievement at the fourth grade: An analysis by item reading demands. In: Martin, M.O. and Mullis, I.V.S. eds. *TIMSS and PIRLS 2011: relationships among reading, mathematics, and science achievement at the fourth grade—Implications for early learning*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, pp. 67-108.

Mullis, I.V.S., Martin, M.O. Ruddock, G., O'Sullivan, C.Y. and Preuschoff, C. 2009. *TIMSS 2011 assessment frameworks*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Nakayasu, C. 2016. School curriculum in Japan. *The Curriculum Journal*. **27**(1), pp.134-150.

Neumann, K., Fischer, H.E. and Kauertz, A. 2010. From PISA to educational standards: the impact of large-scale assessments on science education in Germany. *International Journal of Science and Mathematics Education*. **8**(3), pp.545-563.

Ni, M. 2012. Fahui lvse zhebiao de duopan zuoyong. *Modern Teaching.* 3, p.1.

Niemann, D., Martens, K. and Teltemann, J. 2017. PISA and its consequences: shaping education policies through international comparisons. *European Journal of Education*. **52**(2), pp.175-183.

Niss, M. 2003. Mathematical competencies and the learning of mathematics: the Danish KOM project. In: Gagatsis, A. and Papastavridis, S. eds. *3rd mediterranean conference on mathematics education*, *January 2003, Athens, Hellas 3-4-5*. Athen: Hellenic Mathematical Society, pp. 115-124.

Niss, M. and Højgaard, T. eds. 2011. *Competencies and mathematical learning: ideas and inspiration for the development of mathematics teaching and learning in Denmark*. IMFUFA tekst, No. 485. Roskilde: Roskilde University.

Noyes, A. 2012. It matters which class you are in: student-centred teaching and the enjoyment of learning mathematics. *Research in Mathematics Education*. **14**(3), pp.273-290.

Oates, T. 2011. Could do better: using international comparisons to refine the National Curriculum in England. *The Curriculum Journal*. **22**(2), pp.121-150.

OECD. 1999. *Measuring student knowledge and skills: a new framework for assessment*. Paris: OECD Publishing.

OECD. 2002. *PISA 2000 technical report*. Paris: OECD Publishing.

OECD. 2004. *PISA 2003 results: learning for tomorrow's world*. Paris: OECD publishing.

OECD. 2005. *PISA 2003 technical report*. Paris: OECD Publishing.

OECD. 2006. *Education at a glance.* Paris: OECD Publishing.

OECD. 2009. *PISA 2006 technical report*. Paris: OECD Publishing.

OECD. 2010. *PISA 2009 results: what students know and can do – Student performance in reading, mathematics and science (Volume I)*. Paris: OECD Publishing.

OECD. 2012. *PISA 2009 technical report*. Paris: OECD Publishing.

OECD. 2013a. *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing.

OECD. 2013b. *PISA 2012 results: excellence through equity: giving every student the chance to succeed (Volume II)*. Paris: OECD Publishing.

OECD. 2013c. *PISA 2012 results: what makes schools successful? Resources, policies and practices (Volume IV)*. Paris: OECD Publishing.

OECD. 2013d. *PISA 2012 results: ready to learn: students' engagement, drive and self-beliefs (Volume III).* Paris: OECD Publishing.

OECD. 2014a. *PISA 2012 technical report*. Paris: OECD Publishing.

OECD. 2014b. *PISA 2012 results: what students know and can do – Student performance in mathematics, reading and science (Volume I, Revised edition)*. Paris: OECD Publishing.

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

OECD. 2016b. *PISA FAQ: understanding the results*. [Online]. [Accessed 29 November 2016]. Available from: http://www.oecd.org/pisa/pisafaq/

OECD. 2016c. *About the OECD: who does what*. [Online]. [Accessed 29 November 2016]. Available from: https://www.oecd.org/about/whodoeswhat/

OECD. 2016d. *PISA 2015 assessment and analytical framework: science, reading, mathematic and financial literacy*. Paris: OECD Publishing.

OECD. 2016e. *About PISA*. [Online]. [Accessed 12 November 2016]. Available from: http://www.oecd.org/pisa/aboutpisa/

OECD. 2017. *PISA 2015 technical report*. Paris: OECD Publishing.

OECD. 2019a. *PISA 2018 results (Volume I): what students know and can do*. Paris: OECD Publishing.

OECD. 2019b. *PISA 2018 assessment and analytical framework*. Paris: OECD Publishing.

OECD. 2019c*. PISA website*. [Online]. [Accessed 24 September 2019]. Available from http://www.oecd.org/pisa/

Pajares, F. and Miller, M.D. 1994. Role of self-efficacy and self-concept beliefs in mathematical problem solving: a path analysis. *Journal of Educational Psychology*. **86**(2), pp.193-203.

Patton, M.Q. 2002. *Qualitative research & evaluation methods*. 3rd ed. Thousand Oaks: Sage.

Pehkonen, J., Viinikainen, J., Böckerman, P., Pulkki-Råback, L., Keltikangas-Järvinen, L. and Raitakari, O. 2015. Relative age at school entry, school performance and long-term labour market outcomes. *Applied Economics Letters*. **22**(16), pp.1345-1348.

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

Phillips, D. and Ochs, K. 2003. Processes of policy borrowing in education: some explanatory and analytical devices. *Comparative Education*. **39**(4), pp.451-461.

Pons, X. 2012. Going beyond the “PISA shock” discourse: an analysis of the cognitive reception of PISA in six European countries, 2001-2008. *European Educational Research Journal*. **11**(2), pp.206-226.

Pons, X. 2016. Tracing the French policy PISA debate: a policy configuration approach. *European Educational Research Journal*. **15**(5), pp.580-597.

Praetorius, A.-K., Klieme, E., Herbert, B. and Pinger, P. 2018. Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *ZDM*. **50**(3), pp.407-426.

Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K. and Klieme, E. 2014. One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*. **31**, pp.2-12.

Prais, S.J. 2003. Cautions on OECD’s recent educational survey (PISA). *Oxford Review of Education*. **29**(2), pp.139-163.

Purves, A.C. 1987. The evolution of the IEA: a memoir. *Comparative Education Review*. **31**(1), pp.10-28.

Rasch, G. 1960. *Probabilistic models for some intelligence and attainment tests*. Expanded edition. Chicago: University of Chicago Press.

Rautalin, M., Alasuutari, P. and Vento, E. 2019. Globalisation of education policies: does PISA have an effect? *Journal of Education Policy*. **34**(4), pp.500-522.

Reeve, J. 2012. A self-determination theory perspective on student engagement. In: Christenson, S.L., Reschly, A.L. and Wylie, C. eds. *Handbook of student engagement*. New York: Springer, pp.149-172.

Ren, Z. and Chen, A. 2013. Research on ability test and content reform in College Entrance Examination. *Curriculum, Teaching Material and Method*. **33**(10), pp.88-93.

Riley, K., Torrance, H. 2003. Big change question: as national policy-makers seek to find solutions to national education issues, do international comparisons such as TIMMS and PISA create a wider understanding, or do they serve to promote the orthodoxies of international agencies?. *Journal of Educational Change*. **4**, pp.419-425.

Ringarp, J. 2016. PISA lends legitimacy: a study of education policy changes in Germany and Sweden after 2000. *European Educational Research Journal*. **15**(4), pp.447-461.

Robertson, R. 1995. Glocalization: time-space and homogeneity-heterogeneity. In: Featherstone, M. and Robertson, R. eds. *Global modernities*. London: Sage, pp.25-44.

Robson, C. 1993. *Real world research*. Oxford: Blackwell.

Robson, C. 2011. *Real world research*. 3rd ed. Chicherster: Wiley.

Rubin, D. B. 1987. *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rutkowski, D. 2015. The OECD and the local: PISA-based Test for Schools in the USA. *Discourse: Studies in the Cultural Politics of Education*. **36**(5), pp.683-699.

Rutkowski, D. and Delandshere, G. 2016. Causal inferences with large scale assessment data: using a validity framework. *Large-scale Assessments in Education*. **4**(1), p.6.

Rutkowski, L. and Rutkowski, D. 2016. A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher*. **45**(4), pp.252-257.

Rutkowski, L., Rutkowski, D. and Zhou, Y. 2016. Item calibration samples and the stability of achievement estimates and system rankings: another look at the PISA model. *International Journal of Testing*. **16**(1), pp.1-20.

Rutkowski, L. and Zhou, Y. 2014. Using structural equation models to analyze ILSA Data. In: Rutkowski, L., von Davier, M. and Rutkowski, D. eds. *Handbook of international large-scale assessment*. Boca Raton: CRC Press, pp.425-449.

Ryan, R.M. and Deci, E.L. 2000. Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology*. **25**(1), pp.54-67.

Ryan, R.M. and Deci, E.L. 2009. promoting self-determined school engagement: motivation, learning and well-being. In: Wentzel, K.R. and Wigfield, A. eds. *Handbook of motivation at school*. New York: Taylor Francis, pp.171-196.

Şahin, M.G. and Yıldırım, Y. 2016. The examination of the variables affecting mathematics behavior and mathematics literacy by multi-group hybrid model in the sample of PISA 2012 Turkey. *Education and Science*. **41**(187), pp.181-198.

Schatz, M., Popovic, A. and Dervin, F. 2017. From PISA to national branding: exploring Finnish education®. *Discourse: Studies in the Cultural Politics of Education*. **38**(2), pp.172-184.

Scheerens, J. 1990. School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement*. **1**(1), pp.61-80.

Scheerens, J. and Bosker, R. 1997. *The foundations of educational effectiveness*. Oxford: Pergamon.

Scherer, R., Nilsen, T. and Jansen, M. 2016. Evaluating individual students' perceptions of instructional quality: An investigation of their factor structure, measurement invariance, and relations to educational outcomes. *Frontiers in Psychology*. **7**, article no: 110 [no pagination].

Schiefele, U. 2009. Situational and individual interest. In: Wentzel, K.R. and Wigfield, A. eds. *Handbook of motivation at school*. New York: Routledge, pp.197-222.

Schleicher, A. and Zoido, P. 2016. The policies that shaped PISA, and the policies that PISA shaped. In: Mundy, K., Green, A., Lingard, B. and Verger, A. eds. *Handbook of global education policy*. Chichester: Wiley, pp.374-384.

Schmidt, W.H. and Burroughs, N.A. 2015. Puzzling out PISA: what can international comparisons tell us about American education? *American Educator*. **39**(1), pp.24-31.

Schmidt, W.H. and Maier, A. 2009. Opportunity to learn. In: Sykes, G., Schneider, B. and Plank. D. N. eds. *Handbook of education policy research*. New York: Routledge, pp.541-559.

Schmidt, W.H., McKnight, C.C., Houang, R.T., Wang, H., Wiley, D.E., Cogan, L.S. and Wolfe, R.G. 2001. *Why schools matter: a cross-national comparison of curriculum and learning.* San Francisco: Jossey-Bass.

Schmidt, W.H., Zoido, P. and Cogan, L. 2013. *Schooling matters: opportunity to learn in PISA 2012*. *OECD Education Working Papers*, No.95. Paris: OECD Publishing.

Schunk, D.H. and Pajares, F. 2005. Competence perceptions and academic functioning. In: Elliot, A.J. and Dweck, C.S. eds. *Handbook of competence and motivation*. New York: Guilford Publications, pp.85-104.

Schunk, D.H. and Pajares, F. 2009. Self-efficacy theory. In: Wentzel, K. R. and Wigfield, A. eds. *Handbook of motivation at school*. New York: Routledge, pp.35-53.

Schweisfurth, M. 2013. *Learner-centred education: whose pedagogy for whose development?* New York: Routledge.

Schwille, J. et al. 1982. *Teachers as policy brokers in the content of elementary school mathematics*. *Research Series*, No. 113. East Lansing, Michigan: Institute for Research on Teaching, Michigan State University.

Scott, D. 2014. Curriculum standards: performance and competence. *European Journal of Curriculum Studies*. **1**(1), pp.5-14.

Seaton, M., Marsh, H.W., Yeung, A.S. and Craven, R. 2011. The big fish down under: examining moderators of the 'Big-Fish-Little-Pond' effect for Australia's high achievers.*Australian Journal of Education*. **55**(2), pp.93-114.

Sellar, S. and Lingard, B. 2013a. Looking East: Shanghai, PISA 2009 and the reconstitution of reference societies in the global education policy field. *Comparative Education*. **49**(4), pp.464-485.

Sellar, S and Lingard, B. 2013b. The OECD and global governance in education. *Journal of Education Policy*. **28**(5), pp.710-725.

Sellar, S. and Lingard, B. 2014. The OECD and the expansion of PISA: new global modes of governance in education. *British Educational Research Journal*. **40**(6), pp.917-936.

Sellar, S., Thompson, G. and Rutkowski, D. 2017. *The global education race: taking the measure of PISA and international testing*. Canada: Brush Education Inc.

Shanghai Academy of Educational Sciences.2012. *Zhongguo-Finland PISA yantaohui—— PISA results: jichu jiaoyu zhong de jingyan yu tiaozhan*. [Online]. [Accessed 15 October 2018]. Available from: http://www.cnsaes.org/homepage/html/researchnews/7320.html

Shi, Z. and Zhang, X. 2008. Education reform: experience from China. *Journal of Beijing Normal University (Social Sciences)*. 5, pp.22-32.

Shih, C-M. 2009. How tests change teaching: a model for reference. *English Teaching: Practice and Critique*. **8**(2), pp.188-206.

Shin, J., Lee, H. and Kim, Y. 2009. Student and school factors affecting mathematics achievement: international comparisons between Korea, Japan and the USA. *School Psychology International*. **30**(5), pp.520-537.

Shohamy, E. 1992. Beyond proficiency testing: a diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal*. **76**(4), pp.513-521.

SHPISA, 2010. *Quality and equity: research report of the Programme for International Student Assessment 2009 in Shanghai, China*. Shanghai: Shanghai Education Publishing House.

SHPISA, 2013. *Quality and equity: research report of the Programme for International Student Assessment 2012 in Shanghai, China*. Shanghai: Shanghai Education Publishing House.

Shulman, L. S. 1986. Those who understand: Knowledge growth in teaching. *Educational Researcher*. **15**(2), pp.4-14.

Shulman, L. S. 1987. Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*. **57**(1), pp.1-22.

Sirin, S.R. 2005. Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research*. **75**(3), pp.417-453.

Sjøberg, S. 2015. PISA and global educational governance – A critique of the project, its uses and implications. *Eurasia Journal of Mathematics, Science & Technology Education*. **11**(1), pp.111-127.

Song, N. 2015. Woguo yiwu jiaoyu zhiliang jiance de “Sige chuangxin”. [Press release]. [Accessed 7 May 2017]. Available from: http://www.moe.gov.cn/jyb_xwfb/xw_fbh/moe_2069/xwfbh_2015n/xwfb_150415/150415_sfcl/201504/t20150415_187153.html

Spillane, J.P. 1999. External reform initiatives and teachers’ efforts to reconstruct their practice: the mediating role of teachers’ zones of enactment. *Journal of Curriculum Studies*. **31**(2), pp.143-175.

Sprietsma, M. 2010. Effect of relative age in the first grade of primary school on long-term scholastic results: international comparative evidence using PISA 2003. *Education Economics*. **18**(1), pp.1-32.

StataCorp. 2013. *Stata Statistical Software* (Version 13). [Software]. [Accessed 1 October 2016].

State Council. 1993. *Zhongguo jiaoyu gaige he fazhan gangyao*. [Online]. Beijing: State Council. [Accessed 8 January 2020]. Available from: http://old.moe.gov.cn/publicfiles/business/htmlfiles/moe/moe_177/200407/2484.html

State Council. 1994. *Zhonggong zhongyang guanyu jinyibu jiaqiang he gaijin xuexiao deyu gongzuo de ruogan yijian*. [Online]. Beijing: State Council. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/jyb_sjzl/moe_177/tnull_2479.html

State Council. 1999. *Zhonggong zhongyang guowuyuan guanyu shenhua jiaoyu gaige, quanmian tuijin suzhi jiaoyu de jueding*. [Online]. Beijing: State Council. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/jyb_sjzl/moe_177/tnull_2478.html

State Council. 2010. *Guojia zhongchangqi jiaoyugage he fazhan guihua gangyao (2010-2020)*. [Online]. Beijing: State Council. [Accessed 8 January 2020]. Available from:http://www.moe.gov.cn/srcsite/A01/s7048/201007/t20100729_171904.html

State Council. 2014. *Guowuyuan guanyu shenhua kaoshi zhaosheng zhidu gaige de shishi yijian*. [Online]. Beijing: State Council. [Accessed 8 January 2020]. Available from: http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/201409/t20140904_174543.html

Steiner-Khamsi, G. 2003. The politics of league tables. *JSSE - Journal of Social Science Education*. [Online]. **2**(1). [Accessed 19 September 2018]. Available from: http://www.jsse.org/index.php/jsse/article/view/470.

Steiner-Khamsi, G. and Waldow, F. eds. 2012. *World yearbook of education 2012: Policy borrowing and lending in education*. NY: Routledge.

Stevens, T., Olivarez, A., Lan, W.Y. and Tallent-Runnels, M.K. 2004. Role of mathematics self-efficacy and motivation in mathematics performance across ethnicity. *The Journal of Educational Research*. **97**(4), pp.208–222.

Stoet, G. and Geary, D.C. 2013. Sex differences in mathematics and reading achievement are inversely related: within- and across-nation assessment of 10 years of PISA data. *PLOS ONE*. **8**(3), p.e57988.

Sullivan, K., McConney, A. and Perry, L.B. 2018. A comparison of rural educational disadvantage in Australia, Canada, and New Zealand using OECD’s PISA. *SAGE Open*. **8**(4), p.2158244018805791.

Suter, L.E. 2017. How international studies contributed to educational theory and methods through measurement of opportunity to learn mathematics: *Research in Comparative and International Education*.

Takayama, K. 2008. The politics of international league tables: PISA in Japan’s achievement crisis debate. *Comparative Education*, **44**(4), pp.387-407.

Tao, J. and Wang, Z. 2014. Fangshan jiaoyu dingceng sheji yu tixi goujian – quyu jiaoyu fazhan de “Fangshan moshi”, *People's Tribune*, **6**(2), pp.68-71.

Tashakkori, A. and Teddlie, C. eds. 2010. *SAGE Handbook of mixed methods in social & behavioral research*. 2nd ed. Los Angeles: SAGE.

Teddlie, C. and Johnson, R.B. 2009. Methodological thought since the 20th century. In: Teddlie, C. and Tashakkori, A. eds. *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences*. 2nd ed. Los Angeles: SAGE, pp.62-82.

Teltemann, J. and Klieme, E. 2016. The impact of international testing projects on policy and practice. In: Brown, G.T.L. and Harris, L.R. eds. *Handbook of human and social conditions in assessment*. New York: Routledge, pp. 369-386.

Tikly, L. 2015. What works, for whom, and in what circumstances? Towards a critical realist understanding of learning in international and comparative education. *International Journal of Educational Development*. **40**, pp.237-249.

Törnroos, J. 2005. Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation*. **31**(4), pp.315-327.

Tucker, M., ed. 2011. *Surpassing Shanghai: An agenda for American education built on the world's leading systems*. Cambridge: Harvard Education Press.

UNESCO. 2019. Data for the sustainable development goals. [Online]. [Accessed 21 November 2019]. Available from: http://uis.unesco.org/en/country/cn

van Buuren, S., H. C. Boshuizen, and D. L. Knook. 1999. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine.* **18**(6), pp.681-694.

van der Schoot, M., Bakker Arkema, A.H., Horsley, T.M. and van Lieshout, E.C.D.M. 2009. The consistency effect depends on markedness in less successful but not successful problem solvers: an eye movement study in primary school children. *Contemporary Educational Psychology*. **34**(1), pp.58-66.

Vestheim, O.P., Husby, M., Aune, T.K., Bjerkeset, O. and Dalen, T. 2019. A population study of relative age effects on national tests in reading literacy. *Frontiers in Psychology*. **10**, article no: 1761 [pagination].

Vista, A. 2013. The role of reading comprehension in maths achievement growth: investigating the magnitude and mechanism of the mediating effect on maths achievement in Australian classrooms. *International Journal of Educational Research*. **62**, pp.21-35.

Waldow, F. 2017. Projecting images of the 'good' and the 'bad school': top scorers in educational large-scale assessments as reference societies. *Compare: A Journal of Comparative and International Education*. **47**(5), pp.647-664.

Waldow, F., Takayama, K. and Sung, Y. 2014. Rethinking the pattern of external policy referencing: media discourses over the 'Asian Tigers'' PISA success in Australia,Germany and South Korea. *Comparative Education*. **50**(3), pp.302-321.

Walkington, C., Clinton, V. and Shivraj, P. 2017. How readability factors are differentially associated with performance for students of different backgrounds when solving mathematics word problems. *American Educational Research Journal*. **55**(2), pp.362-414.

Wall, D. 1997. Impact and washback in language testing. In: Clapham, C. and Corson, D. eds. *Language testing and assessment*. Amsterdam: Kluwer Academic Publishers, pp.291-302.

Walpole, M. 2003. Socioeconomic status and college: how SES affects college experiences and outcomes. *Review of Higher Education*. **27**(1), pp.45-73.

Wang, C. 2015. PISA dui shuxue jiaoyu jiaoxue de qishi. *Teachers' Journal*. 10, pp.25-26.

Wang, L. 2007. Woguo daguimo jiaoyu pingjia xiangmu tanjiu yu shijian. *Educational Science Research*. 11, pp.25-28.

Wang, L. 2009. China large-scale education assessment reform: lessons learned from PISA China Trial research. *China Examinations*. 5, pp.17-25.

Wang, L. and Jing, A. 2013. Women cong PISA xuedaole shenme: jiyu PISA zhongguo shice de yanjiu. *Peking University Education Review*. 1, pp.172-180.

Wang, L., Jing, A. and Tong, W. 2017. Transition from PISA China Trial to education evaluation system with Chinese characteristics: based on PISA2009 China Trial. *Educational Research (China)*. **38**(1), pp.114-123.

Wang, L. and Tong, W. 2015. Competency assessment: implications of item development in PISA for College Entrance Examination. *Tsinghua Journal of Education*. 6, pp.93-98.

Wang, L. and Tong, W. 2016. Cong PISA bianma kan Gaokao wangshang yuejuan. *Educational Science Research*. 4, pp.33-38+45.

Westbrook, J., Durrani, N., Brown, R., Orr, D., Pryor, J., Boddy, J. and Salvi, F. 2013. *Pedagogy, curriculum, teaching practices and teacher education in developing countries. Education Rigorous Literature Review*. London: Department for International Development.

White, K.R. 1982. The relation between socioeconomic status and academic achievement. *Psychological Bulletin*. **91**(3), pp.461-481.

Wigfield, A. and Eccles, J.S. 2000. Expectancy–Value theory of achievement motivation. *Contemporary Educational Psychology*. **25**(1), pp.68-81.

Wu, K. 2008. On the special arenas restricting educational reform in China. *Educational Research (China)*. 12, pp.16-20.

Wu, M. 2009. A comparison of PISA and TIMSS 2003 achievement results in mathematics. *Prospects*. **39**(1), pp.33.

Wu, M. 2010. *Comparing the similarities and differences of PISA 2003 and TIMSS*. *OECD Education Working Papers*, No.32. Paris: OECD Publishing.

Wu, M. 2014. Evidence-based policy making in education. *International Journal of Contemporary Educational Research*. **1**(1), pp.1-8.

Wu, Y. Z. 2015. Beijing Fangshan: jiyu PISA ceshi de quyu guanli biange. *Zhongxiaoxue Guanli*, 8, pp.13-16.

Xia, X. 2004. Xin kegai huhuan xin Gaokao. *China Examinations*. 10, pp.4-7.

Yang, D.-C. and Sianturi, I.A. 2017. An analysis of Singaporean versus Indonesian textbooks based on trigonometry content. *Eurasia Journal of Mathematics, Science and Technology Education*. **13**(7), pp.3829-3848.

Yang Hansen, K. and Strietholt, R. 2018. Does schooling actually perpetuate educational inequality in mathematics performance? A validity question on the measures of opportunity to learn in PISA. *ZDM*. **50**(4), pp.643-658.

Yang, J. 2012. Fundamental spirit of curriculum criterion and instruction expectancy. *Jiangsu Education Research*. 10A, pp.3-9.

Yang, Y. ed. 2008. *China education examinations yearbook (2007)*. Beijing: Communication University of China Press.

Yang, Y. ed. 2014. *China education examinations yearbook (2013)*. Beijing: Communication University of China Press.

Yang, Y. 2016. Jiejian PISA jingyan de xueye kaoshi mingti ji fenxi. *Journal of Mathematics (China)*. 6, pp.9-11.

Yang, X. 1989. Gaokao de chengjiu ji qushi. *People's Education*. 3, pp.9-10.

Ye, L. 2017. Gaokao feilianxu wenben demingti xianzhuang he ceshi kongjian. *Yuwen Jianshe*. 6, pp.52-57.

Yi, H.S. and Lee, Y. 2017. A latent profile analysis and structural equation modeling of the instructional quality of mathematics classrooms based on the PISA 2012 results of Korea and Singapore. *Asia Pacific Education Review*. **18**(1), pp.23-39.

Yin, H. 2014. Shanghai "Lvse zhibiao": jiaoyu zhiliang zonghe pingjia gaige de sikao he shijian. *Jichu Jiaoyu Kecheng*. **127**(4), pp.6-21.

Yin, H., Xu, D., Ji, M., Wang, M., Zhu, L., Zhou, Z. and Liu, Z. 2012. Secure the health and happy growth of students: the reform and practice of Shanghai Green Indicators for academic performance of primary and secondary school students. *Educational Policy Observatory*. 5, pp.105-154.

Yore, L.D., Anderson, J.O. and Chiu, M.-H. 2010. Moving PISA results into the policy arena: perspectives on knowledge transfer for future considerations and preparations. *International Journal of Science and Mathematics Education*. **8**(3), pp.593-609.

Yu, H. 2017. A historical review on 40 years' content reform of the College Entrance Examination after its resumption. *China Examinations*, 3, pp. 1-5.

Yu, H., Lai, P. and Shi, J. 2017. Zhongguo de jiaoyu zixin zai nair?. *China Education Daily*. 3 March, p.1.

Yvonne Feilzer, M. 2010. Doing mixed methods research pragmatically: implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*. **4**(1), pp.6–16.

Zhao, J. 2017. Connotation evolution of reading ability and strategies of Chinese subject for College Entrance Examination: pectination and problem analysis about measuring course of Chinese reading ability in College Entrance Examination for 40 years. *Curriculum, Teaching Material and Method.* 12, pp.43–49.

Zhang, J. and Liang, Y. 2015. PISA yanjiu zhutui shuxue jiaoyu biange. *Teachers' Journal*. 10, pp.20-24.

Zhang, M. and Huang, H. 2016. Self-confidence, self-reflection, self-consciousness: PISA 2012 maths test and characters of Shanghai maths education. *Educational Research (China)*. **37**(1), pp.35-46.

Zhang, M. and Yan, W. 2015. Secret of Chinese mathematics education in the eye of English teachers – an interview with Professor Zhang Minxuan, Dean of the Research Institute of International and Comparative Education at Shanghai Normal University. *Primary & Secondary Schooling Abroad*. 1, pp.1-5.

Zhang, X. 2018. PISA shi chenggong jiaoyu de shijinshi ma? *Ethnic Education of China*, 2, pp.16.Zhang, Z. 2016. Yingyu zhongkao dui chuzhong yingyu jiaoxue de fanbo zuoyong. *Teaching Research for Primary and Middle Schools*. 6, pp.41-44+57.

Zheng, R. 2002. Shixi Gaokao de zhihuibang zuoyong. *Journal of Xiamen University(Arts & Social Sciences)*. 2, pp.7-10.

Zhou, Y., Fan, X., Wei, X. and Tai, R.H. 2017. Gender gap among high achievers in math and implications for STEM Pipeline. *The Asia-Pacific Education Researcher*. **26**(5), pp.259-269.

Zhu, Y., Kaiser, G. and Cai, J. 2018. Gender equity in mathematical achievement: the case of China. *Educational Studies in Mathematics*. **99**(3), pp.245-260.

Appendices

A.1 Interview guides

Part I To collect data about Fangshan's participation in PISA, PISA-motivated initiatives, the general picture of local educational policymakers' perceived impact of PISA on students' learning

Interviewee: an educational official of Fangshan Education Commission who is responsible for the administration of the three cycles of PISA and the research on translating PISA outputs in informing policies and practices in Fangshan

1. Could you tell me something about the background of Fangshan's participation in PISA 2009 China Trial?
   (General start, and probe the educational context of pre-participation, and the aim)

2. What about the participation in PISA 2012 China Trial and PISA 2015?
   (Probe the context and the aim, and the difference)

3. What did you identify from Fangshan PISA results?
   (Probe the results of each PISA cycle, and the trends)

4. In what ways did Fangshan do to improve students' learning?
   (Probe details and "how")

5. What do you think about the effect of these efforts on improving students' learning?
   (Probe supporting details)

Part II To collect data about what the projects did, how PISA outputs (e.g. framework, results) were used and translated in these projects' initiatives, project key informants' perceptions of the impact of PISA on students' reading/mathematics/science learning

Interviewee: the reading project leader

1. Could you introduce the initiation of the reading project?

2. Could you give me a general description about how reading was instructed in school before 2009?

3. What do you think about students' previous reading performance?

4. What factors do you think would affect students' previous reading performance?
   (Probe into "why")

5. What does the project have done or will it do to improve students' reading learning?
   (Probe details, "why" and "how")

6. What do you think about the effect of what the project did?
   (Probe supporting details)

Interviewee: the mathematics project leader

1. Could you introduce the initiation of the mathematics project?
2. What do you think about students' mathematics performance in PISA 2012 China Trial?
3. What about the performance in PISA 2015 compared with PISA 2012?
4. What does the project have done or will it do to improve students' mathematics learning?
   (Probe details, "why" and "how")
5. What do you think about the effect of what the project did?
   (Probe supporting details)

Interviewee: the science project assistant

1. Could you introduce the initiation of the mathematics project?
2. What do you think about students' science performance in PISA 2012 China Trial?
3. What about the performance in PISA 2015 compared with PISA 2012?
4. What does the project have done or will it do to improve students' mathematics learning?
   (Probe details, "why" and "how")
5. What do you think about the effect of what the project did?
   (Probe supporting details)

Part III To investigate the school enactment of the PISA-motivated initiatives, and school-level practitioners' perceptions of these initiatives' impact on students' learning at schools

Interviewees: school leaders

1. Could you briefly introduce your school?
2. Could you describe the mathematics teaching and learning in your school?
3. Has your school done anything or been involved in any local initiatives related to facilitate students' mathematics learning?
   (Probe "how")

4. What about the teaching and learning of reading and science? (Probe what has the school done or been involved in for improving students' reading and science learning) 5. What change of students' learning have you seen brought by these efforts? (Probe supporting details, particularly details about mathematics) 6. What other initiatives which are launched by Fangshan Education Commission are carried out in your school?
Part IV To investigate mathematics teachers' awareness of the PISA-motivated mathematics initiatives and the enactment in teaching practices, and mathematics teachers' perceptions of these initiatives' impact on students' mathematics learning at schools
Interviewees: mathematics teachers 1. Could you describe your mathematics instruction in your classroom? (Probe instruction design and teaching practices) 2. What did you do in terms of improving students' mathematics learning? 3. In your class, what change of students' mathematics learning have you seen brought by these efforts? 4. What professional training did you attend in recent years? 5. What do you think about them in terms of improving students' mathematics learning? 6. What would you recommend about the training or other ways to improve students' mathematics learning?

A.2 Phone scripts of inviting potential participants

For the local educational official

Hello, Mr/Ms ***. My name is Huiming Ding. I am a PhD student of the School of Education at University of Leeds in the United Kingdom, and previously worked in China PISA 2015 National Centre. My proposed PhD project is “The Impact of PISA on Students’ Learning: a Chinese Perspective”. Since Fangshan has rich experience in participating in PISA and utilizing its instrument and results in local educational practice, my research will be undertaken with a case study focusing on Fangshan. Your rich knowledge about the PISA-related work and initiatives in Fangshan could significantly contribute to the understanding of the research topic. I wonder if I could invite you to participate in a 1-hour interview for this research.

If it is convenient, could you please tell me your email address? I could then send you an information sheet describing in detail about my research aim, what will be involved in the interview, and how will I use and protect your data.

I plan to collect data in the workplace during working days in September and October. If you are willing to participate, you could select a date and time which is convenient for you. If you decide to participate or would like to know more information, please contact me on this number. I would appreciate if you kindly agree to participate. Thank you for your time.

For the team leaders of PISA-driven projects

Hello, Mr/Ms ***. My name is Huiming Ding. I am a PhD student of the School of Education at University of Leeds in the United Kingdom, and previously worked in China PISA 2015 National Centre. My proposed PhD project is “The Impact of PISA on Students’ Learning: a Chinese Perspective”. Since Fangshan has rich experience in participating in PISA and utilizing its instrument and results in local educational practice, my research will be undertaken with a case study focusing on Fangshan. Your rich knowledge about the PISA-related work and initiatives in Fangshan could significantly contribute to the understanding of the research topic. I wonder if I could invite you to participate in a 40-minute interview for this research.

If it is convenient, could you please tell me your email address? I could then send you an information sheet describing in detail about my research aim, what will be involved in the interview, and how will I use and protect your data.

I plan to collect data in the workplace during working days in September and October. If you are willing to participate, you could select a date and time which is convenient for you. If you decide to participate or would like to know more information, please contact me on this number. I would appreciate if you kindly agree to participate. Thank you for your time.

For head teachers

Hello, Mr/Ms ***. My name is Huiming Ding. I am a PhD student of the School of Education at University of Leeds in the United Kingdom, and previously worked in China PISA 2015 National Centre. You may still remember that your school and all the other secondary schools of Fangshan took part in PISA 2015. My proposed PhD project “The Impact of PISA on Students’ Learning: a Chinese Perspective” will be undertaken with a case study focusing on Fangshan. Your rich knowledge about schooling practice is important for investigating the actual influence of PISA-driven initiatives at school level. I wonder if I could invite you to participate in a 40-minute interview for this research.

If it is convenient, could you please tell me your email address? I could then send you an information sheet describing in detail about my research aim, what will be involved in the interview, and how will I use and protect your data.

I plan to collect data in the workplace during working days in September and October. If you are willing to participate, you could select a date and time which is convenient for you. If you decide to participate or would like to know more information, please contact me on this number. I would appreciate if you kindly agree to participate. Thank you for your time.

For mathematics teachers

Hello, Mr/Ms ***. My name is Huiming Ding. I am a PhD student of the School of Education at University of Leeds in the United Kingdom, and previously worked in China PISA 2015 National Centre. You might also know that your school took part in PISA 2015 together with all the other secondary schools of Fangshan. My proposed PhD project “The Impact of PISA on Students’ Learning: a Chinese Perspective” will be undertaken with a case study focusing on Fangshan. Your rich experience in teaching mathematics and full understanding about the students is important for investigating the actual influence of PISA-driven initiatives on students’ mathematics learning from the view of classroom instruction. I wonder if I could invite you to participate in a 40-minute interview for this research.

If it is convenient, could you please tell me your email address? I could then send you an information sheet describing in detail about my research aim, what will be involved in the interview, and how will I use and protect your data.

I plan to collect data in the workplace during working days in September and October. If you are willing to participate, you could select a date and time which is convenient for you. If you decide to participate or would like to know more information, please contact me on this number. I would appreciate if you kindly agree to participate. Thank you for your time.

UNIVERSITY OF LEEDS

A.3 Letter seeking permission from school gatekeepers

Huiming Ding
School of Education
University of Leeds
Leeds, LS2 9JT
UK

School head teacher
Fangshan District of Beijing
China
22.05.2017

Request for permission to conduct research in the school

Dear Sir/Madam,
My name is Huiming Ding. I am a PhD student of School of Education at University of Leeds in the UK. The proposed title of my PhD research is “The Impact of PISA on Students’ Learning: a Chinese Perspective”. It will be undertaken with a case study focusing on Fangshan. I am writing to ask if it would be possible to recruit one participant from the mathematics teachers who have rich experience (more than 5 years) in teaching grade 9 (or grade 10) students mathematics in your school for this research.
The research aims to investigate the perceived impact of PISA on students’ learning and specifically explore to what extent students’ mathematics learning outcomes have been influenced by the initiatives driven in light of PISA. The rich experience of the participant could contribute to the understanding of the actual influence of PISA-driven initiatives on students’ mathematics learning from the view of class instruction. I plan to interview the participant face-to-face for 40 minutes. Your school name and the participant’s name will be anonymised and will not be disclosed in the outcomes of the research. I have prepared a description of the research and what will be involved for the participant, and I have attached a copy for you to read (please see “Information sheet for mathematics teachers”).

I would like to carry out the interview between September and October, but I am happy to follow your advice on this. I would anticipate that the interview will be done during the break in working days in a quiet place in your school, and there should be no disturbance to teaching. If you wish to know any further information, please feel free to contact me.

Many thanks for reading this letter and looking forward to obtaining your support.

Yours faithfully,

Huiming Ding
PhD student
University of Leeds
Email: edhd@leeds.ac.uk
Phone: (+86) 15210980641

Dr Matt Homer
Supervisor
University of Leeds
Email: M.S.Homer@leeds.ac.uk
Tel: (+44) 0113 3434654

A.4 Background questionnaire for mathematics teachers (English version)

Thank you for participating in this doctoral research project “The Impact of PISA on Students’ Learning: a Chinese Perspective”. Before the start of the interview, please briefly answer the following questions.

1. How old are you?

☐26-30 ☐31-35 ☐36-40 ☐41-45 ☐46-50 ☐above 50

2. When did you start teaching mathematics in this school?

____

3. How many years have you taught students in 9th grade / 10th grade in this school?

____

4. Your academic major is

□mathematics □other (please specify)

____

UNIVERSITY OF LEEDS

A.5 Information sheet for the local educational official (English version)

Information Sheet

The title of the research project

The Impact of PISA on Students’ Learning: a Chinese Perspective

Invitation

You are being invited to take part in a research project. Before you decide whether or not to participate, it is important for you to understand the aim of the research and what your participation will involve. Please take time to read the following information carefully. Please ask me if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

The purpose of the project

This is a PhD research project, which is supported by the University of Leeds in the United Kingdom. The project aims to investigate the perceived impact of PISA on students’ learning and specifically explore to what extent students’ mathematics learning outcomes have been influenced by the initiatives driven in light of PISA. Since Fangshan District of Beijing has rich experience in participating in PISA and making use of PISA instrument as well as results, this project will be undertaken with a case study focusing on Fangshan. The research is expected to be completed by September 2019. Data will be collected in Fangshan in the autumn of 2017.

Why you were invited to participate?

Your experience and understanding about the administration of PISA in Fangshan and the PISA-driven initiatives at the local level are significant for this research to investigate the general picture of the perceived impact of PISA. Other participants including the team leaders of Fangshan PISA-driven projects, and head teachers as well as mathematics teachers of secondary academic schools will also be involved in this research.

Do I have to participate?

Participation in this research is entirely voluntary. It is up to you to decide whether or not to take part. If you do decide to take part you will be asked to keep this information sheet and sign a consent form. During the process of the interview, you are free to withdraw at any time without giving any reason and without there being any negative consequences. After the interview, you could also withdraw your interview data within 2 weeks.

What will happen to me if I take part?

You will be asked to take part in a face-to-face individual interview discussing Fangshan's participation in the 3 cycles of PISA, PISA results and the initiated measures for improving students' learning. The interview will take 1 hour. With your permission, the interview will be audio recorded. If you feel uncomfortable to answer any particular question or questions, you are free to decline.

What will happen to your data?

The audio recordings collected from your interview and field notes will be transcribed into texts, and then coded and anonymised for analysis. Your name will be replaced with the code "the official". Your data will be used in this research and may be used in my relevant future research in an anonymised form. Your data will not be shared with others. Due to the necessary process of supervising the quality of my PhD thesis, sections of the anonymized data collected from your interview may be looked at by the other members of this project, i.e. my supervisor and my co-supervisor. Without your written permission, your data will not be used for other purposes.

For security, the digital data (e.g. the original audio recordings, transcripts) generated from your interview will be stored in my personal storage area on University of Leeds server without anyone else having access to it. The paper-based data (e.g. field notes) will be stored in a locked place.

What will happen to the results of the research project?

The results of this research project will be my PhD thesis, and also would be other research outcomes, such as reports, publications and presentations. As described above, your information will be anonymised in any of these research outcomes. In these outcomes, some quotes of your words may be involved in
these research outcomes, however, your personal details (e.g. your position in Fangshan Education Commission) which might reveal your identity will not be disclosed.

What are the possible disadvantages and risks of taking part?

This research is not expected to bring you any direct disadvantages or risks. You will not be identified or identifiable in the outcomes of this research by general audience; it is possible that you might be identified in the outcomes of this research by people who are quite familiar with the administration of PISA-driven initiatives in Fangshan. During the interview, it is possible that you may feel uncomfortable to answer one or some questions, or not feel free to talk. To avoid this case, before the interview, if you have any questions about this research, please feel free to ask me, we could also discuss the ways to diminish the discomforts you may feel.

What are the possible benefits of taking part?

Whilst there are no direct benefits to yourself from participating in this research, a short report based on the data collected from the interviews with you and other participants will be provided to Fangshan Education Commission. This report would be a reference for Fangshan educational reforms in terms of the use of the instruments and results of educational assessments (PISA and even others)

Who is funding the research?

This research is funded by the School of Education Scholarship at University of Leeds.

Ethics Approval Reference____AREA 16-156___

Date of Approval_________15 June 2017________

Contact for further information

If you would like to know more about this research, please do not hesitate to contact me or my supervisor.

Doctoral researcher's name: Huiming Ding
School of Education

University of Leeds
Leeds, LS2 9JT, UK
Phone number: (+86) 15210980641
Email address: edhd@leeds.ac.uk

Supervisor's name: Dr Matt Homer
School of Education
University of Leeds
Leeds, LS2 9JT, UK
Email address: M.S.Homer@leeds.ac.uk

Thank you very much for taking time to read through this information.

Project title	*Document type*	Version #	Date
The Impact of PISA on Students' Learning: a Chinese Perspective	Information sheet for the official	V2	2017/08/18

UNIVERSITY OF LEEDS

A.6 Consent form for the local educational official (English version)

Participant's consent to take part in the PhD research: ***The Impact of PISA on Students' learning: a Chinese Perspective*** Name of Researcher: Huiming Ding	Add your initials next to the statement if you agree
I confirm that I have read and understand the information sheet explaining the above research project and I have had the opportunity to ask questions about the project.	
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences during the process of my interview. In addition, should I not wish to answer any particular question or questions, I am free to decline.	
I understand that I also could withdraw my interview data within the two weeks after I have been interviewed without giving any reason.	
I understand that my name will not be linked with the research materials.	
I understand that I will not be identified or identifiable in the outcomes (e.g. PhD thesis, research reports, publications, presentations, etc.) of this research by general audience; it is possible that I might be identified in these research outcomes by people who are quite familiar with PISA-driven initiatives in Fangshan.	
I agree for the data collected from me to be stored and used in relevant future research of this researcher in an anonymised form.	
I understand that this researcher may quote my words in her PhD thesis, reports, publications, presentations, and other research outcomes.	
I understand that relevant sections of the data collected during the study may be looked at by select individuals from the University of Leeds, i.e. the supervisors of this researcher, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my anonymised records.	
I agree to take part in the above research project and will inform the researcher should my contact details change.	

Name of participant	
Participant's signature	
Date	
Name of researcher	
Signature	
Date*	

*To be signed and dated in the presence of the participant.
Once this has been signed by all parties the participant should receive a copy of the signed and dated participant consent form, the letter/ pre-written script/ information sheet and any other written information provided to the participants. A copy of the signed and dated consent form should be kept with the project's main documents which must be kept in a secure location.

Project title	*Document type*	*Version #*	*Date*
The Impact of PISA on Students' Learning: a Chinese Perspective	Consent form for the official	V2	2017/05/22

A.7 Application for using Fangshan PISA data (English version)

Application for using Fangshan PISA data

I, Ding Huiming, was the staff of China PISA 2015 National Centre, engaging in PISA-related work, such as sampling, data management and analysis, etc. Now I am doing PhD in the School of Education at University of Leeds in the UK. Inspired by the work experience obtained from National Centre, the proposed title of my PhD thesis is *The Impact of PISA on Students' Learning: a Chinese Perspective*. This research will be carried out with a case study focusing on Fangshan District of Beijing. Within this local educational context, this research is going to investigate the use of PISA data and results, and the corresponding measures initiated to improve students' learning, with the goal to generate outcomes meaningful for educational practice. I also would like to make use of Fangshan data to do other research besides of my PhD thesis. Hereby, I apply for the approval of National Education Examinations Authority of China (NEEA) for using Fangshan data of PISA 2009 China Trial, PISA 2012 China Trial, PISA 2015 and Fangshan lists of participating schools in these 3 cycles (see the appendix).
The data I apply for will mainly be used in my PhD thesis, and may also be used in my own other related research. In my thesis and the other research which would use the data, the name of the researched region, namely, Fangshan, will be specified; the research results would be accessible in the ways such as publication.
I affirm that I will comply with the confidentiality in using the data, without the permission of NEEA, I will not share these data to any individuals or institutions; I will reasonably note the data source in any of my research which uses the data; I will not disclose any school names of Fangshan in my thesis or other research which use the data.

Appendix: Directory of the requested data

Applicant: Ding Huiming (Signature)

5 January 2017

Appendix: Directory of the requested data

1. PISA 2009 China Trial Fangshan student scaled database;
2. PISA 2009 China Trial Fangshan student response database;
3. PISA 2009 China Trial Fangshan school questionnaire database;
4. PISA 2009 China Trial Fangshan list of participating schools;
5. PISA 2012 China Trial Fangshan student scaled database;
6. PISA 2012 China Trial Fangshan student response database;
7. PISA 2012 China Trial Fangshan school questionnaire database;
8. PISA 2012 China Trial Fangshan list of participating schools;
9. PISA 2015 China Trial Fangshan student scaled database;
10. PISA 2015 Fangshan student response database;
11. PISA 2015 Fangshan school questionnaire database;
12. PISA 2015 Fangshan teacher questionnaire database;
13. PISA 2015 Fangshan list of participating schools.

A.8 Confidentiality statement on using PISA data (Chinese version)

教育部考试中心

关于使用 PISA 数据的保密声明

姓　　名：丁慧明　　　职　　业：博士研究生

单位名称：利兹大学　教育学院

研究领域：教育评价

E-mail：edhd@leeds.ac.uk　　　联系电话：(+44)07731902097

需求数据： PISA 2009 中国独立研究房山学生量尺化数据库；PISA 2009 中国独立研究房山学生作答数据库；PISA 2009 中国独立研究房山学校数据库；PISA 2009 中国独立研究房山参测学校名单；PISA 2012 中国独立研究房山学生量尺化数据库；PISA 2012 中国独立研究房山学生作答数据库；PISA 2012 中国独立研究房山学校数据库；PISA 2012 中国独立研究房山参测学校名单；PISA 2015 房山学生量尺化数据库；PISA 2015 房山学生作答数据库；PISA 2015 房山学校数据库；PISA 2015 房山教师数据库；PISA 2015 房山参测学校名单。

使用范围： 用于本人博士论文（拟题为《PISA 对学生学习的影响：中国视角》），以及本人基于该数据的其他学术研究。本人博士论文及基于该数据的其他研究将指明所研究地区的名称，房山；研究成果将可能用于发表等公开途径。

第 1 页，共 2 页

教育部考试中心

本人承诺：本人愿意并保证遵守对所申请数据的保密规定，不向任何个人或机构泄露或分享该数据；在本人任何基于该数据的研究中将合理标注数据来源；在本人博士论文以及基于该数据进行的其他研究中，不透露具体学校名称。

承诺人签名：丁慧明

2017 年 1 月 6 日

A.9 Confidentiality statement on using PISA data (English version)

Name:	Ding Huiming	**Occupation:**	PGR student
Institution:	University of Leeds, School of Education		
Research area:	Educational assessment		
E-mail:	edhd@leeds.ac.uk	**Phone number:**	(+44) 07731902097

Data needed:

PISA 2009 China Trial Fangshan student scaled database;
PISA 2009 China Trial Fangshan student response database;
PISA 2009 China Trial Fangshan school questionnaire database;
PISA 2009 China Trial Fangshan list of participating schools;
PISA 2012 China Trial Fangshan student scaled database;
PISA 2012 China Trial Fangshan student response database;
PISA 2012 China Trial Fangshan school questionnaire database;
PISA 2012 China Trial Fangshan list of participating schools;
PISA 2015 China Trial Fangshan student scaled database;
PISA 2015 Fangshan student response database;
PISA 2015 Fangshan school questionnaire database;
PISA 2015 Fangshan teacher questionnaire database;
PISA 2015 Fangshan list of participating schools.

Use of the data:

The data listed above will be used in my own PhD thesis (currently, the title of my thesis is drafted as *The impact of PISA on students' learning: a Chinese perspective*) and my own other related researches. In my thesis and the other researches which would use the above data, the name of the researched region, namely, Fangshan, will be specified; the research results would be published or open accessed.

Confidentiality undertaking:

I, the undersigned, agree and ensure that I will comply with the confidentiality in using the data listed above, will not disclose or share these data to any individuals or institutions; I will reasonably note the data source in any of my research which uses the above data; I will not disclose any school names in my thesis or other researches which use the above data.

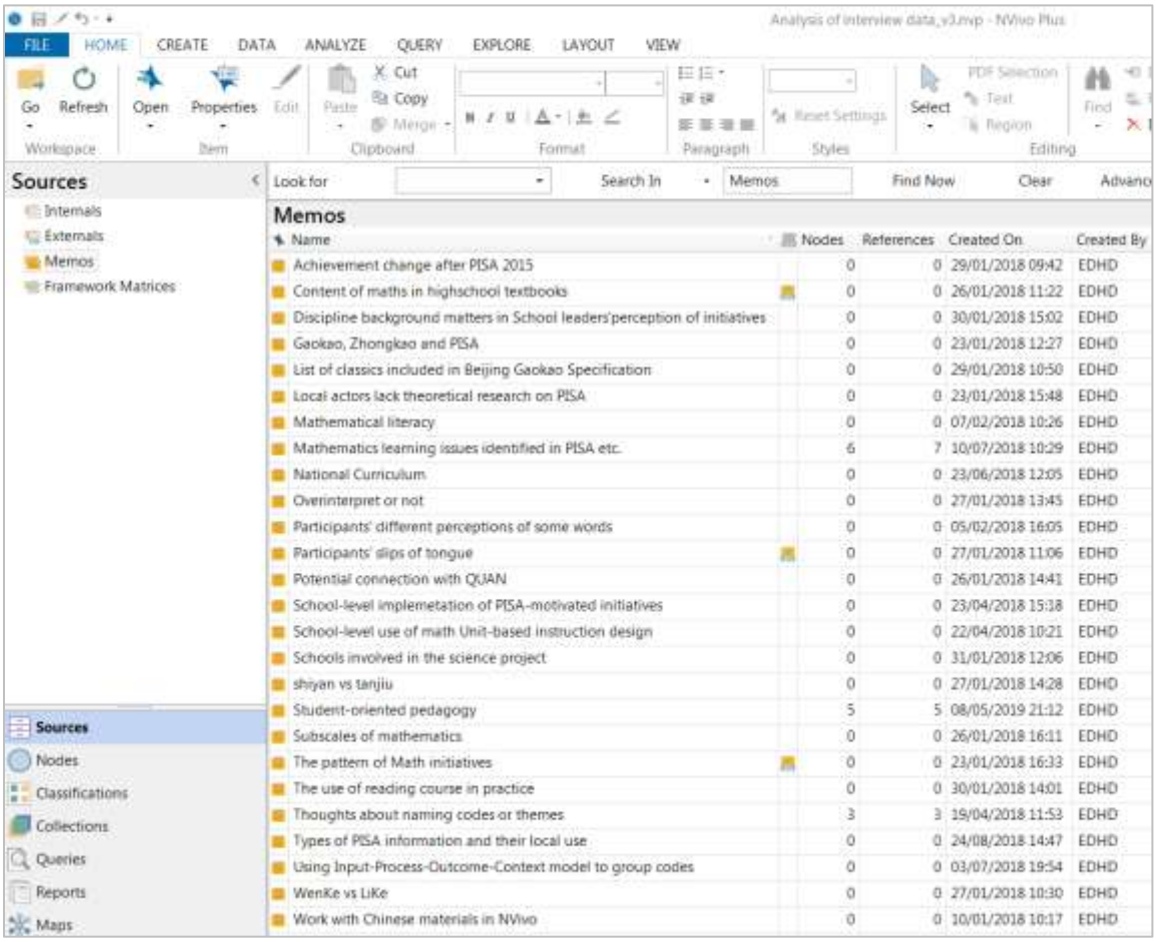
Committed signature: Ding Huiming

6 January 2017

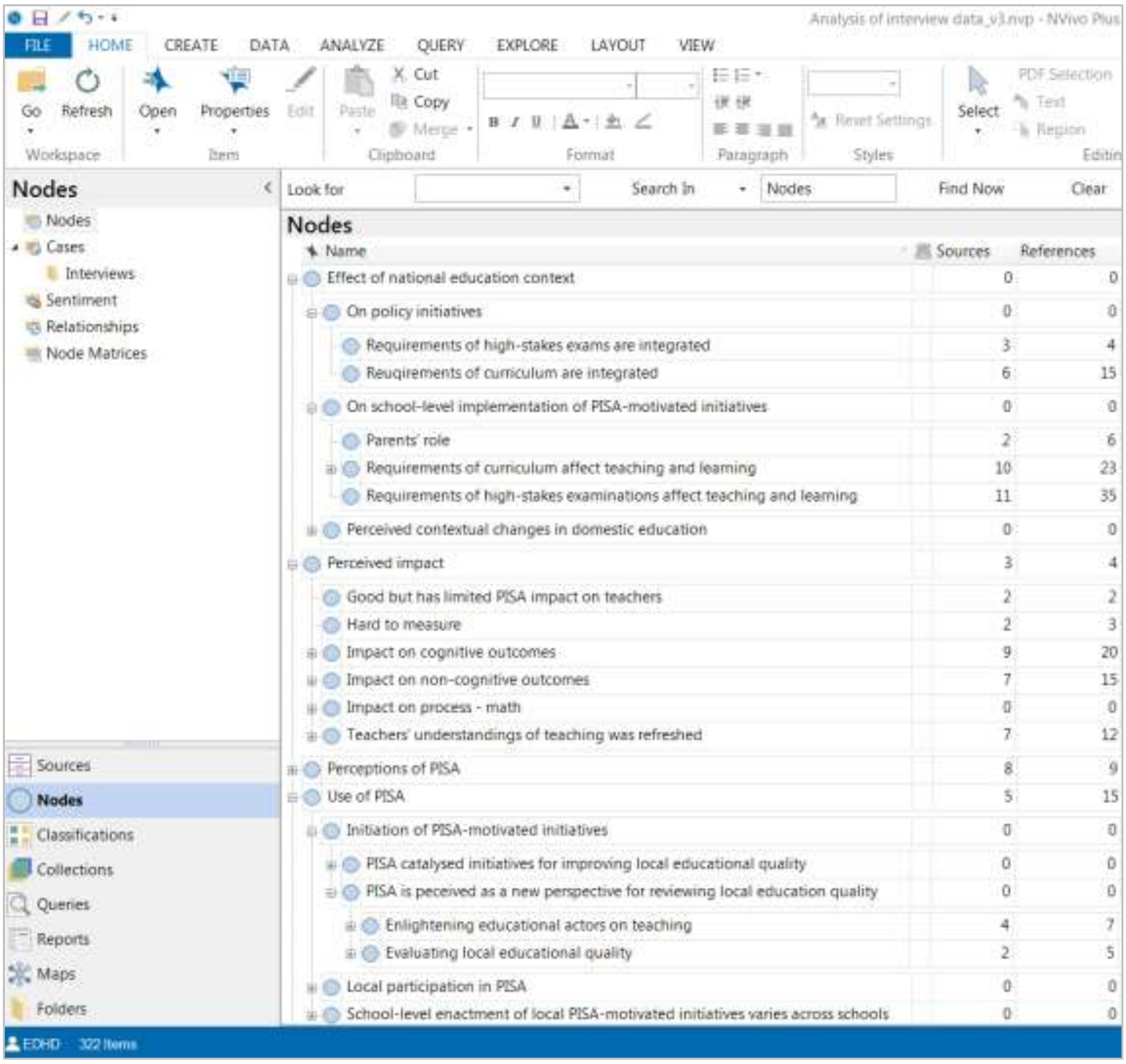
A.10 Transcription notation

<table>
<tr><td>Symbol</td><td>Description</td></tr>
<tr><td>[ ]</td><td>Response to the speaker.</td></tr>
<tr><td>__[ ]</td><td>Parallel speech. ”__” refers to the speech uttered by the speaker; “[ ]” refers to the speech uttered at the same time by the listener.</td></tr>
<tr><td>{ }</td><td>Speakers’ nonverbal expression and behaviour. (Exceptions: {A}, {A，B}, {A，B，C} refer to the examples of “Set” in mathematics)</td></tr>
<tr><td>（ ）</td><td>Missing words, uncompleted speech.</td></tr>
<tr><td>__（ ）</td><td>Correction of the slip of tongue. The text in ”（ ）” refers to the content correcting for that in “__”.</td></tr>
<tr><td>（？）</td><td>Unclear syllable.</td></tr>
<tr><td>|</td><td>Short pause.</td></tr>
<tr><td>|||</td><td>Long pause.</td></tr>
<tr><td>Bold font</td><td>Emphasised words.</td></tr>
<tr><td>——</td><td>Prolonged sounds.</td></tr>
<tr><td>X</td><td>Interrupted.</td></tr>
</table>

A.11 Memos of qualitative analysis



A.12 Screenshot of categorising codes



A.13 Ethical approval letter

The Secretariat
Level 11, Worsley Building
University of Leeds
Leeds, LS2 9JT
Tel: 0113 343 4873
Email: ResearchEthics@leeds.ac.uk

UNIVERSITY OF LEEDS

Huiming Ding
School of Education
University of Leeds
Leeds, LS2 9JT

ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee
University of Leeds

15 June 2017

Dear Huiming

Title of study: **The Impact of PISA on Students' learning: a Chinese Perspective**
Ethics reference: **AREA 16-156**

I am pleased to inform you that the above research application has been reviewed by the ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee and I can confirm a favourable ethical opinion as of the date of this letter. The following documentation was considered:

Document	Version	Date
AREA 16-156 Ethical review form_Huiming Ding_v3_signed.doc	1	25/05/17
AREA 16-156 Application for using Fangshan PISA data_Huiming Ding_English version.pdf	1	25/05/17
AREA 16-156 Confidentiality statement on using PISA data_Huiming Ding_English version.pdf	1	25/05/17
AREA 16-156 Data_management_plan_Huiming Ding_v1.docx	1	25/05/17
AREA 16-156 Draft interview guide_Huiming Ding_v1.docx	1	25/05/17
AREA 16-156 Fieldwork assessment form_MED risk_ESSL_Huiming Ding_v2_signed.doc	1	25/05/17
AREA 16-156 Information sheet for head teachers_Huiming Ding_v1.docx	1	25/05/17
AREA 16-156 Information sheet for mathematics teachers_Huiming Ding_v1.docx	1	25/05/17
AREA 16-156 Information sheet for the official_Huiming Ding_v2.docx	1	25/05/17
AREA 16-156 Information sheet for the team leaders of PISA-driven projects_Huiming Ding_v1.docx	1	25/05/17
AREA 16-156 Letter seeking permission from school gatekeepers_Huiming Ding_v1.docx	1	25/05/17
AREA 16-156 Participant_consent form for head teachers and mathematics teachers_Huiming Ding_v2.doc	1	25/05/17
AREA 16-156 Participant_consent form for the official and leaders of PISA-driven projects_Huiming Ding_v2.doc	1	25/05/17
AREA 16-156 Phone script of inviting potential participants_Huiming Ding_v1.docx	1	25/05/17

Committee members made the following comments about your application:

General comments		
A carefully thought-out and thorough application. Thank you. The comments below are minor points for your consideration.		
Application section	**Comment**	**Response required/ amended application required/ for consideration**
7.2 and Draft interview guide	It appears that there is only one official of Fanshan Education Committee who is responsible for local admin of PISA… Likewise, it seems there are only 3 leaders of the projects driven by PISA. If this is the case, they would be easily identifiable by anyone who knows the area, even if a pseudonym is used. How will you protect their identity?	For consideration
8.8 and Draft interview guide	You are right to note that the interviewees may not feel able to give you the whole story – you might also want to consider the potential consequences for participants if they do say something critical of the Chinese government. Consider how you will ensure that they are protected in such a case. You might find Milroy's work in Northern Ireland during the Troubles useful: Milroy, L. (1980). *Language and Social Networks*. Oxford: Blackwell.	For consideration

Please notify the committee if you intend to make any amendments to the information in your ethics application as submitted at date of this approval as all changes must receive ethical approval prior to implementation. The amendment form is available at http://ris.leeds.ac.uk/EthicsAmendment.

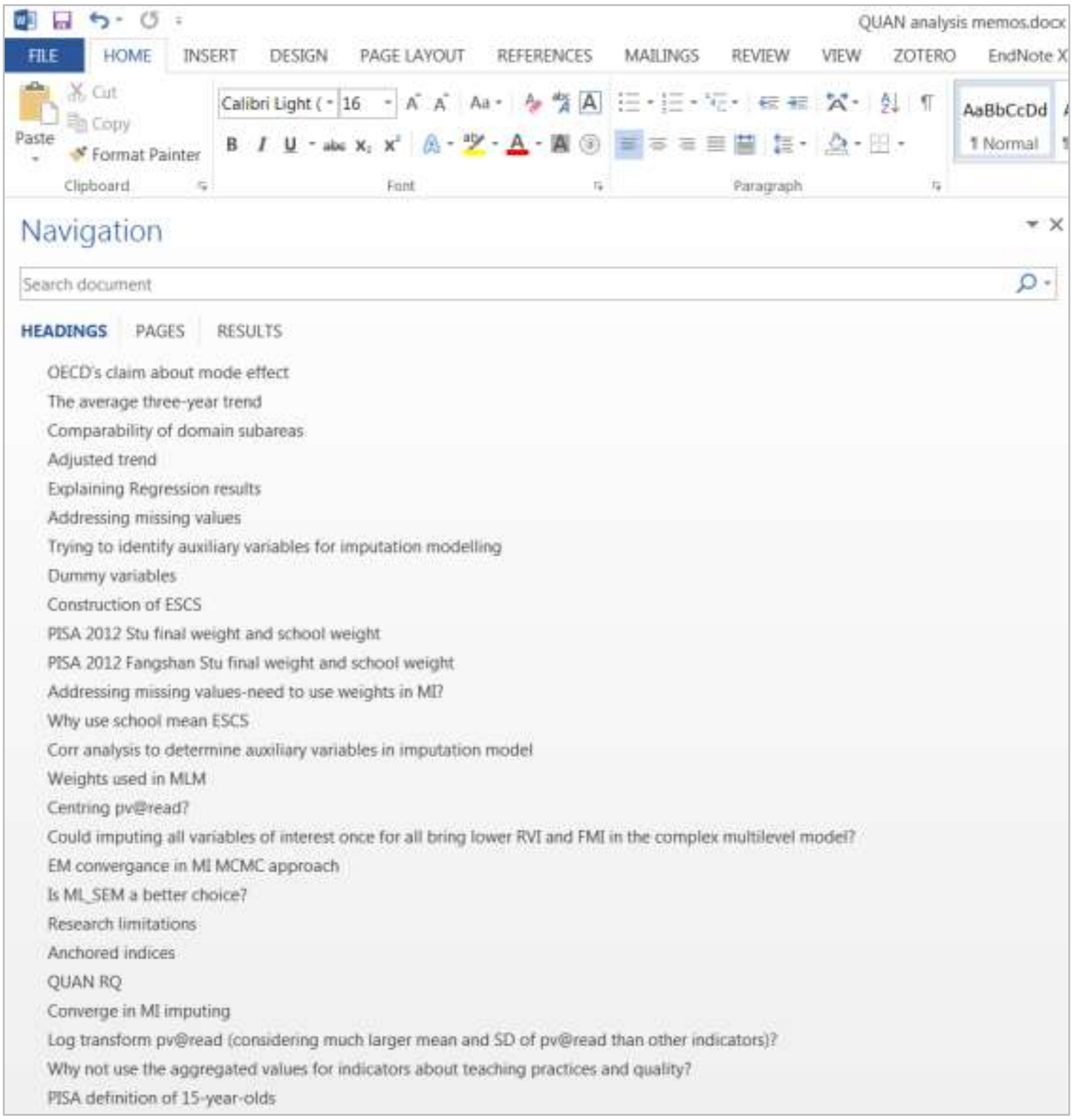
Please note: You are expected to keep a record of all your approved documentation and other documents relating to the study, including any risk assessments. This should be kept in your study file, which should be readily available for audit purposes. You will be given a two week notice period if your project is to be audited. There is a checklist listing examples of documents to be kept which is available at http://ris.leeds.ac.uk/EthicsAudits.

We welcome feedback on your experience of the ethical review process and suggestions for improvement. Please email any comments to ResearchEthics@leeds.ac.uk.

Yours sincerely

Jennifer Blaikie
Senior Research Ethics Administrator, the Secretariat
On behalf of Dr Kahryn Hughes, Chair, AREA Faculty Research Ethics Committee
CC: Student's supervisor(s)

A.14 Journal of quantitative analyses



A.15 The process of conceptualising PISA's impact on students' learning