Moderator effect of reading proficiency on science performance across 15-year-old boys and girls in the Netherlands, Finland, and Singapore

Bo Zhang
Educational Science & Technology
Faculty of Behavioral Sciences
University of Twente, Enschede, the Netherlands
August 2017

Moderator effect of reading proficiency on science performance across 15-year-old boys and girls in the Netherlands, Finland, and Singapore

Author

Researcher:	Bo Zhang
Student Number:	s1657879
Email:	b.zhang@student.utwente.nl
Faculty:	Master of Educational Science and & Technology, Behavioral Science, University of Twente

Supervisors

1st Supervisor:	Dr. Martina Meelissen, m.r.m.meelissen@utwente.nl Assistant Professor Department of Research Methodology, Measurement and Data Analysis Enschede, Netherlands
2nd Supervisor:	Dr. Hans Luyten, j.w.luyten@utwente.nl Associate Professor Faculty of Behavioural Sciences, University of Twente Enschede, Netherlands

Date: August 2017

ACKNOWLEDGEMENTS

I am very lucky to be in this research group. I learnt more than I expect in this memorable research experience!

I give my sincere thanks to my first supervisor Dr. Martina Meelissen not only for granting me this great opportunity to study in this project but timely, insightful, open and concrete suggestions at each stage of the thesis writing. With her guidance, my research gradually goes further and clearer in this process.

I thank my second supervisor Dr. Hans Luyten for his nice patience on my thesis modification and constructive guidance on the data processing part of this project. I learned practical skills from him on dealing with data and make convincing research. Equally importantly, I learn to practice conscientious scholarship under the guidance of them. This is the starting point and also one of the criteria of true research.

I thank Ms. Anniek for her generous help and suggestions in my thesis writing.

I thank my friends Kelly, Zhen, Li, Ning, and Marieke for their kind help on the childcare for my daughter. Without their help, I would be unable to concentrate on my study.

I thank my husband Xuelong for his financial and psychological support for my study.

My final thanks go to my parents and my daughter for their understanding and unconditional support for my study.

ABSTRACT

Purpose: The major aim of the present study is to investigate students' science self-efficacy and its relationship with science performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore. On the other hand, whether and in what way students' reading proficiency moderates the relationship between science self-efficacy and science performance was explored across boys and girls in the three countries

Methods: Literature review and moderated regression analysis with SPSS programming software are major tools to address the research questions.

Results: Data analysis based on PISA 2015 dataset reveals a significant gender difference in science self-efficacy in favor of boys across the Netherlands, Finland and Singapore. In addition, Dutch girls demonstrate lower science self-efficacy than Finnish girls and Singaporean girls. Significant gender difference in science performance in favor of boys is demonstrated except Finland. Significant gender difference in reading performance in favor of girls is demonstrated across the three countries. A significant and positive relationship between science self-efficacy and science performance was found across 15-year-old boys and girls in the three countries. Moreover, moderator effect of reading proficiency is satisfied across boys and girls in the three countries. As reading proficiency increases, the relationship between science self-efficacy and science performance is strengthened across gender in the three countries.

Keywords: science self-efficacy, science performance, reading proficiency, gender difference, moderator effect.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

1.1 RESEARCH BACKGROUND

STEM is the acronym for four separate disciplines of science, technology, engineering and mathematics. It has caught the attention from researchers since it originated at the National Science Foundation (NSF) in the US in 1990s (Bybee, 2013). Innovation in STEM is regarded to be vital for a country's economic competitiveness and technological innovations (Mullis & Martin, 2014; van Aalderen-Smeets & van der Molen, 2016). In educational contexts, STEM has been "one of the newest slogans' (Bybee, 2013). The significance of STEM education has been increasingly recognized by educational administrators and legislators (White, 2014). However, a lack of student enrolment in STEM-related programs has confronted many western countries (van Aalderen-Smeets & van der Molen, 2016). The low representation of students in STEM programs, especially girls, has called for much concern (Watt et al. 2012).

According to Salmon (2015), girls' underrepresentation in STEM programs could be traced back to their school achievement. It was suggested that the low participation rates of female students in STEM-related careers may be due to delays of one gender with respect to the other in the related subject achievements (Langen, Bosker, and Dekkers, 2006). Girls' lower achievement than boys in STEM-related subjects persists in many countries (Hango, 2013; Modi, Schoenberg, & Salmond, 2012; Smeding, 2012; OECD, 2016). Triennial assessment from Program for International Student Assessment (PISA) among 15-year-old students in 2015 shows that on average girls' performance in science and mathematics is significantly lower than boys across most of OECD countries (OECD, 2016). Moreover, girls are consistently underrepresented than boys among top performers of science and mathematics except Finland (OECD, 2016).

Girls' lower performance in STEM-related subjects has been concerned for decades, such as in the Netherlands (Booy, Jansen & Joukes, 2012), the US (Rossi, 1965), France (De Peslouan, 1974) and Turkey (Bybee, 2010). A variety of researchers explore girls' low performance in STEM subjects from the perspective of the cognitive ability (Meyer & Koehler, 1990), self-efficacy (Tseng, Chang, Lou & Chen, 2013; MacPhee, Farro, & Canetto, 2013) and threat of stereotypes from teachers and parents (Shapiro & Williams, 2012) and so on. For instance, science performance has been related with self-efficacy in many studies as summarized in Chapter 2 (see Table 1), attempting to understand girls' underperformance in STEM-related subjects. It is found that girls' self-efficacy is a reliable and significant predictor of science performance (Bandura, 1986; Motlagh, Amrai, Yazdani, altaib Abderahim & Souri, 2011).

1.2 RESEARCH GAP

Despite substantial research concerning girls' underperformance in STEM-related subjects, research gap as well as research inconsistency exists. First, among the literature investigating the low achievement of girls in STEM performance, most researchers attempt to study the issue from the perspective of gender gap in STEM related subjects independent from the non-STEM subjects. Little research is conducted on the possible role of non-STEM subjects such as reading on the STEM-related subject performance. For instance, whether and to what extent students' science performance changes with different reading level is little concerned. The possible moderator effect from a non-STEM subject on a STEM subject performance is also little studied.

Recently, the widening gender gap on reading performance started to draw researchers' attention to non-STEM subjects in addressing gender gap in STEM subjects (Wang, Eccles, & Kenny, 2013; OECD, 2016). Parallel with girls' declining performance in STEM subjects, boys are lagging behind girls in reading in consecutive years with a larger gender gap (van Langen, Bosker & Dekkers, 2006; OECD, 2012). Low reading performance by male students is suggested to decrease the reading performance of a country (OECD, 2016). Consequently, a contrast condition occurs. Overall, the STEM subject science on which boys perform well is not well performed by girls. The non-STEM subject reading on which boys perform weak is well performed by girls. Whether and in what way reading proficiency affects students' science performance is worthwhile to be explored.

Second, there is abundant research investigating students' performance from the perspective of self-efficacy in a broad way, investigation on science self-efficacy is insufficient. According to a study investigating sources of science self-efficacy beliefs among middle school students, it was suggested that science self-efficacy was the most consistent and reliable predictor of students' science performance (Britner & Pajares, 2006). Thus, investigating students' science self-efficacy instead of the broad concept of self-efficacy in the correlation between self-efficacy and science performance could be more valid.

In addition, though abundant literature focuses on the relationship between self-efficacy and academic performance, consensus is not reached. According to Bandura (1986), self-efficacy is a significant and reliable predictor of academic performance since he proposed social cognitive theory (1986). This was supported by a number of researchers who also found out a significant and direct relationship between the two factors in high school (Motlagh, Amrai, Yazdani, altaib Abderahim & Souri, 2011) and among freshmen (Chemers, Hu, & Garcia, 2001). However, in recent years, increasing research is focused on an indirect relationship between students' self-efficacy and their academic performance. It was found that a third variable such as a high level of narcissism can transform the relationship between self-efficacy and academic performance (Beattie et al., 2017).

To summarize, the research gap on girls' underperformance of STEM subjects is mainly concerned with three respects. First, the role of non-STEM subject such as girls' superior ability in reading has been insufficiently considered when investigating girls' underperformance in STEM subjects. Second, the study of self-efficacy in the prediction of STEM subject performance appears to be broad compared with the study of science self-efficacy. But science self-efficacy is little concerned in previous study on gender difference of STEM subject performance. Third, inconsistency exists as to the relationship between students' self-efficacy and academic performance. Thus, the above-mentioned research gap and research inconstancy makes the starting point of this project.

1.3 RESEARCH AIMS & OBJECTIVES

This project aims to explore the gender gap of students' science performance from the perspective of girls' superior ability of a non-STEM subject reading in countries of the Netherlands, Finland and Singapore. Specifically, there are three major objectives.

1. Identify the difference of science performance, reading performance and science self-efficacy across 15-year-old boys and girls in the Netherlands, Finland and Singapore (O1).
2. Identify how science self-efficacy correlates with students' science performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore (O2).
3. Identify in what way a third variable reading proficiency affects the strength and/or the direction of the relationship between science self-efficacy and science performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore (O3)

In this project, Finland and Singapore were selected to make a comparison with the Netherlands. Reasons for why they are selected are explained in Chapter 3.

1.4 RESEARCH QUESTIONS

Based on the above three research objectives, the thesis investigates the gender gap in science performance mainly by exploring the moderator effect of a non-STEM subjects (i.e. reading). To do this, first, the overall gender gap in science performance among the three countries is studied. Second, the relationship between science self-efficacy and science performance is investigated. Third, an indirect relationship between science self-efficacy and science performance is modeled through the moderator analysis on reading proficiency. Accordingly, the three primary research questions are as follows:

> **RQ1.** To what extent is science performance, reading performance and science self-efficacy different across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA 2015 data?
>
> **RQ2.** To what extent does science self-efficacy correlate with science performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA 2015 data?
>
> **RQ3.** To what extent does reading proficiency weaken, strengthen or transform the relationship between science self-efficacy and science performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA data 2015?

The following diagram illustrates the order and content of the three research questions.

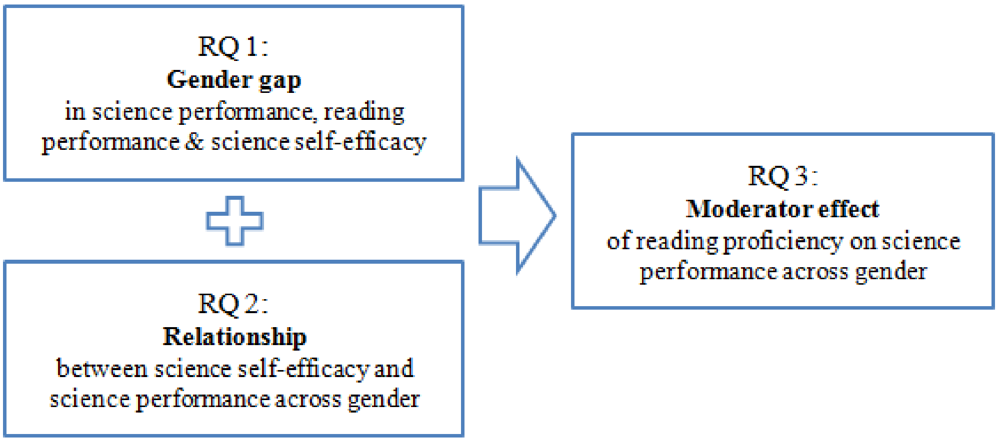


Figure 1 Overview of research questions

The above research questions connect with each other in a logical way. RQ1 and RQ2 are prerequsite for RQ3, the moderator effect of reading proficiency on the relationship between science self-efficacy and science performance. Specifically, RQ1 takes a close look at gender gap among 15-year-old students' science performance, reading performance and science self-efficacy in the three countries, aiming to identify the gender gap differences between the Netherlands and the other two countries of Finland and Singapore. In addition, gender difference in reading performance and science performance is compared so as to identify the difference of gender gap across the two domains. RQ2 focuses on the relationship between science self-efficacy and science performance, of which are presumptions of moderator effect in RQ3. Based on the relationship between science self-efficacy and science performance in RQ2, the moderator effect of reading proficiency on science performance (RQ3) can be analyzed.

1.5 SCIENTIFIC & PRACTICAL RELEVANCE

SCIENTIFIC RELEVANCE

Scientifically, comparing the gender gap between a non-STEM subject reading and a STEM subject science in the three countries not only reveals the gender gap on performance of each subject but also the difference of gender gap across the two domains among the three countries. It makes a useful background on whether and to what extent students' reading proficiency can affect their science performance. Second, it reveals to what extent students' science self-efficacy is associated with their science performance. This study based on the latest PISA data can add additional evidence on the applicability of Bandura's self-efficacy theory (1997). This would provide additional evidence on existing relationship study between self-efficacy and academic performance. Furthermore, whether and in what way reading proficiency moderates the relationship between science self-efficacy and science performance is also investigated. This adds additional evidence on the moderator effect of a third variable on the relationship between self-efficacy and academic performance.

PRACTICAL RELEVANCE

Practical implications can be offered to practitioners and school teachers. It can shed light to the other countries who wish to achieve better education equity especially on science subject. Precisely, first, this research can guide practitioners and teachers to have a correct understanding of gender difference of science performance and reduce the stereotypes towards girls' underperformance in science subject. In addition, by correlating science self-efficacy with science performance based on the PISA 2015 data in the Netherlands, Finland and Singapore, school teachers can be acknolwedged on the significance of science self-efficacy on science performance as well as the gender effect of this relationship. This could be beneficial to the school instruction on science. For instance, attention and/or intervention can be given to foster girls' science self-efficacy in science instruction and to check whether it is beneficial in improving their science performance. This research can shed light to the EU-wide benchmark target by 2020 that less than 15% of 15-year-old students should be under-achieved in science ('ET 2020'). Moreover, the moderator analysis in this project is beneficial for teachers to get aware of the role of reading proficiency on science performance.

1.6 RESEARCH STRUCTURE

To address the three primary research questions, the thesis is divided into 7 chapters.

In **Chapter 1**, the research background, the research gaps, aims, objectives, research questions, as well as scientific and practical implications are introduced.

In **Chapter 2**, the conceptual framework is developed. Literature on gender difference of STEM-related subject performance and reading performance is described together with two important indicators (i.e. ability, self-efficacy). The relationship between self-efficacy and academic performance is introduced to exemplify the significance of science self-efficacy and identify the development of the study on their relationship.

In **Chapter 3**, a comparison is made between the Netherlands, Finland, and Singapore in terms of students' science study time, teachers and teachers' quality as well as parental involvement. This part makes a necessary background for the later comparative analysis in statistics.

In **Chapter 4**, Data collection method, participants, and sampling method of PISA is introduced. Data analysis plan is made as well.

In **Chapter 5**, descriptive and regression analysis results are revealed and interpreted.

In **Chapter 6**, discussion and conclusion based on the comparative study and statistic results is conducted.

In **Chapter 7**, scientific implications for future research are recommended. In practical, brief implications with regard to the STEM and reading education are given to teachers, schools, and policy makers.

CHAPTER 2 LITERATURE REVIEW

2.1 Overview

The goal of this chapter is to provide important literature to build a theoretical framework for this project, relationship between science self-efficacy and science performance as well as the moderator effect of reading proficiency on science performance. Based on the research objectives and research questions, on one hand, important studies concerning self-efficacy and academic performance are described. Meanwhile, gender gap in science performance, reading performance and self-efficacy are explored. On the other hand, the significance of self-efficacy and its relationship with academic performance is investigated.

A systematic review was conducted to collect literature. First, database of Web of Science, and Science Direct were used to search for articles from 2007 until the search date in June of 2017. The year 2007 was set as the starting year because a meta-analysis with 242 studies between 1990 and 2007 had revealed the trend in gender and STEM-related subject performance before 2007 (Lindberg, Hyde, Petersen, & Linn, 2010). Several combinations of the following key words were searched in the database of Web of Science and Science Direct.

- Gender difference, gender gap;
- Science self-efficacy;
- Ability, performance;
- Reading, science;.
- STEM.

Only the articles that meet at least one of the following criteria were included.

- The article studies the relationship between students' (science) self-efficacy and (science) performance;
- The article reports the gender gap in science performance, reading performance, (science) self-efficacy in the Netherlands, and/or Finland, and/or Singapore;
- The article studies gender gap in STEM subjects from the perspective of non-STEM subject such as verbal ability and/or readisng ability.

Consequently, thirty-nine articles were selected in the database of Web of Science with the combinations of key words above. Five articles were found with keywords of "reading", "science", "gender" and "STEM". Another round of selection is conducted together with the articles included in the meta-analysis (Lindberg, Hyde, Petersen, & Linn, 2010). The number of articles was narrowed down to thirty-two.

2.2 Ability

Ability can be defined as the natural aptitude or acquired proficiency of an individual (*Meriram-Webser*, 2017). Various studies explored gender gap on science-related subjects from the perspective of innate cognitive skills (Bhana, 2005; Fennema, Peterson, Frome & Eccles, 1998; Furnham, Reeves & Budhani, 2002). It was found that females tend to outperform in verbal abilities (Halpem et al., 2007). Males outperform females on visual-spatial abilities which are implicated to contribute to mathematics and science performance in standardized exams (Halpem et al., 2007). In addition, in a gender difference study on mathematical cognition using brain imaging (Keller & Menon, 2009), there showed significant differences in brain responses during mathematical cognition tasks. All these suggest that girls' underperformance in STEM subjects might be related with their inherent cognitive abilities. Moreover, the belief of inferior ability of girls on STEM-related subjects is prevalent among parents and teachers (Blascovich, Spencer, Quinn, & Steele, 2001; Else-Quest, Hyde & Linn, 2010).

However, in recent years, there are increasing discussions on gender similarities in math ability. Gender similarity hypothesis claims that male and female students are actually more similar than different in mathematics ability (Hyde, 2005; Hyde, Lindberg, Linn, Ellis, & Williams, 2008). According to Booij, Jansen, and Joukes (2012), there is no gender gap in terms of STEM aptitude as revealed by Neuroscientific research. It is suggested that "in principle, girls and boys can perform at the same level in science subjects" (Booy, Jansen & Joukes, 2012, p. 9).

In line with it, a trend in recent years shows that gender gap in science is narrowed considerably across nations (OECD, 2016). Data shows that individual science difference is significantly larger than gender difference (OECD, 2016). Girls' performance becomes equal or even higher than boys in some countries (Booy et al., 2012), such as Finland where girls outperform boys in both mathematics and science (OECD, 2016). In the US, girls are found to be equal or higher than boys in computation (Hyde, 2005) and the standardized tests (Hyde et al., 2008). Therefore, it is likely that gender gap in STEM subject performance may not be attributed to students' inherent ability but other factors, such as explicit or implicit beliefs (Booy, Jansen & Joukes, 2012). One of the possible explanations could be gender difference in self-efficacy.

2.3 Science self-efficacy

An important implicit belief among students related to gender gap in academic performance is self-efficacy. In this part, the definition, origin and significance of self-efficacy and science self-efficacy are described. In addition, its relationship with academic performance is explored as well.

2.3.1 Definition & Significance

Self-efficacy, coined by Albert Bandura forty years ago (1977), is defined as

> "people's judgment of their capabilities to organize and execute courses of action required to attain designated types of performances" (1986, p. 391).

To put it in another way, self-efficacy refers to individual belief on one's ability in performing a specific task in a given context (Pajares, 2005; Zimmerman, 2000). Self-efficacy is regularly measured in the large-scale PISA assessment, science self-efficacy, to be exact. According to PISA 2015, science self-efficacy is defined as:

> "the extent to which students believe in their own ability to handle science tasks effectively and overcome difficulties" (OECD, 2017, p.3).

The significance of self-efficacy has been recognized by researchers with its relationship with students' academic performance (Nugent et al., 2015; Kıran & Sungur, 2012). Bandura (1997) has suggested that students with low self-efficacy are very likely to underperform in science regardless of their abilities. In addition, female students who are proficient in mathematics frequently fail to pursue mathematics-related careers is also associated with their low self-efficacy (Zeldin & Pajares, 2000). Therefore, the role of self-efficacy cannot be ignored in the exploration in student performance.

It should be noted that self-efficacy is not concerned with individual interest, ability, or motivation (Sharma & Nasa, 2014), but they are related in a strong way (Kozlowski & Salas, 2010). Self-efficacy is regarded to provide the foundation for human motivation and personal achievement (Sharma & Nasa, 2014). Learning engagement and duration of persistence in the tasks is suggested to be considerably influenced by students' self-efficacy beliefs (Pintrich & Schunk, 2002; Pajares 1996). Namely, people are willing to select a task on which they have strong self-efficacy. If learners have no confidence in their ability to fulfill particular tasks, they will not engage required effort on the tasks (OECD, 2013; Kıran & Sungur, 2012; Pintrich & Schunk 2002). Though factors such as motivation and self-control can motivate and guide students, but learners with low self-efficacy are unlikely to be motivated to learn (Klassen & Usher, 2010).

2.3.2 Gender gap in science self-efficacy

PISA 2012 report shows that girls in the Netherlands have a "significantly lower self-efficacy" with regard to science than boys (Booy, Jansen, & Joukes, 2012). Meanwhile, Dutch students present a greater gender gap with regard to self-efficacy than the remaining 39 countries except Switzerland (Booy et al., 2012). Gender gap still exists in the latest PISA assessment in 2015 (see Chapter 5). Though previous assessment data shows that gender gap is not evenly distributed among all the countries and economies, the mean index of science self-efficacy of boys is significantly higher than that of girls in 41 countries and economies (OECD, 2016). The opposite condition happens in eight countries. The remaining 23 countries, there is no significant gender difference in science self-efficacy (OECD, 2016). These suggest that gender gap in self-efficacy in favor of boys appears to be a more common issue.

The "confidence gap" has been found between male and female students for a long time (Sadker & Sadker, 1994). In the literature exploring female students' underrepresentation in STEM subjects (van Aalderen-Smeets & van der Molen, 2016), it is found that girls' self-efficacy in relation to STEM subjects is often not in consistent with their ability (Booy, 2012). According to PISA 2015 report, girls tend to be more likely to have low self-efficacy than boys (OECD, 2016). In a study conducted by Zeldin and Pajares (2000), a girl who is proficient in science might have an inconsistently low science self-efficacy. In contrast, it frequently occurs that boys tend to overestimate their ability in mathematics and science (Watt, 2006). In line with it, a meta-analysis by Else-Quest, Hyde and Linn (2010) shows that boys are "extrinsically and intrinsically" more confident and have higher self-efficacy than girls. However, to what extent is students' self-efficacy associated with their academic performance? This should be investigated first before addressing the underperformance of girls in STEM-related subjects.

2.3.3 Self-efficacy & academic performance

The relationship between self-efficacy and academic performance is investigated in sufficient literature. The major findings are listed (see Table 1). From the table, it can be seen that Bandura's self-efficacy theory was most influential and frequently used as the basic theoretical framework for self-efficacy research. Based on Bandura's self-efficacy theory, self-efficacy is recognized as a positive and reliable predictor of academic performance (Bandura, 1986; Chemers, Hu, & Garcia, 2001; OECD, 2013; Motlagh, Amrai et al., 2011; Sharma & Nasa, 2014). By Bandura's self-efficacy theory (1993), five behaviors have been classified suggesting that students with high academic self-efficacy tend to carry a positive attitude towards problems, commitment to the goals and affect academic performance in a positive way. It is believed that self-efficacy beliefs can "lead to specific behaviors and motivations that can encourage or discourage performance" (Sharma & Nasa, 2014, p. 60).

In line with the above research findings, data shows that science self-efficacy is positively related with science performance across OECD countries according to the latest PISA 2015 results, (OECD, 2016). What's more, the relation between self-efficacy and performance is positive and significantly stronger among the highest-achieving students than the lowest-achieving students in all countries except Algeria and the Dominican Republic (OECD, 2016). Students with low science self-efficacy tend to perform weaker in science application knowledge and skills than students with high science self-efficacy (OECD, 2017).

However, the direct relationship between self-efficacy and performance was challenged in recent years. Increasing studies are concerned with an indirect relationship between self-efficacy and performance (see Table 1). For instance, it was found that a high level of narcissism can transform the relationship between self-efficacy and subsequent performance into negative (Beattie et al., 2017). Moreover, in a meta-analysis directed by Vancouver & Purl (2016), both positive and negative relationships were found between self-efficacy and performance in different conditions. It was suggested that their relationship could vary according to different types of students and study characteristics. What's more, in a comparative study investigating the moderator effect of feedback on science-efficacy-performance relationship, it is found that type of feedback is an important moderator of the science-efficacy-performance relationship. Namely, the relationship can be transformed into negative with minimal feedback but positive with detailed feedback (Beattie, Woodman, Fakehy & Dempsey, 2016). These research findings show that the relationship between self-efficacy and academic performance can be moderated by a third variable.

However, a great proportion of research is focused on the influence of self-efficacy on academic performance. The study on the potential prediction from academic performance to self-efficacy is not discovered in previous literature. It should be noted that the significance of self-efficacy is not only restricted to its relationship with academic performance, but also other factors such as students' career aspirations. Due to the limitation of the project topic, relationship with other factors would not be dealt in detail.

The following table shows the important research results on the relationship between self-efficacy and academic performance.

Table 1

Research findings on the relationship between self-efficacy & academic performance

	Author(s)	Self-efficacy theory	Outcome variable	Predictive variable	Value	Subjects	Conclusion
1	(Motlagh et al., 2011)	Bandura, (1986), Pajares, (2002)	Academic achievement	Self-efficacy	(see Table 1, p. 766)	High school	"Self-efficacy is a considerable factor in academic achievement" (p. 765)
2	(Sharma & Nasa, 2014)	Bandura, (1977)	Educational performance	Academic self-efficacy	(Not applicable)	(Not applicable)	"Self-efficacy leads to specific behaviors and motivations that can encourage or discourage effective performance" (p. 58)
3	(Meral, Colak, & Zereyak, 2012)	Bandura, (1986, 1997)	Academic performance	Self-efficacy	r= .45 (p< .01)	University	"Significant correlation may show self-efficacy is an important variable on students' academic performance and effects their achievement positively".
4	(Motlagh, Amrai, Yazdani, Abderahim & Souri, 2011)	Bandura (1986)	Academic performance	Self-efficacy	(see Table 1, p. 766)	High school	"Self-efficacy is a considerable factor in academic achievement" (p.765).
5	(Chemers, Hu & Garcia, 2001)	Bandura (1997)	academic self-efficacy and optimism	Classroom performance	Academic performance (standardized coefficient = .17, p < .01	Freshmen	"Academic self-efficacy was significantly and directly related to academic performance" (p. 61).
7	(Pajares & Graham, 1999)	Bandura (1986)	mathematics self-efficacy	mathematics performance	(.57 fall, .59 spring	Middle school students	"Students' task-specific self-efficacy was the only motivation variable to predict performance and did so both at start and end of year" (p. 133).
8	(Skaalvik, Federici &	Bandura (1986)	Motivation	Mathematics performance	(see Table 2, p. 133)	Middle school students	Self-efficacy serves as a mediator between mathematics performance and motivation

	Klassen, 2015).	Zimmerman & Cleary (2006)					
9	(Wu, 2017)	(Bandura, 2006)	Learning performance	Media multi-tasking self-efficacy	(Table 3, p. 947)	University students	“”…revealed significant negative indirect relationship between media multitasking self-efficacy and learning performance via both students’ perceived attention problems and self-regulation strategies” (p. 56).
10	(Beattie, S., Dempsey, C., Roberts, R., Woodman, T., & Cooke, A., 2017)	Bandura, 1997). (Beattie et al., 2011)	Self-efficacy	Performance	Significance between performance & science efficacy b20=.18, p<.001 Moderator effect of narcissism b21 = .21, p<.08	University students (mean age=22)	“High levels of narcissism may contribute to the negative relationship between self-efficacy and subsequent performance” (p. 65).
11	(Vancouver, J. B., & Purl, J. D., 2016)	Bandura and Locke (2003) 2013, Sitzmann and Yeo Bandura (2012, 2015)	Performance	Self-efficacy	Not applicable	Not applicable	Self-efficacy is positively related to performance and effort for individuals in the low ambiguity condition.
12	Andrew, S. (1998).	(Andrew, 1995) (Bandura, 1986, 1977)	Academic performance	Self-efficacy for science (SEFS)	SCIE110 (r=0.49, P=0.0001) CIE111 (r=0.43, p=0.0005)	Freshmen	“Self-efficacy was indeed statistically significantly correlated with academic performance in the two bioscience subjects” (p. 599).

13	(Bergey, Ketelhut., Liang, Natarajan & Karakus, 2015)	Paj ares 1996 Bandura (1977, 1986, 1997)	students' scientific inquiry self-efficacy	performance on a science assessment	(see Table 1, p. 702)	middle school students (sixth, seventh, & eighth grades)	"we found significant differences in boys' and girls' computer game self-efficacy, consistent with Ketelhut" (p. 705) computer game self-efficacy was not predictive of or predicted by performance in a virtual environment. (p.706)
14	(Pajares & Graham, 1999)	Bandura's (1986)	Performance	Students' task specific self-efficacy	Self-efficacy: Beta=0.267, p=0.0001 Gender: beta=-0.087	grade 6 in middle school	"Mathematics self-efficacy was the only motivation variable to predict mathematics performance both at beginning and end of year" (p.133).
15	(Zimmerman. 2000)	Bandura (1977, 1997)	Achievement outcomes	Self-efficacy	Not applicable	Not applicable	"Self-efficacy has proven to be responsive to improvements in students' methods of learning and predictive of achievement outcomes" (p. 89). "The strength of efficacy beliefs accounted for more than 13% of the variance in their final math grades" (p. 87).
16	(Multon, Brown, & Lent, 1991)	Lent et al., 1987; Siegel, Galassi & Ware, 1985)	Academic performance	Self-efficacy beliefs	Effect size estimates: .38	Not applicable-met a analysis	Self-efficacy beliefs account for approximately 14% of the variance in students' academic performance" (p. 34)

2.4 A BROADER VIEW: READING VS. SCIENCE

Girls outperform boys in reading in all nations in consecutive years with a much larger gender gap than mathematics (OECD, 2009). What's more, boys in general have a relatively lower reading self-efficacy than girls (Freeman, & Garces-Bascal, 2015). The increasingly larger gender gap in reading ability offers additional insights on girls' underrepresentation in STEM subjects (Wang, Eccles, & Kenny, 2013; OECD, 2015). For instance, girls' superior ability (i.e. verbal ability) has offered them more confidence and more opportunities in non-STEM careers (Wang, Eccles, & Kenny, 2013). In contrast, as has been introduced, girls generally have a weaker motivation and performance in mathematics (Skaalvik & Skaalvik, 2004). Therefore, investigation into the gender gap in reading ability may contribute to the study of girls' low interest and performance in mathematics.

A longitudinal study coincides with the so-called paradoxical correlation between gender gap of reading ability and gender gap of mathematics ability initiated by Stoet and Geary (2013). The paradoxical association between gender gap in mathematics and reading was explored based on 10 years of PISA data (Stoet & Geary, 2013). It was found that gender gap in a STEM-related subject (i.e. mathematics) were consistently associated with gender gap in reading in an inverse way (Stoet & Geary, 2013). Countries with a small gender gap in mathematics usually have a large gender gap in reading and vice versa (Stoet & Geary, 2013). This result has been supported by the latest data, e.g. Finland, where there is a good track of mathematics performance based on PISA assessment, has the largest gender gaps of reading performance by 47 points among all participating countries (Peña-López, 2016). In this research, gender difference in science and reading will be compared.

Some researchers claim that students with good reading ability are more likely to have good achievement in mathematics (Grimm, 2008; Guthrie & Wigfield, 2000). Whether and in what way reading ability affect STEM-related subjects (e.g. science) deserves to be studied and tested against PISA 2015 data to confirm their relationship. Based on the PISA data in recent years, gender gap of reading in favor of girls is narrowed by 12 points across OECD countries from 2009 to 2015 (OECD, 2016). Boys' performance improved, especially among the highest-achieving boys, while girls' performance decreased among the lowest-achieving girls in particular (OECD, 2015). In addition, there exists a cross-national difference as to reading and STEM-related subject performance. For instance, in Finland, PISA 2012 shows a small gender gap in reading with a small gender gap in mathematics. By contrast, in counties like Chile, a narrower gender gap in reading in favor of girls is parallel with a wider gender gap in mathematics in favor of boys (OECD, 2015). The relationship between gender gap in reading performance and that of STEM subjects (i.e. science) should be further tested.

2.5 CONCEPTUAL MODEL

Based on Bandura's social cognitive theory (1986) and inspiration from widening gender gap in reading performance among 15-year-olds, a research framework of moderator effect of reading proficiency on science performance is developed (see Figure 2). It mainly focuses on the relationship between science self-efficacy and science performance and in what way reading proficiency moderates science performance across 15-year-old boys and girls. The dependent variables in the framework is science performance. Students' science self-efficacy serves as independent variable.

It is hypothesized that

- science self-efficacy is a predictor of science performance among 15-year-old boys and girls in the Netherlands, Finland, and Singapore;
- reading proficiency works as a moderator in the relationship between students' science self-efficacy and' science performance. Namely, the relationship between science self-efficacy and science performance can be weakened, strenthened, or transformed by reading proficiency among 15-year-old boys and girls in the Netherlands, Finland, and Singapore.

Accordingly, the research framework is expected to be achieved in this order.

First, gender gap is explored among the three variables, science performance, reading performance and science self-efficacy respectively.

Second, the variable of science self-efficacy is correlated with science performance in order to check the association between these two factors.

Third, the core task is to investigate the moderator effect of reading proficiency on the relationship between science self-efficacy and science performance.

Research framework is illustrated as follows:

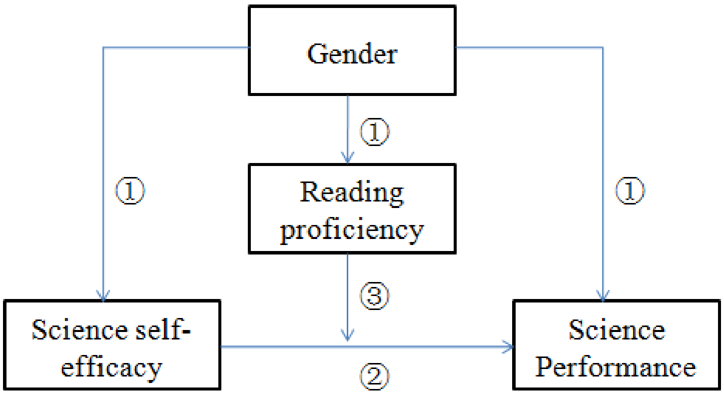


Figure 2 Research Framework

CHAPTER 3 CROSS-NATIONAL COMPARISONS

3.1 Overview

This section aims to take a a comparative perspective to explore gender differences in the Netherlands, Finland, and Singapore. In the first part, the reasons why Finland and Singapore are selected as comparison countries are described. In the second part, a comparison is conducted in terms of gender gap in science performance, reading performance, and science self-efficacy. Meanwhile, in order to deepen the understanding of the underperformance of female students in STEM subjects, comparison is made with regard to teacher and teacher quality, students engagement in science study, and parental involvement. Comparison on these factors can shed light on plausible reasons that make the difference between the three countries.

3.1.1 Finland

The primary reason to select Finland in this comparative study lies in the top performance of Finnish students among OECD countries. Finland continues to be one of the best performing countries with excellent performance in mathematics, science, and reading (“PISA 2015”, 2017). Finland has been viewed as a success story in education (Hargreaves, 2008; Hargreaves, Halász, & Pont, 2007; OECD, 2008). Its educational systems have received broad international attention from researchers and practitioners because of its widespread equity (Kupari, 2008; Linnakylä & Välijärvi 2005). It can be helpful and interesting to identify the major differences between the two countries so as to transform the inferior condition of girls a successful STEM education for the Netherlands and other countries.

The second reason to select Finland is because of its top performance of female students. Finland is a country that has girls’ performance higher than boys in all three subjects. In 2015, girls perform higher than boys by 8 points, 19 points in science and reading respectively (OECD, 2015). In addition, Unlike the Netherlands, Singapore and most of the rest of countries in which girls are higher in reading but lower in science, Finland set a good example that girls are able to surpass boys regardless of STEM or non-STEM subjects. Moreover, like the Netherlands, Finland is also a country with a vernacular language (Andere, 2015). Finnish schools, like schools in the Netherlands, enjoy much autonomy too (Doorman et al., 2007). Thus STEM education boosting measures, policies employed by Finland can shed light to the Netherlands and other countries.

3.1.2 Singapore

Unlike Finland and the Netherlands, Singapore is a non-OECD country but is also selected. This is based on the consideration that some Asian non-OECD countries stand out and perform relatively better than OECD countries. It is worthwhile to understand why and in what way Asian countries have an overall good performance in STEM-related subjects. Singaporean students achieve top scores in mathematics, science, and reading literacy consistently in the international assessment of PISA (Bautista, Wong, & Gopinathan, 2015). Top performers in science, namely, students who are proficient at Level 5 or 6, account for 8% of students across OECD countries as opposed to 24% in Singapore (OECD, 2016). In OECD countries, arround 20% of students are below the baseline level of science proficiency (OECD, 2016)

Though Singapore does not share much in common with the Netherlands on history, culture, and education system, identifying the major differences on STEM education can help researchers and policy makers understand why Asian countries top the world STEM education consistently, which would shed light to the Netherlands and other OECD countries as to improve students’ performance in STEM-related subjects and decrease gender gap in STEM education.

3.2 Comparison

Selecting appropriate perspectives to make comparisons between countries is not easy because many factors intertwined in the success story of education (Niemi, 2012; Sahlberg, 2011). They twisted all together so that it is hard to distinguish the most important one from the others (Linnakylä, 2008). Among the various factors, three perspectives were chosen to delve into the differences between the Netherlands, Singapore, and Finland. i.e. teachers and teacher quality, student study time on science, and parental involvement. These three perspectives are regarded as three of primary perspectives for consistent and high performance of STEM education (Andere, 2015).

3.2.1 Teachers & Teacher Quality

While researchers and policy makers provide many explanations for Finnish education success, according to Andere (2015) teacher and teacher quality is regarded as the primary reason accounting for the high STEM outcome in Finland. What's more, teachers and teachers' quality is selected as a perspective for comparison lies also in the significance of teachers' role in students' self-efficacy beliefs. It is claimed that the teachers play a significant role in instilling perceptions of self-efficacy in their students (Sharma & Nasa, 2014). While in what way perceptions of self-efficacy is instilled in students is not the focus of this project, significance of teachers and their quality cannot be ignored in success of students' science

In most European and Asian countries teachers enjoy high status, which was explicitly emphasized in the consultants' reports of countries like Finland, Singapore, France, Chinese, Korean, and Japanese (Marginson et al., 2013). But the required educational degree and competitiveness of the teaching profession vary across countries. In Finland, teaching profession is among the most popular and competitive professions (Laukkanen, 2008; McKinsey & Company, 2010). Its selection is highly competitive. Primary school teachers should not only have a discipline-specific master's degree but also enrolled in university pre-service programs to be trained as a highly qualified teacher (Parveva et al., 2011). Only the most talented and dedicated candidates, around one out of ten, could finally become a teacher because of its intense popularity and competition (Parveva et al., 2011).

Identical with Finland, in Singapore, teaching profession is also highly respected (Bautista, Wong & Gopinathan, 2015). It is regarded as a phenomenon influenced by Confucian traditions. The Finland consultants' report also emphasizes high respect for teachers (Parveva et al., 2011). Recruiting is strict with various strategies in order to guarantee high teacher quality and select the best candidates for primary and secondary schools (Darling-Hammond, 2010). Candidates are selected from each cohort of graduating students with highest performance. After a thorough application process and a series of examinations, consequently, only about one out of eight of the candidates are accepted in Singapore (Darling-Hammond, 2010).

In the Netherlands, primary school teachers do not necessarily have a high educational degree. In PISA 2015, only around 30% of science teachers of 15-year-old students in the Netherlands have a University degree. This may gender some problems such as confidence in teaching STEM subjects (Parveva et al., 2011). Only the first-level qualified teachers are less likely to lack confidence in STEM subjects. However, the second-level qualified teachers are likely to suffer from low confidence in STEM subjects (Parveva et al., 2011). In addition, Dutch teachers are found to have higher gender-stereotypes associations than other countries (Cheryan, Siy, Vichayapai, Drury & Kim, 2011). Teachers in the Netherlands may not motivate girls to engage in STEM subjects. Boys and girls with the same level in mathematics and physics would probably receive different advice from career advisors (VHTO 2008). In comparison with it, for instance, America girls are more likely to pursue STEM majors because teachers are expected to motivate and assist students to pursue what they desire (Bieri Buschor, Berweger, Keck Frei & Kappler, 2014).

The following figure shows the difference between the Netherlands, Finland and Singapore with regard to percentage of science teachers of 15-year-old students with certification and university diploma. The Netherlands is lagging behind Finland, Singapore, OECD average and the world average in terms of the percentage of university degrees. As for proportion of teachers' certification, Finland

and Singapore take the lead. The Netherlands is slightly lower than the average of OECD countries and 72 countries.

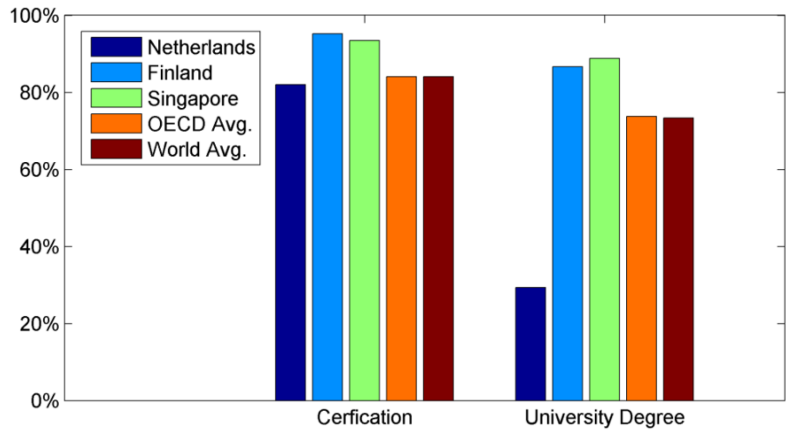


Figure 3 Teachers' certification and University degrees

In contrast with the high competitivity of teaching profession in Finland and Singapore, popularity of teaching profession in the Netherlands are not high. PISA 2015 questionnaire among school principals provides convincing complementary for it. Result shows that the Netherlands are prominent with regard to "the lack of teaching staff". Around 61% of principals in Finland and 46% Singaporean principals say that they are not lack of teaching staff at all. However, the Netherlands only 32% of principals agree that they are not lack of teaching staff, lagging behind the OECD average (38%). Therefore, the relatively low educational degree of teachers and the lack of teaching staff in the Netherlands make a dilemma. Namely, the lack of teaching staff is hard to be resolved with a weak popularity of teaching profession. This can be detrimental to Dutch students' STEM subject performance improvement without tackling this dilemma.

Table 2

A lack of teaching staff

	A lack of teaching staff							
	Not at all		Very little		To some extent		A lot	
	%		%		%		%	S.E.
		S.E.		S.E.		S.E.		
Finland	60.5	(4.4)	36.7	(4.4)	2.8	(1.2)	0.0	C
Netherlands	32.2	(4.0)	40.7	(4.0)	24.9	(4.0)	2.1	(1.5)
Singapore	45.9	(1.2)	43.4	(1.2)	10.7	(0.1)	0.0	C
OECD average	38.8	(0.5)	31.7	(0.5)	25.2	(0.5)	4.3	(0.2)

Note. Extracted from PISA online dataset ("PISA 2015: full selection of indicators", 2017)

3.2.2 Parental involvement

Parental involvement is selected to make a comparison between the three countries considering its close relationship with academic performance and self-efficacy.

It has been recognized that "parents have significant impacts on students' learning and developmental processes" (Fan & Williams, 2010, p. 53). The role of parental involvement in students' academic success has been an important issue in educational research and policy (Kernan, 2012). Increasing research findings show the existence of a positive relationship between parental involvement and educational performance (Catsambis, 2001; Kernan, 2012; Wang & Sheikh- Khalil, 2014). Based on a meta-analysis, this positive relationship has been found in 50 studies (Hill & Tyson, 2009). This implies students whose parents involve themselves in children's education would achieve higher learning outcomes in mathematics (Sirvani, 2007; Yan & Lin, 2005). It was found that parental involvement can influence children's achievement in mathematics (Sanders & Sheldon, 2009; Yan & Lin, 2005). Fan and Chen (2001) found out the strong association between parental involvement and student learning achievement and further distinguished the type of parental involvement which is beneficial to positive student learning attitude. More than that, it is found that parental involvement in students' learning is related with other respects including more positive attitude towards school and higher self-regulation (Fan & Chen, 2001).

Furthermore, in a study concerning parental involvement and students' self-efficacy among 10th grade students, it is found that parental advice can predict students' academic self-efficacy in English in a positive way. Sanders and Sheldon (2009) expanded the benefits of parental involvement. They found that parental involvement could affect children's reading, mathematics and science achievement. In line with it, it is found that parental involvement and educational outcome is related with each other regardless of students' socioeconomic background, ethnicity and type of school (Catsambis, 2001).Therefore, it could be predicted that parental involvement is an important strategy to enhance education quality (Driessen, Smit & Sleegers, 2005). Many schools would make effort to promote parental participation so as to increase students' academic performance and reduce educational inequity (Kerman, 2012).

Strong evidence has been found among Singaporean parents as evidenced in the findings by Raytheon's (2010) study. The high parental involvement paid off in the PISA assessment. Different from the other countries in Europe and America, Singapore demonstrates a higher parental support. It is reported that 51% of parents in Singapore assist children with mathematics by extra instruction from educators on practical methods. 42% of Singaporean parents would employ family tutors to help children with mathematics. In addition, 92% of Singapore students are engaged in some form of after-school mathematics learning (Raytheon, 2010). Moreover, according to Raytheon's report, 26 percent of Singapore students participated in different types of extracurricular activities (i.e. competitions, games, and camps etc) (Raytheon, 2010). Besides these, more types of parental involvement activities appear in Singapore in recent years. For instance, at least one orientation day for new entrants is in most primary schools. Workshops and briefing sessions are held in some schools on curriculum developments and the way that parents might assist their children's learning at home (Clarke, 2001). In this way, Singapore is the only country where parents receive instructions on how to assist children at home among the three countries.

As for Finland and the Netherlands, parents are very much involved in both of the two countries and they both have a long history of encouraging parental involvement. Compared with the Netherlands, Finland is among the earliest countries to have legislation that parents particiapte in school education system since 1931 (Parveva et al., 2011). A close cooperation wih family was set as a goal in 1957 Act, which regulated that the primary schools should be run and supervised by a school management board comprised by parents. The Netherlands has a shorter history compared with Finland and other European countries, there was not legislation of parental participation until 1981 which is 50 years later than Finland and 10 years later than the majority of European members ("EURYDICE," 1997). In 1992, a clear distinction was set between rights and role of parents and teachers (ib.id). In both countries, there are various types of associations for parents to play the participation role, namely, 4 major associations in the Netherlands and over 800 local parents' associations called Home and School Associations in Finland. Overall, the major parental participation focused on consultation concerning school issues and activities at school level.

However, parental involvement demonstrates differences especially in decision-making power. Parents do not have equal powers across the two countries. In Finland, parents not only participate in consultation of curriculum development, school atmosphere enhancement and family-school cooperation but implement some practical power, e.g. school timetables drawing, school work plans drawing, and school curriculum drawing within the national framework (ib.id). In the Netherlands, It is open for parents to express their opinion and consultation on most issues about education and school both at school level and national level. But parents play little impact on the system. In the Netherlands, the responsibility of curriculum and examinations are taken by the state. Rights of the parents were clearly defined and restricted. Parents' participation in the work of the relevant authorities is minor. The major forms of parental involvement are assistance and support on school activities, school travel and cleaning. Parents find them in highly dominated position and they become "objects that help professionals to cope with their own disabling position" (van den Berg & van Reekum, 2011).

3.2.3 Study time after-class & regular lessons

It is necessary to get acknowledge of the science study time across the three countries.There is an interesting phenomenon that relationship between study time and academic achievement is positive in some countries, such as Korea, Canada and Singapore (Kuehn & Landeras, 2012), while in other countries such as Finland, students spend less time in regular lesson and after-school learning but receive high achievement (see Figure 4). In addition, it was found that in Mexico, Turkey, and Greece, students spend more time studying science than students from better performing countries but ranked the worst among OECD countries based on PISA data in 2006. (Kuehn & Landeras, 2012). Figure 3 is the average study hour in science in regular lesson and after school in the three countries in 2015, which shows a consistent result with 2006 as observed by Kuehn and Landeras (2012) who found out a positive relationship between study time and performance for Singapore but negative relationship for Finland. Specifically, in Finland, where students top in science performance but with far less study time both in regular lessons and after school. Study time in Singapore tops among the three countries with the top performance among the three countries.

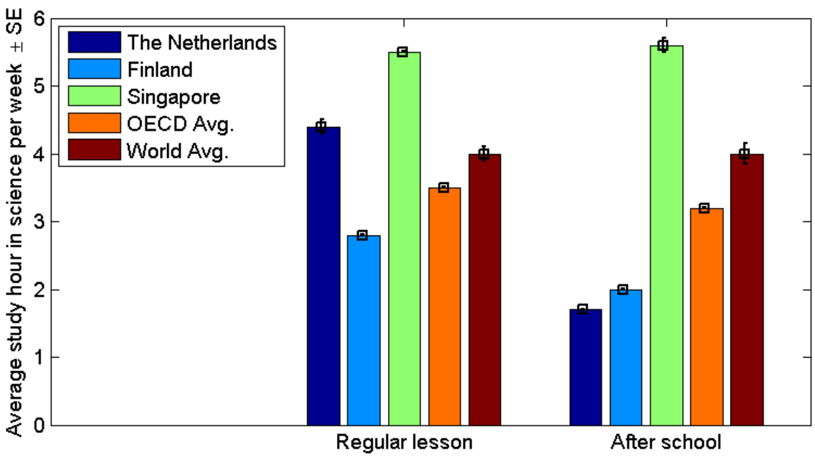


Figure 4 Average science study hour in regular lesson and after school

It should be mentioned that the comparison from the above three perspectives allows an overall understanding on the differences between the three countries. However, no further statistical analysis is conducted since it mainly aims to provide background knowledge for the subsequent quantitative analysis and shed light for the plausible reasons that might bring about the differences between the countries. They are not the focus of this project though each perspective can make a unique topic in the exploration of girls' underperformance in STEM-related subjects by cross-national comparison.

CHAPTER 4 METHODS

Since the project is based on the latest PISA data in 2015. An introduction to PISA is given prior to the methods. Respondents, sampling methods, instruments, research design, and data analysis plan are described after the introduction.

4.1 What is PISA?

PISA, acronym for the Program for International Student Assessment, is a triennial survey launched by OECD countries in 1997 aiming to assess 15-year-old students' core school subjects of reading, mathematics and science (OECD, 2016). The goal of PISA is to assess to what extent students at the end of their compulsory education have gained the required knowledge and skills that are important for their later participation in modern societies (OECD, 2016). The assessment does not only focus on what students know or reproduce knowledge but also to what extent they can apply the knowledge in different contexts. Each triennial assessment focuses on one domain. In 2015, the focus of assessment is on science. Around 540 000 students participated in the assessment in 2015, representing approximately 29 million 15-year-old students of the 72 participating countries (OECD, 2017).

PISA is unique with its comprehensiveness in assessing international students' performance and collecting plausible reasons on family and institutional factors that may help explain differences (OECD, 2016). Namely, aside of a 2-hour test on three domains, questionnaires are also finished by students, teachers, principals, and/or parents to gather information on students' family background, learning environment and learning approaches. Now not only OECD countries but also many other countries participate in the PISA assessment based on an internationally agreed framework. While PISA cannot demonstrate a cause-and-effect relationship between student outcomes and educational policies, PISA findings have been used by policy makers around the world to understand the strengths and weaknesses of one's own education systems, set benchmarks for improvement, and/or make comparison with other economies in terms of education policies, students' learning outcomes, teaching quality etc. PISA data as well as its reports are available on PISA official website (http://www.oecd.org/pisa/) from which the data could be downloaded.

4.2 Respondents and sampling

The respondents for this study are 15-year-old students studying in the Netherlands (N=5385, 2700 females, 2685 males), Finland (N=5582, 2863 females, 3019 males), and Singapore (N=6115, 2973 females, 3142 males). The age of respondents is between 15 years 3 months and 16 years 2 months at the time of the test ("Background and basics," 2016). They are randomly selected within schools so they have different learning environments, teachers and level of instruction. Respondents sampling follows a strict technical standards. For instance, in order to lower the bias, school teachers are not allowed to inspect their own students (OECD, 2016).

PISA does not draw simple random samples of students from all the 15-year-old student lists. Instead, it is a two-stage sampling (OECD, 2016). Namely, first, a sample of schools is selected from a complete list of schools. Then, a simple random sample of students or classes is drawn from the selected schools. Usually 35 students from the population of 15-year-olds are randomly selected. If less than 35 15-year-old students attend a selected school, then all of the students would be invited to participate (OECD, 2009). In this way, the selection of schools and students can be kept as inclusive as possible (OECD, 2016). Further information on the implementation of PISA sampling can be found in the PISA 2015 Technical Report (OECD, 2016).

4.3 Instruments

4.3.1 PISA Tests

Science test

The science test, the focus of PISA 2015 (OECD, 2016), is a total of 1-hour computer-based test. The aim is to test students' scientific literacy, such as the scientific thinking and scientific discovery. It requires learners not only have the knowledge of concepts, how they are derived, and how they can advance the science (OECD, 2016). As defined by PISA 2015, scientific literacy refers to the competency to:

- explain phenomena scientifically
- evaluate and design scientific inquiry
- interpret data and evidence scientifically

(OECD, 2016, p. 19)

Specific science test items and test criteria for PISA 2015 and previous years can be accessed in the PISA website (http://www.oecd.org/pisa/test/). Specific description of scientific literacy and desired distribution of items by the three competencies is depicted in chapter 2 of PISA analytical framework (OECD, 2016). Seven proficiency levels in science are classified in PISA 2015 (see Appendix A). A sample of science test items is listed (see Appendix B). Moreover, the test items are a mixture of multiple-choice questions, for which students are required to construct their own responses (OECD, 2016). Around 810 minutes of test items were covered for science, mathematics, reading and collaborative problem solving, with different students from different countries taking different combinations of test items (OECD, 2016).

Reading test

Reading literacy is a minor domain assessed in PISA 2015 (OECD, 2016). It is also a computer-based 1-hour test to test students' reading literacy, which is defined by PISA 2015 as "understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, develop one's knowledge and potential, and participate in society" (OECD, 2016, p. 49). Detailed information on reading literacy, please check Chapter 3 of PISA 2015 Analytical Framework (OECD, 2016). Seven proficiency levels in reading are classified in PISA 2015 (see Appendix C). A sample of computer-based reading test item is provided (see Appendix D).

4.3.2 Questionnaire

Student questionnaire

Information is gathered concerning students' self-efficacy and background information in the form of student questionnaire (OECD, 2016). Students were asked to complete a questionnaire which lasts 35 to 45 minutes. Detailed information is collected concerning students' family backgrounds, learning engagement, science self-efficacy, instruction approaches, and so on (OECD, 2016). For instance, students were asked to indicate how many hours they spend each week on regular science lessons and science learning after school. As for the concept of science self-efficacy, 8 items were given to allow students to indicate to what extent they have confidence in resolving the scientific problems. These data are especially for the exploration on the association between reading performance and science performance as well as the moderator effect of reading on science self-efficacy. Data about the questionnaire is available on official website of OECD (http://www.oecd.org）.

4.4 VARIABLES

4.4.1 Science self-efficacy

Students' science self-efficacy is taken as independent variable in this research. It was measured in student questionnaire by asking students to indicate his/her confidence level towards eight science-related items (see Appendix E). For instance, students were asked to indicate the confidence level with the item "identify the better of two explanations for the formation of acid rain". The answer is an ordinal variable with a 4-point Linkert scale ranging from "I could do this easily" to "I couldn't do this".

4.4.2 Students' science performance

Students' science performance is a continuous and dependent variable in this research. It should be noted that cognitive data in PISA are scaled with the Rasch Model and performance is not represented by EAP scores but denoted by plausible values. Precisely, the science test results are not represented in traditional scores but 1 to 10 plausible values (PVs). This implies that the average of the ten plausible values cannot be calculated to process the data which will make a fatal error (OECD, 2016) and will consequently underestimate the standard deviation (Monseur, 2009). Plausible values are dealt according to the manual book of OECD so as to be valid and minimize the bias (OECD, 2016).

4.4.3 Students' reading proficiency

Identical with science performance, reading performance is also denoted by plausible value 1 to 10. Students' reading proficiency is classified into three levels in order to check whether reading proficiency can moderate students' science performance. Specifically, the reading performance of 15-year-old students from the the Netherlands, Finland, and Singapore are classified into three levels: high level, medium level and low level. The cutpoint is based on seven levels of reading proficiency of PISA (see Appendix C). The classification follows PISA assessment in which performance lower than level 2 (reading score <407.47) is defined as low level.) Reading level 5 and 6 (reading score > 625.61) is defined as high level (PISA, 2016). The left is medium level.

4.5 Data analysis & steps

4.5.1 Research design

This is an exploratory research which is mainly comprised of three tasks based on the three research questions. The first task is descriptive comparison across boys and girls between the Netherlands, Finland and Singapore in terms of gender gap in science performance, science self-efficacy, and reading performance. The comparison is mainly 1) to check the gender gap of 15-year-olds in the subjects of reading and science as well as science self-efficacy; 2) gender gap in reading and science performance is also compared to check the strengths and weaknesses of each gender. Meanwhile, differences between the Netherlands and the other two top-performing countries (i.e. Finland and Singapore) with good tracks of educational equity and science performance are also revealed in this process.

The second task is a correlational study between science self-efficacy and science performance. This is prerequisite for the third task, the regression analysis to check the moderator effect of reading proficiency on science performance. Namely, to what extent reading proficiency can weaken, strengthen or transform the relationship between science self-efficacy and science performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore. In this process, to what extent gender interacts with reading in the prediction of science performance among 15-year-old students in the three countries is checked. Therefore, statistical analysis is a primary tool and complementary to the former literature study. Whether the superior ability of girls in reading would change the relationship between science self-efficacy and science performance would be revealed in the quantitative data.

The following Table provides a readers' matrix on methods for research questions and in which chapter they can be accessed.

Table 3
Readers guide on research questions and methods

	Research Questions	**Method**	**Chapter**
RQ1	To what extent is science performance, reading performance and science self-efficacy different across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA 2015 data?	Literature review Two-way ANOVA Linear regression	Chapter 2 Chapter 5
RQ2	To what extent is science self-efficacy correlated with science performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA 2015 data?	Literature review Correlation	Chapter 2 Chapter 5
RQ3	To what extent does reading proficiency moderate the relationship between science self-efficacy and science	Literature review Linear regression	Chapter 2 Chapter 5

	performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA data 2015?

4.5.2 Data Analysis & Procedures

Data analysis is conducted in SPSS programming software (IBM Statistics Version 22) and IEA IDB Analyzer (Version 4.0.11) in order to address to what extent 15-year-old girls' reading proficiency moderates their their science performance. ANOVA is used to check gender gap in science performance among 15-year-old students in the three countries. Regression analysis is conducted to test the moderator effect of girls' reading proficiency on science performance. Major effect of reading proficiency and interaction effect between reading proficiency and science self-efficacy is also checked in regression analysis.

IDB Analyzer is introduced with the purpose of dealing with plausible values and weights in the data in an appropriate way. Weights are associated to each subject since students in a particular country may not have the equal probability to be selected. The variables with PVs such as science performance can be processed in IDB Analyzer with which a syntax can be directed to SPSS programming for further processing. As for the sampling weight, The IDB Analyzer can select the weight and replication variables automatically.

Steps

Data is processed in 4 steps. Step 1 is data preparation. Data for the three countries of the Netherlands, Finland, and Singapore is selected and extracted using SPSS from the datafile which was downloaded from OECD official website. There are 857 out of 17382 cases deleted because they are labeled invalid or missing. In addition, to cope with the requirement of regression analysis, the categorical variable gender is dummy coded to 0 (female) and 1 (male). The continuous variable reading performance PV1-10 and science performance PV1-10 are standardized in SPSS.

Step 2 is to make descriptive tables and charts in order to get an overview of gender gap in science, reading performance and self-efficacy compared with the average of OECD countries. Meanwhile, in order to check the distribution of science self-efficacy in different reading proficiency level, reading performance is transferred from continuous variablle to categorical variable by classifying into three levels separately (i.e. HIGH, MEDIUM, and LOW). Classification criteria is based on PISA standard in dealing with the data. Specifically, reading performance is originally classified into 7 levels from level 1b to level 6. Specific criteria for each level and scores are demonstrated (see Appendix C). According to PISA 2015 report (OECD, 2016), level 1b to level 2 is classified as low level. Level 3 and level 4 are classified as medium level. Level 5 and level 6 are classified as high level. Specific classifying criteria for each level can also be found in the Appendix of the Help file for IDB Analyzer. The cut point for the three levels are 407.47 and 625.61.

Step 3 is assumption checks. Prior to the regression analysis, assumptions must be met in order to make the analysis valid. The assumption check is comprised of 6 sub-steps. Namely, the continuousness of the dependent variable, the continuousness of independent variable, independent observations, linear relationship between dependent variable of science performance and independent variable of science self-efficacy in three reading level groups and gender groups, Homoscedasticity aiming to check whether the error variances are identical for all combination of dependent variable of science performance and independent variables of science self-efficacy, reading level, and gender. Multi-collinearity was checked since there are three independent variables. No significant outliers were found. Standardized deleted residuals and ANOVA tests are checked too. Approximate normality is met for residuals.

Step 4 is to use SPSS programming and IDB Analyzer to conduct descriptive analysis and regression analysis. In order to address the first research question, means of science performance PV1 to PV10 and reading performance PV1 to PV10 are calculated in IDB Analyzer to show gender difference in raw scores. Gender gap of science self-efficacy is plotted across the three countries in two-way

ANOVA using science self-efficacy as the dependent variable. The following figure represents a statistical model in the research. Specific results can be seen in Chapter 5.

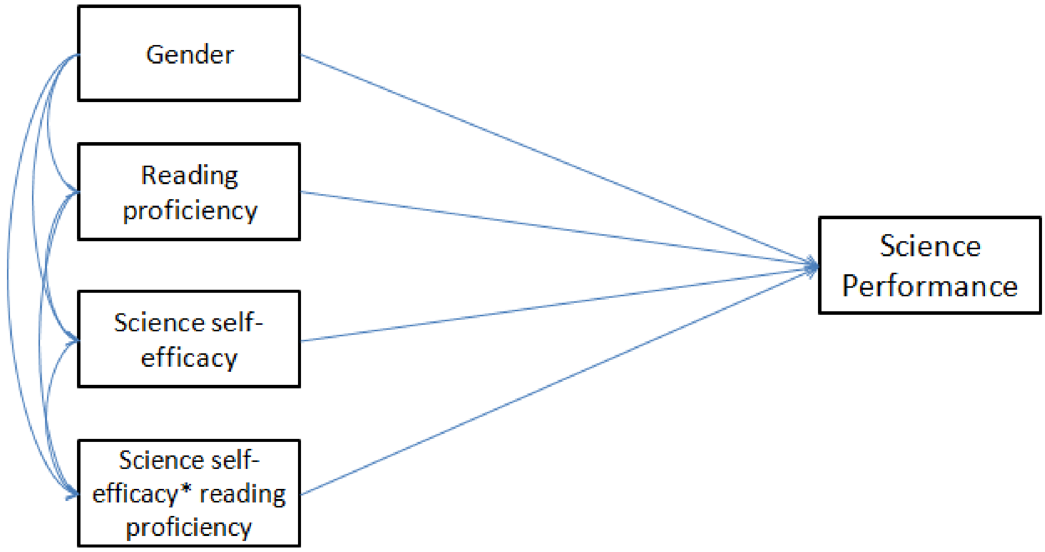


Figure 5. Statistical model of the research framework

CHAPTER 5 RESULTS

The research results are described in this chapter. The following figure 6 shows the overall research results for each research question. At the bottom of the figure, the value of "R" represents correlation coefficient between science self-efficacy and science performance.

Further information is elaborated in this chapter.

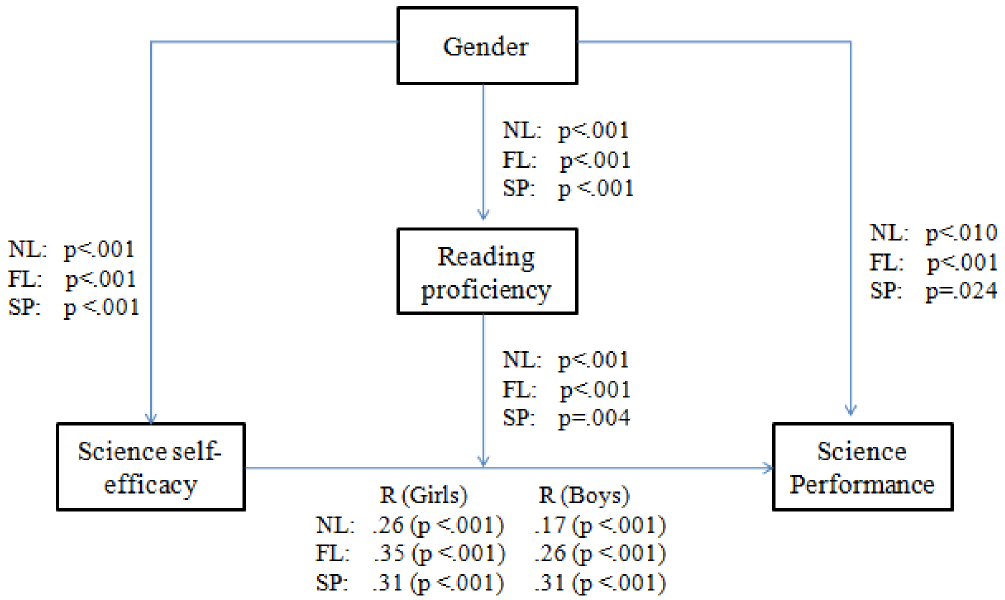


Figure 6 Overview of the research results

It should be mentioned that, in order to keep uniform of the data results and address the limitation of SPSS which does not allow the calculation for PV1-10 once for all, the analysis for science and reading performance are calculated from PV1 to PV10 separately. The interpretation is mainly based on PV1. Results for PV2-10 are given in Appendix. Considering the results for PV2-10 in each analysis is different with PV1 with little difference, it will not influence the determination and judgment for the result interpretation of each research question.

5.1 GENDER GAP

5.1.1 Gender gap in science performance & reading performance

RQ1. To what extent is science performance and reading performance different across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA 2015 data?

Means and gender difference of science and reading performance for PV1 across the three countries are shown in Table 4. Results for PV2-10 for science are shown in Appendix F. There exists gender difference both in science and reading performance with different degrees. In addition, gender difference in reading performance is strikingly higher than science performance across the three countries. Statistical significance exists on gender difference in both science and reading performance for PV1 (see Table 4). Statistical significance also exists in science performance from PV2 to PV10 (see Appendix G).

As for country difference, gender difference exists in the three countries with different degrees. In Table 4, it can be seen that Finland is different from the other two countries with its largest gender gap in favor of girls. Namely, Finnish girls surpass boys in both reading and science with the largest gender difference compared with the other two countries. The mean difference in reading is -45.20 in favor of girls in Finland as opposed to -18.81 in the Netherlands and -17.22 in Singapore. The mean difference in science performance is the largest with -15.08 in Finland, 7.02 in the Netherlands and 2.71 in Singapore.

Table 4

Mean of science and reading performance by gender (PV1)

		Gender	N	Mean	SD	Mean difference (Boys-Girls)
Finland	Science	Female	2707	544.97	89.17	-15.08
		Male	2763	529.89	97.97	
	Reading	Female	2707	555.95	80.82	-45.20
		Male	2763	510.75	90.36	
Netherlands	Science	Female	2570	513.11	93.54	7.02
		Male	2486	520.13	100.42	
	Reading	Female	2570	521.07	90.31	-18.81
		Male	2486	502.26	96.29	
Singapore	Science	Female	2921	554.08	97.67	5.96
		Male	3078	560.04	106.72	
	Reading	Female	2921	543.98	94.46	-17.22
		Male	3078	526.76	100.92	

Note. Weighted Mean & Standard Deviation-Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGHT.

Linear regression is conducted to check whether the gender gap of science and reading performance in the three countries is significant (see Table 5).

Result shows that boys' science performance is significantly higher than girls in the Netherlands ($p = .010$) and Singapore ($p = .024$) and significantly lower than girls in Finland ($p < .001$) for PV1. Statistical significance for science performance also exists from PV2 to PV10 in science performance (see Appendix 7). Boys' reading performance is consistently and significantly lower than girls in Finland ($p < .001$), the Netherlands ($p < .001$), and Singapore ($p < .001$) for PV1. Statistical significant results for reading performance are also shown from PV2 to PV10 (see Appendix H).

Table 5

Significance of gender difference in science performance (PV1) and reading performance (PV1)

	Country	Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	T	Sig.
Finland	Science	1	(Constant)	560.06	4.03		138.92	.00
			Gender	-15.08	2.54	-.08	-5.95	.000
	Reading	1	(Constant)	601.15	3.69		162.93	.000
			Gender	-45.20	2.32	-.26	-19.47	.000
Netherlands	Science	1	(Constant)	506.09	4.29		117.90	.000
			Gender	7.02	2.73	.04	2.57	.010
	Reading		(Constant)	539.87	4.13		130.74	.000
			Gender	-18.81	2.63	-.10	-7.16	.000
Singapore	Science	1	(Constant)	548.13	4.23		129.68	.000
			Gender	5.96	2.65	.029	2.25	.024
	Reading		(Constant)	561.20	4.04		139.01	.000
			Gender	-17.22	2.53	-.09	-6.81	.000

Note. a. Weighted Least Squares Regression-Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGHT b. Science performance: R Squared (Finland) = .060, R Squared (Netherlands) = .010, R Squared (Singapore) = .010. d. Reading performance: R Squared (Finland) = .065, R Squared (Netherlands) = .010, R Squared (Singapore) = .080

5.1.2 Gender gap in science self-efficacy

RQ1. To what extent is science self-efficacy different across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA 2015 data?

Gender gap in science self-efficacy is first explored by means and standard deviations in the following Table 6. Science self-efficacy of girls is consistently lower than boys across the three countries with the largest gender gap in the Netherlands, followed by Finland and Singapore.

Table 6
Gender difference in science self-efficacy in the three countries

Country	Gender	N.	Mean	S.D.	Mean difference (boys - girls).
Finland	Female	2707	-.175	1.074	.262
	Male	2763	.087	1.264	
Netherlands	Female	2570	-.208	1.254	.269
	Male	2488	.061	1.264	
Singapore	Female	2921	-.003	1.053	.212
	Male	3078	.209	1.209	

Note. Weighted Least Squares Regression - Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGHT

Significance of gender gap in science self-efficacy is demonstrated in Table 7, which shows statistically significant gender gap in science self-efficacy in Finland (p < .001), the Netherlands (p < .001), and Singapore (p < .001).

Table 7
Significance of gender difference of science self-efficacy in the three countries

	Model	Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	T	Sig.
NL	(Constant)	-.477	.056		-8.559	.000
	Gender	.269	.035	.106	7.596	.000
FL	(Constant)	-.437	.051		-8.651	.000
	Gender	.262	.032	.111	8.242	.000
SP	(Constant)	-.215	.047		-4.595	.000
	Gender	.212	.029	.093	7.222	.000

Note. a. Dependent Variable: Science self-efficacy (WLE); b. Weighted Least Squares Regression - Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGH; c. R Squared (Finland) = .120 c. R Squared (Netherlands) = .110 d. R Squared (Singapore) = .090.

Plotted result in Figure 7 gives a visual image on gender gap in science self-efficacy across the three countries. It shows that Singaporean students surpass Finnish and Dutch students in terms of students' science self-efficacy. Dutch students' science-self-efficacy is the lowest among the three countries. As for the gender difference, the science self-efficacy of male students is systematically higher than female students across the three countries. Furthermore, Dutch girls' science self-efficacy is the lowest among the girls in the three countries.

To summarize, in response with the RQ1, girls' science self-efficacy is significantly higher than girls across Finland ($p < .001$), the Netherlands ($p < .001$), and Singapore ($p < .001$).

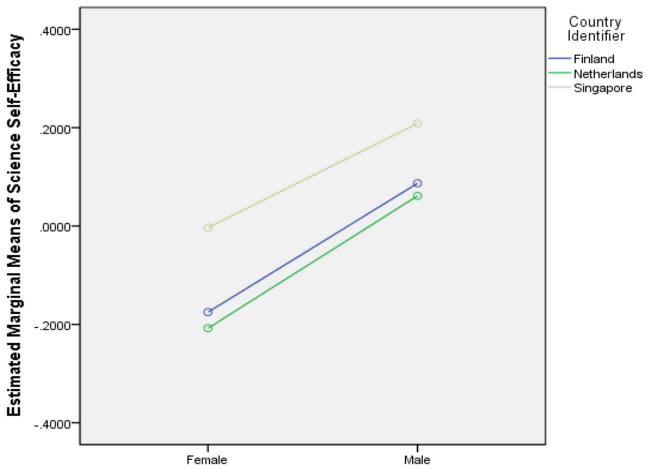


Figure 7 Science self-efficacy by gender in the three countries

5.2 Association between science self-efficacy and science performance

The second task for this project is to check the relationship between science self-efficacy and science performance as illustrated in research question 2.

RQ2. To what extent is science self-efficacy correlated with science performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA 2015 data?

The following table shows that science self-efficacy is positively and significantly related with science performance among 15-year-old students in the three countries. The relationship is higher among girls than boys within each of the three countries. It is highest among Finnish girls with .35, followed by Singaporean female students. The correlation for female students in the Netherlands is lower than Singapore and Finland.

Table 8

Significance & correlation between science self-efficacy & science performance (PV1-10) by gender

Country	Gender	N.	Correlation	S.E.	Sig. (2-tailed)
Netherlands	Female	2570	.26	.03	.000
	Male	2486	.17	.03	.000
Finland	Female	2707	.35	.02	.000
	Male	2763	.26	.02	.000
Singapore	Female	2921	.31	.02	.000
	Male	3078	.30	.02	.000

5.3 Moderator effect of reading proficiency

Since the positive relationship between science self-efficacy in research question 2 is identified, the final task is the moderator effect in RQ3.

RQ3. To what extent does reading proficiency weaken, strengthen or transform the relationship between science self-efficacy and science performance across 15-year-old boys and girls in the Netherlands, Finland and Singapore based on PISA data 2015?

Moderated regression results for PV1 are shown in Table 9. Model 1 represents regression model summary for the main effect from three variables of gender, science self-efficacy, and reading proficiency. Interaction effect between reading proficiency and science self-efficacy is represented in model 2. It should be noted that science and reading performance in this regression model is centered in convenience of the interpretation and comparison of the data. Results of PV2 to PV10 are presented in Appendix I. It not only reveals a significant major effect of science self-efficacy ($p < .001$) and significant gender effect ($p < .001$) on students' science performance across the three countries. Interaction effect between science self-efficacy and reading performance is also significant for the Netherlands ($p < .001$), Finland ($p < .001$) and Singapore ($p = .004$).

The moderator effect of reading proficiency on the relationship between science self-efficacy and science performance is satisfied. The result shows that as the reading proficiency increases, the correlation between science self-efficacy and science performance increases. Precisely, the standardized coefficient for science self-efficacy systematically increases across Finland (from .073 to .083), the Netherlands (from .093 to .113) and Singapore (from .067 to .068) as the interaction effect is added into the regression model. Moreover, Singaporean students demonstrate the least moderation effect of reading proficiency on science performance among the three countries.

Table 9

Regression of reading proficiency on science performance (PV1)

Country	Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	T	Sig.
FL	1	(Constant)	-.177	.009		-19.256	.000
		Dummy_Gender	.250	.013	.135	19.501	.000
		Science self-efficacy	.057	.005	.073	10.665	.000
		Zscore: PV 1 in Reading	.894	.007	.877	124.174	.000
	2	(Constant)	-.183	.009		-19.950	.000
		Dummy_Gender	.244	.013	.132	19.133	.000
		Science self-efficacy	.065	.005	.083	11.934	.000
		Zscore: PV 1 in Reading	.891	.007	.874	124.309	.000
		Read * Efficacy	.040	.005	.053	8.032	.000
NL	1	(Constant)	-.167	.008		-20.264	.000
		Dummy_Gender	.220	.012	.115	18.572	.000
		Zscore: PV 1 in Reading	.882	.006	.886	142.442	.000
		Science self-efficacy	.070	.005	.093	14.917	.000
	2	(Constant)	-.172	.008		-20.832	.000
		Dummy_Gender	.218	.012	.114	18.518	.000
		Zscore: PV 1 in Reading	.879	.006	.883	142.404	.000
		Science self-efficacy	.085	.005	.113	16.674	.000
		Read * Efficacy	.036	.005	.049	7.306	.000
SP	1	(Constant)	.010	.008		1.272	.203
		Dummy_Gender	.205	.011	.102	19.459	.000
		Zscore: PV1 in Reading	.902	.005	.899	167.058	.000

	Science self-efficacy	.059	.005	.067	12.477	.000
2	(Constant)	.007	.008		.873	.383
	Dummy_Gender	.204	.011	.101	19.373	.000
	Zscore: PV1 in Reading	.900	.005	.897	165.187	.000
	Science self-efficacy	.060	.005	.068	12.556	.000
	Read * Efficacy	.012	.004	.015	2.866	.004

Note. a. Dependent Variable: Zscore Plausible Value 1 in Science b. Weighted Least Squares Regression: Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGHT c.

It should be noted that ANOVA tests of regression residual are checked to guarantee the validity of the moderator analysis, (see Appendix J). Results show that model 1 without the interaction is significant with $F(3, 5466) = 5868.828$, $p < .001$. Model 2 in which interaction is added, is also significant with $F(4, 5465) = 4468.892$, $p < .001$. Significant results were also shown for the Netherlands and Singapore (see Table 9).

R^2 is checked with satisfactory results in model statistics of regression (see Table 10). Reliable and satisfactory results are demonstrated for PV2 to PV10 too (see Appendix K). Specifically, the following table shows that in model 1, the main effect of the three variables accounted for 76.3%, 81.1%, and 84.0% of variance on students' science performance in Finland, the Netherlands, and Singapore respectively. Model 2 with the interaction effect accounted for additional variance than model 1 without including the interaction significantly. Namely, R^2 increases as interaction is added in the model. (Finland: R^2 change = .003, $p < .001$; the Netherlands: R^2 change = .002, $p < .001$; Singapore: R^2 change < .001, $p < .001$). This indicates that potentially there is significant moderation between reading proficiency and science self-efficacy on students' science performance.

Table 10
Model Statistics summary of the regression model

					Changed statistics
	Model	R	R^2	Adjusted R^2	R^2 Change
FL	1	.874[a]	.763	.763	.763
	2	.875[b]	.766	.766	.003
NL	1	.900[a]	.811	.810	.811
	2	.901[b]	.813	.812	.002
SP	1	.917[a]	.840	.840	.840
	2	.917[b]	.840	.840	.000

The following figure also shows moderator effect of reading proficiency in a different way. As reading performance increases, the correlation between science self-efficacy and science performance increases systematically across the three countries. At a low reading level the correlation is very weak but turns out to be stronger as the reading proficiency increases. Therefore, based on the regression result, reading proficiency significantly moderates the relationship between science self-efficacy and science performance.

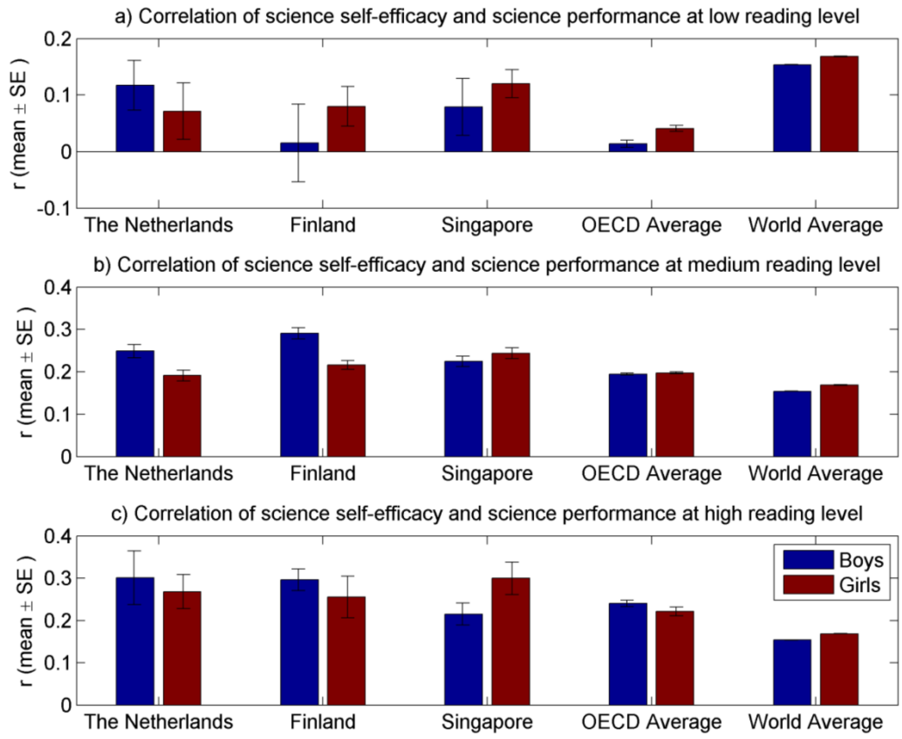


Figure 8 Correlation in different reading proficiency level

CHAPTER 6 DISCUSSION

The major research results are discussed in this part, gender gap in science performance, science self-efficacy, the relationship between science self-efficacy and science performance, and the moderator effect of reading proficiency on science performance across 15-year-old students in the three countries. It should be mentioned that no cause-and-effect relationship can be established between science self-efficacy and science performance. But science self-efficacy and the moderator effect can provide further insights on gender gap in science performance.

On the other hand, important research results are discussed in combination with the comparison study from the three perspectives of teacher quality, students' study time, and parental involvement that have been described in Chapter 3. It should also be noted that the three perspectives explored in Chapter 3 mainly aims to provide a rich background for the three countries and shed light to gender difference in science performance. Though no cause-and-effect relationship can be set up between gender gap in science performance and the three perspectives, discussions on the three perspectives can provide insightful ideas on girls' low performance in science.

6.1 Gender gap in science performance & science self-efficacy

As for gender gap in science performance, Finland is the only country that has girls surpassing boys in science based on PISA 2015 results. High performance of Finnish students has attracted the attention from researchers with a very small amount of home assignments and regular lesson hours (Reinikainen, 2012). It is hard to make plausible interpretation on gender gap in science performance from this respect, because the length of science study time and science performance is not necessarily associated in a fixed way. Different countries show different features. For instance, in the world top-performing country China, students in Shanghai receive a high science performance with the most study hours of 13.8 in 2012 (OECD, 2013). But Finnish students, who top the OECD countries, only use 2.8 hours each week for homework (OECD, 2013). However, based on the result of PISA 2015, it is clear that Dutch students' study time after school is shorter than Finland, Singapore and the OECD average at a large margin. In spite of this, future research is recommended that a comparison to be made to check whether significantly different study hours can be identified between girls and boys. In this way, to what extent the length of study hours is associated with gender gap in science performance can be revealed.

Dutch girls are not only lower than boys in science performance and science self-efficacy, but also lower than girls in Finland and Singapore with regard to science performance, reading performance and science self-efficacy. Identical interpretation might apply the underperformance of Dutch 15-year-old girls. As has been compared across the three countries in Chapter 3, teachers in the Netherlands in terms of educational degree are lower than Finland and Singapore. The PISA 2015 questionnaire shows only around 30% of science teachers have University degree. In contrast, Finnish and Singaporean school teachers have a master degree with a high competition and popularity of teaching profession in these two countries. Therefore, whether teachers' low educational degree and a lower competition of teaching profession in the Netherlands hinders a high quality of science teaching which in return affect students' science performance should be further clarified.

Meanwhile, whether teachers' low educational degree is associated with a high gender-stereotypes association among teachers should be further explored. Based on previous research results, stereotypes are found to be prevalent among Dutch teachers and parents that boys have talents in STEM-related subjects and refer scientific majors as masculine (Booy, Jansen, & Joukes, 2012). In addition, Dutch teachers are found to have higher gender-stereotypes associations than some other countries (Cheryan, Siy, Vichayapai, Drury & Kim, 2011). Stereotypes in relation to gender and science among Dutch people rank right after Tunisia, (Nosek e.a. 2009). It is suggested that the stronger the gender stereotypes in relation to STEM subjects and gender, the greater the gender difference in STEM subject performance (Booy et al., 2012). Furthermore, stereotypes may negatively influence students' confidence and performance in science (Blascovich, Spencer, Quinn, & Steele, 2001; Else-Quest, Hyde & Linn, 2010). However, there is lack of self-report from teachers concerning gender-stereotypes by PISA 2015 questionnaires. Therefore, more research should be done to get empirical evidence to show whether teachers' low educational degree is associated with gender-stereotypes and finally affect students' science self-efficacy which in return affect students' science performance.

Significant gender gap in science self-efficacy exists among the three countries. Boys are significantly higher than girls with regard to science self-efficacy in the Netherlands, Finland and Singapore. This is consistent with the previous research findings that boys are more confident in STEM subjects than girls (Booy, Jansen, & Joukes, 2012). It is in line with the meta-analysis by Else-Quest, Hyde and Linn (2010) which showed that boys were "extrinsically and intrinsically" more confident and have higher self-efficacy than girls. Identifying the reasons that boys have higher science self-efficacy than girls and its relationship with science performance is important in addressing gender gap in science performance. For detailed discussion, please go to the section in Chapter 6.2. However, it was pointed out boys tend to overestimate their ability in mathematics and science (Watt, 2006). Therefore, whether boys' higher science self-efficacy than girls is the result of overestimation or a real reflection of their confidence in their ability should be further explored. For further discussion, please see Chapter 6.3

As for parental involvement, this factor has been recognized in over 50 studies to be positively associated with academic outcomes (Hill & Tyson, 2009; Kernan, 2012; Wang & Sheikh- Khalil, 2014). Comparison study in Chapter 3 shows a country difference in degree of parental involvement. However, there is little research as to what extent a parental involvement might have distinguished science performance across genders. To this point, further research could be made on whether the content and style of parental involvement is different across boys and girls. In this way, parental involvement and gender gap in science performance is related.

6.2 Relationship between science self-efficacy & science performance

As for the relationship between science self-efficacy and science performance, positive relationship exists in the three countries. The exploration on the positive relationship between science self-efficacy and science performance, plausible interpretations for it have been described in Chapter 2. The positive relationship of science self-efficacy and science performance in the research result is in line with the previous research in support of Bandura's self-efficacy theory which claims that self-efficacy is associated with science performance reliably and positively (Bandura, 1986; Motlagh, Amrai et al., 2011; Sharma & Nasa, 2014; Denissen Zarrete & Eccles, 2007). It implies that increasing students' science self-efficacy is associated with greater odds of higher science performance. The pity is that no interactive relationship can be explored in the research due to the limitation of the dataset. Therefore to what extent science performance can predict science self-efficacy deserves to be explored in future research.

Furthermore, plausible explanations for this positive relationship might lie in of the influence of a third variable (i.e. learning engagement, motivation) on students' science self-efficacy which in turn influences students' science performance. For instance, self-efficacy can influence individual learning engagement (Sharma & Nasa, 2014; Kozlowski & Salas, 2010) and duration of persistence in the tasks (Pintrich & Schunk, 2002) which might bring about higher academic performance. Thus, future research is recommended to conduct in what way that science self-efficacy can directly or indirectly contribute to a higher science performance.

Based on the result, girls have lower science self-efficacy and science performance than boys in the Netherlands and Singapore. Dutch boys have the least correlation, followed by Finnish boys and Singaporean boys. The degree of correlation can be interpreted in this way. A higher and positive correlation implies that students' science self-efficacy can positively and better predicts their science performance. Meanwhile, it can also be a case that a lower science self-efficacy correlates with a higher science performance among some girls. Or a higher science self-efficacy correlates with a lower science performance among some boys especially in the Netherlands where boys have the least correlation compared with the other two countries. However, how to interpret the gender gap in the relationship, namely, how to interpret a higher correlation between science self-efficacy and science performance among girls in the Netherlands and Finland but an equal correlation between boy and girls in Singapore could only be explored based on data from more countries.

As for the gender difference in the relationship, a higher correlation exists among girls than boys in Finland and the Netherlands. No gender gap in the correlation between science self-efficacy and science performance exists in Singapore. But no implications can be found as to the gender gap in the correlation from the three perspectives discussed in Chapter 3, i.e. teachers and teacher quality, student's science study time, and parental involvement. However, considering the plausible relationship between teachers and students' science self-efficacy and the relationship between parental involvement and students' self-efficacy that has been discussed in Chapter 3, whether teachers and parental involvement can affect the relationship between science self-efficacy and science performance is recommended to be explored in the future study.

6.3 Moderator effect of reading proficiency

The moderator analysis shows that as the reading proficiency increases, the correlation between science self-efficacy and science performance becomes higher. Based on the result, the moderator effect of reading proficiency on science performance is satisfied. Reading proficiency can weaken the relationship between science self-efficacy and science performance at a low reading level but strengthens the relationship at medium and high level. To put it in another way, the relationship between science self-efficacy and science performance becomes stronger as the reading proficiency increases. In addition, a low reading level does not only imply a low science self-efficacy but also implies a low correlation between science self-efficacy and science performance. So reading proficiency plays a role in the relationship between science self-efficacy and science performance. A higher reading proficiency is associated with great odds to get higher science performance and a higher association level between a student's science self-efficacy and the science performance.

To put it in another way, the result implies that a student's science self-efficacy becomes more reliable in predicting the science performance as his/her reading proficiency increases. From boys' perspective, there might be a case that they do not pay sufficient attention to science self-efficacy in a way that girls do. For instance, some boys may overestimate their science self-efficacy level since they are more confident than girls in science. This can make the correlation lower than it should be for the boys with high science self-efficacy but relatively low science performance. However, it is unlikely to interpret the moderator effect of reading proficiency on science performance from the perspectives of teachers and teacher quality, students' study time, and parental involvement. So future research is recommended to investigate to what extent teachers' qualification and instruction could influence students' science self-efficacy and science performance.

Gender gap in reading performance is strikingly higher than that of science performance. This is in line with the research result of PISA 2015 in which gender gap of science performance in favor of boys identified in most countries (OECD, 2016). However, it is not in line with the trend that gender gap in science performance is decreasing (OECD, 2016). In contrast, gender gap in the three countries turned out to be higher than the survey result of PISA 2012 (see Table 5). Finland is the only country in which girls surpass boys in both reading and science performance. This implies that girls can be expected to surpass boys regardless of STEM or non-STEM subjects. As for reading performance, it is prominent with its very large gender gap in favor of girls regardless of country. This is in line with the research results as well as the world trend of gender gap in reading (Booy, Jansen, & Joukes, 2012).

Research Limitations

The limitations for this project are described in this part. First, this study cannot answer why many girls achieve high outcomes in science subject but still select non-STEM careers in the end. This is a question intriguing researchers too. Exploring gender gap on science performance through moderator effect of reading proficiency can only answer 15-year-old students’ gender gap in science performance and in what way reading proficiency moderated students’ science performance. Namely, a high performance in science subject does not necessarily mean that a student will opt for science-related career. In this way, the significance of this study is restricted without taking career aspirations into consideration.

Another limitation of the project originates from the dataset. Reading self-efficacy is not measured in PISA 2015. Thus, whether and to what extent reading self-efficacy interacts with science self-efficacy in the prediction of students’ science performance cannot be studied. A comparison of reading self-efficacy and science self-efficacy cannot be achieved too. This means that investigating science self-efficacy without considering reading self-efficacy may not reveal the whole picture of the moderator effect of reading proficiency on students’ science performance. The good thing is that the core study subject of PISA 2018 is reading. The above-mentioned limitation relationship between reading self-efficacy and science self-efficacy can be expected to be studied and compared across gender.

In addition, the investigation on students’ reading and science performance is based on one test of PISA 2015 survey. There are no pre-test or post-test. Thus the relationship between science self-efficacy and science performance can be only based on one test. Without comparison of pre-test and post-test on science performance, investigation on the interactive relationship between science self-efficacy and science performance cannot be achieved. Consequently, in what way science performance predict students’ science self-efficacy is unknown in this project.

While the comparison study in Chapter 3 offers insightful ideas on the final discussions, it cannot give in-depth information on gender gap in science performance and science self-efficacy. In addition, it offers little plausible interpretation for the relationship between science self-efficacy and science performance and the moderator effect. More evidence has to be collected to explore in what way the three perspectives of teachers and teacher quality, study time, and parental involvement can distinguish students’ science performance and science self-efficacy across genders.

As for the comparison countries, to what extent schools of the Netherlands can get practical insights from the good examples of Finland and Singapore is restricted and complicated. Each country is unique in culture and history. This makes learning from the other country is not easy to resolve the educational problems in one’s own country. For instance, teachers’ educational level in the Netherlands is lower than Finland and Singapore. However, it is not easy to improve the educational level among Dutch teachers since the competitiveness of teaching profession in the Netherlands is much less than Finland and Singapore. In Finland, one out of nine applicants can finally become a teacher where most secondary school teachers have master degrees. In addition, the number of the comparison countries is only two which is too few to represent the whole situation in the world.

Part of research limitation comes from the limitation of the major tool of SPSS programming software and IEA IDB Analyzer. Each of them has its own advantages and disadvantages. For instance, though IDB Analyzer is very efficiency in dealing with means and percentages of plausible values, but interaction effect cannot be introduced into it. Though SPSS is user-friendly in regression analysis but due to its technical limitation dealing with the plausible values, science and reading performance denoted by ten plausible values cannot be analyzed at one time but separately. This makes lengthy data results are produced. In addition, interpretation is mainly based on PV1, though very similar results are produced from PV2 to PV10. Therefore, a more efficient way of processing plausible values is expected to be made.

REFERENCES

Ability. 2017. In *Merriam-Webster.com*. Retrieved June 14, 2017, from https://www.merriam-webster.com/dictionary/ability

Adedokun, O. A., Bessenbacher, A. B., Parker, L. C., Kirkham, L. L., & Burgess, W. D. (2013). Research skills and STEM undergraduate research students' aspirations for research careers: Mediating effects of research self- efficacy. *Journal of Research in Science teaching*, *50*(8), 940-951.

Andere, E. (2015). Are teachers crucial for academic achievement? Finland educational success in a comparative perspective. Education Policy Analysis Archives, 23(39). http://dx.doi.org/10.14507/epaa.v23.1752

Anderson, S. L., & Betz, N. E. (2001). Sources of social self-efficacy expectations: their measurement and relation to career development. Journal of Vocational Behaviour, 58, 98–117.

Background and basics. (2016, December 5). Retrieved from https://www.oecd.org/pisa/aboutpisa/pisafaq.htm

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational psychologist*, *28*(2), 117-148.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: W. H. Freeman and Company

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self- efficacy beliefs as shapers of children's aspirations and career trajectories. *Child development*, *72*(1), 87-206.

Bautista, A., Wong, J., & Gopinathan, S. (2015). Teacher professional development in Singapore: Depicting the landscape.

Beattie, S., Woodman, T., Fakehy, M., & Dempsey, C. (2016). The role of performance feedback on the self-efficacy–performance relationship. *Sport, Exercise, and Performance Psychology*, *5*(1), 1.

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). Women in STEM: A gender gap to innovation.

Betz, N. E., & Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. Journal of Vocational Behavior, 23, 329–345. doi:10.1016/00018791(83)90046-5.

Bhana, D. (2005). "I'm the best in maths. Boys rule, girls drool." Masculinities, mathematics, and primary schooling. Perspectives in Education, 23, 1–10.

Bieri Buschor, C., Berweger, S., Keck Frei, A., & Kappler, C. (2014). Majoring in STEM—what accounts for women's career decision making? A mixed method study. *The Journal of Educational Research*, *107*(3), 167-176.

Bishop, A. (2015). *Career Aspirations of High School Males and Femailes in a Science, Technology, Engineering, and Mathematics Program* (Doctoral dissertation).

Blascovich, J., Spencer, S. J., Quinn, D. M., & Steele, C. M. (2001). Stereotype threat and the cardiovascular reactivity of AfricanAmericans. Psychological Science, 12, 225–229

Booij, C., Jansen, E. J. M., & van Schaik, E. J. (2012). Long term, interrelated interventions to increase women's participation in STEM in the Netherlands. *International Journal of Gender, Science and Technology*, Vol. 5, No. 3.

Booy, C., Jansen, N., & Joukes, G. (2012). Trend analysis gender in higher STEM education, VHTO, file:///D:/EST%20Materials/Final/LITERATURE/DUTCH%20-%20PISA-Trend%20analysis%20gender%20in%20higher%20STEM%20education.pdf
Trend analysis: gender in higher STEM education.

Britner, S. L., & Pajares, F. (2006). Sources of science self- efficacy beliefs of middle school students. *Journal of research in science teaching*, *43*(5), 485-499.

Brown, D. (2002). Career choice and development (4th ed.). San Francisco: Jossey-Bass.

Brusokas, A., & Malinauskas, R. (2014). Career self-efficacy among Lithuanian adolescents in sports schools. *Procedia-Social and Behavioral Sciences*, *116*, 212-216.

Busch, T. (1995). Gender differences in self-efficacy and attitudes toward computers. *Journal of educational computing research*, *12*(2), 147-158.

Bybee, R. W. (2010b). Advancing STEM education: A 2020 Vision. Technology and Engineering Teacher, 70(1), 30-35.

Catsambis, S. (2001). Expanding knowledge of parental involvement in children's secondary education: Connections with high school seniors' academic success. *Social Psychology of Education*, *5*(2), 149-177.

Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM?. *Social Psychological and Personality Science*, *2*(6), 656-664.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2016). Why are some STEM fields more gender balanced than others?

Christou, C., Eliophotou-Menon, M., & Philippou, G. (2004). Teachers' concerns regarding the adoption of a new mathematics curriculum: An application of CBAM. *Educational Studies in Mathematics*, *57*(2), 157-176.

Clark B. J. (2005). Women and science careers: leaky pipeline or gender filter?. *Gender and education*, *17*(4), 369-386.

Clarke, C. (2001). The role of parents in Singapore primary schools. *Teaching and Learning*, *22*(2), 83-92.

Cunningham, B. C., Hoyer, K. M., & Sparks, D. (2015). Gender Differences in Science, Technology, Engineering, and Mathematics (STEM) Interest, Credits Earned, and NAEP Performance in the 12th Grade. Stats in Brief. NCES 2015-075. *National Center for Education Statistics*.

De Peslouüan, G. (1974). *Qui sont les femmes ingénieurs en France?* (Vol. 25). Publication Univ Rouen Havre.

De Vries, M. J., van Keulen, H., Peters, S., & van der Molen, J. W. (Eds.). (2012). *Professional development for primary teachers in science and technology* (Vol. 9). Springer Science & Business Media.

Driessen, G., Smit, F., & Sleegers, P. (2005). Parental involvement and educational achievement. *British educational research journal*, *31*(4), 509-532.

Doorman, M., Drijvers, P., Dekker, T., van den Heuvel-Panhuizen, M., de Lange, J., & Wijers, M. (2007). Problem solving as a challenge for mathematics education in The Netherlands. *ZDM*, *39*(5-6), 405-418.

Dowey, A. L. (2013). *Attitudes, Interests, and Perceived Self-efficacy toward Science of Middle School Minority Female Students: Considerations for their Low Achievement and Participation in STEM Disciplines* (Doctoral dissertation, California State University, San Marcos).

Eccles, J. S. (2007). Where Are All the Women? Gender Differences in Participation in Physical Science and Engineering, in: Stephen J. Ceci & Wendy M. Williams, Why aren't more women in science? American Psychological Association.

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

Eurostat (2009). Education statistics, UOE data collection. European Commission (online http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/de/educ_esms.htm accessed in 2009)

EURYDICE European Unit, Brussels (Belgium). (1997). *The Role of Parents in the Education Systems of the European Union*. [Washington D.C.]: Distributed by ERIC Clearinghouse, http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED426950

Fan, W., & Williams, C. M. (2010). The effects of parental involvement on students' academic self- efficacy, engagement and intrinsic motivation. *Educational Psychology*, *30*(1), 53-74.

Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational psychology review*, *13*(1), 1-22.

Fennema, E., Peterson, P. L., Carpenter, T. P., & Lubinski, C. A. (1990). Teachers' attributions and beliefs about girls, boys, and mathematics. Educational Studies in Mathematics, 21, 55–69.

Forsthuber, B., Motiejunaite, A., & de Almeida Coutinho, A. S. (2011). *Science Education in Europe: National Policies, Practices and Research*. Education, Audiovisual and Culture Executive Agency, European Commission. Available from EU Bookshop.

Freeman, J. & Garces-Bascal, R.M. (2015), ‗Gender differences in gifted children', in M. Neihart, S.I. Pfeiffer & T.L. Cross (Eds.) The Social and Emotional Development of Gifted Children: What do we know?, Waco, Texas: Prufrock Press Inc

Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of personality and social psychology*, *74*(2), 435.

Furnham, A., Reeves, E., & Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *The Journal of genetic psychology*, *163*(1), 24-39.

George A. (2015, June 13).How well are women represented in UK science? *The Guardian*. Retrieved from https://www.theguardian.com/news/datablog/2015/jun/13/how-well-are-women -represented-in-uk-science

George, R., & Kaplan, D. (1998). A structural model of parent and teacher influences on science attitudes of eighth graders: Evidence from NELS: 88. *Science Education*, *82*(1), 93-109.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological science in the public interest*, *8*(1), 1-51.

Hango, D. W. (2013). *Gender differences in science, technology, engineering, mathematics and computer science (STEM) programs at university*. Ottawa: Statistics Canada.

Hedges, L. V., & Nowell, A. (1995, July 7). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science, 269, 41–45.

Herrera, F. A., Hurtado, S., & Chang, M. (2012). Maintaining career aspirations in science, echnology, engineering, and mathematics (STEM) among college students. In a*nnual onference of the Association for the Study of Higher Education, Charlotte, NC*.

Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: a meta-analytic assessment of the strategies that promote achievement. *Developmental psychology*, *45*(3), 740.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A., & Williams, C. (2008, July 25). Gender similarities characterize math performance. Science, 321, 494–495.

Hyde, J. S. (2005). The gender similarities hypothesis. *American psychologist*, *60* (6), 581.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*(5888), 494-495.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*(5888), 494-495.

Jansen, N., & Joukes, G. (2013). Long term, interrelated interventions to increase women's participation in STEM in the Netherlands. *International Journal of Gender, Science and Technology*, *5*(3), 305-316.

Kernan, M. (2012). Parental involvement in early learning. A review of research, policy and good practice. *International Child Development Initiative (ICDI) Leiden on behalf of the Barnard van Leer Foundation, The Hague*.

Kim, E. M., Coutts, M. J., Holmes, S. R., Sheridan, S. M., Ransom, K. A., Sjuts, T. M., & Rispoli, K. M. (2012). Parent Involvement and Family-School Partnerships: Examining the Content, Processes, and Outcomes of Structural versus Relationship-Based Approaches. CYFS Working Paper No. 2012-6. *Nebraska Center for Research on Children, Youth, Families and Schools*.

Kinzie, M. B., & Delcourt, M. A. (1991). Computer technologies in teacher education: The measurement of attitudes and self-efficacy.

Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. In *The decade ahead: Theoretical perspectives on motivation and achievement* (pp. 1-33). Emerald Group Publishing Limited.

Kupari, P. (2008). Mathematics Education in Finnish Comprehensive School: Characteristics Contributing to Student Success. *Lecture presented at ICME-11, Monterrey, Mexico*.

Kuehn, Z., & Landeras, P. (2012). Study time and scholarly achievement in PISA.

Langen, A. V., & Dekkers, H. (2005). Cross- national differences in participating in tertiary science, technology, engineering and mathematics education. *Comparative Education*,

LeBeau, B., Harwell, M., Monson, D., Dupuis, D., Medhanie, A., & Post, T. R. (2012).

Student and high-school characteristics related to completing a science, technology, engineering or mathematics (STEM) major in college. *Research in Science & Technological Education*, *30*(1), 17-28.

Lee, V. E. (2001). *Restructuring High Schools for Equity and Excellence: What Works. Sociology of Education Series*. Teachers College Press, 1234 Amsterdam Avenue, New York, NY 10027.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis.

Linnakylä, P. & Välijärvi, J. (2005). Secrets to literacy success. The Finnish story. Education Canada, 45 (3), 34−39.

Kıran, D., & Sungur, S. (2012). Middle school students' science self-efficacy and its sources: Examination of gender difference. *Journal of Science Education and Technology*, *21*(5), 619-630.

Marginson, Simon, Tytler, Russell, Freeman, Brigid and Roberts, Kelly (2013), STEM: country comparisons: international comparisons of science, technology, engineering and mathematics (STEM) education. Final report. Australian Council of Learned Academies, Melbourne, Vic.

Mastrull, S. (2002). The Mathematics Education of Students in Japan: A Comparison with United States Mathematics Programs.

Modi, K., Schoenberg, J., & Salmond, K. (2012). Generation STEM: What girls say about science, technology, engineering, and math. *A Report from the Girl Scout Research Institute. New York, NY: Girl Scouts of the USA.*

Monseur, C. (2009). PISA Data Analysis Manual: SPSS second edition.

Motlagh, S. E., Amrai, K., Yazdani, M. J., altaib Abderahim, H., & Souri, H. (2011). The relationship between self-efficacy and academic achievement in high school students. *Procedia-Social and Behavioral Sciences*, *15*, 765-768.

Mullis, I.V.S., Martin, M.O., Beaton, A.E., Gonzalez, E.J., Kelly, D.L., and Smith, T.A. (1997). Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS). Chestnut Hill, MA: Boston College.

Mullis, I. V., & Martin, M. O. (2014). *TIMMS Advanced 2015 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *The Journal of Economic Perspectives*, *24*(2), 129-144.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., ... & Kesebir, S. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*(26), 10593-10597.

Nugent, G. et al. (2015), "A model of factors contributing to STEM learning and career orientation", International Journal of Science Education, Vol. 37/7, pp. 1067-1088, http://dx.doi.org/10.1080/09500693.2015.1017863.

OECD (2003). PISA dataset 2003. Paris, France: OECD

OECD. Publishing. (2009). *PISA Data Analysis Manual: SAS*. OECD.

OECD (2013), *PISA 2012 Results: Ready to Learn (Volume III): Students' Engagement, Drive and Self-Beliefs*, OECD Publishing, Paris.

OECD (2016), PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264255425-en

OECD (2017), "PISA: Programme for International Student Assessment", *OECD Education Statistics* (database). DOI: http://dx.doi.org/10.1787/data-00365-en *41*(3), 329-350.

O'Brien, V., Martinez-Pons, M., & Kopala, M. (1999). Mathematics self-efficacy, ethnic identity, gender, and vocational interests related to mathematics and science. *The Journal of Educational Research*, 92, 231–235. Doi:10.1080/00220679909597600.

Organisation de cooperation et de développement économiques. (2015). *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. OECD.
Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory into practice*, *41*(2), 116-125.
Panhuizen, M. (2000). Mathematics education in the Netherlands: A guided tour. *Freudenthal Institute CD-rom for ICME9*, 1-32.
Pajares, F. (2005). Gender differences in mathematics self-efficacy beliefs. In A. M. Gallagher & J. C. Kaufman (Eds.), Gender differences in mathematics: An integrative psychological approach (pp. 294–315). New York: Cambridge University Press
Parveva, T., Noorani, S., Ranguelov, S., Motiejunaite, A., & Kerpanova, V. (2011). *Mathematics Education in Europe: Common Challenges and National Policies*. Education, Audiovisual and Culture Executive Agency, European Commission. Available from EU Bookshop.
Peña-López, I. (2016). PISA 2015 Results (Volume I). Excellence and Equity in Education.
Piercey, R. R. (2013). Reading Self-Efficacy in Early Adolescence: Which Measure Works Best?.
Preckel, F., Goetz, T., Pekrun, R., & Kleine, M. (2008). Gender differences in gifted and average-ability students: Comparing girls' and boys' achievement, self-concept, interest, and motivation in mathematics. *Gifted Child Quarterly*, *52*(2), 146-159.
Ramirez, F. O., & Kwak, N. (2015). Women's Enrollments in STEM in Higher Education: Cross-National Trends, 1970–2010. In *Advancing Women in Science* (pp. 9-49). Springer International Publishing.
Reinikainen, P. (2012). Amazing PISA results in Finnish comprehensive schools. In *Miracle of education* (pp. 3-18). Sense Publishers.
Rittmayer, A. D., & Beier, M. E. (2008). Overview: Self-efficacy in STEM. *SWE-AWE CASEE Overviews*.
Rossi, A. S. (1965). Women in science: Why so Few? *Science*, *148*(3674), 1196-1202.
Rudasill, K. M., & Callahan, C. M. (2010). Academic self-perceptions of ability and course planning among academically advanced students. *Journal of Advanced Academics*, 21, 300–329.
Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles*, *66*(3-4), 175-183.
Sharma, H. L., & Nasa, G. (2014). Academic self-efficacy: a reliable predictor of educational performances. *British Journal of Education*, *2*(3), 57-64.
Schunk, D. H. (1981). Modeling and attributional effects on children's achievement: A self-efficacy analysis. *Journal of educational psychology*, *73*(1), 93.
Schuster, C., & Martiny, S. E. (2017). Not feeling good in STEM: effects of stereotype activation and anticipated affect on women's career aspirations. *Sex Roles*, *76*(1-2), 40-55
Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self- concept, performance expectations, and motivation. Sex Roles: A Journal of Research, 50, 241-252.
Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles*, *67*(11-12), 617-629.
Stigler, J., & Hielbert, J. (1997). Understanding and imporving mathematics instruction: An overview of the TIMSS video study. Phi Delta Kappan, 79*(1)*, 14-21.Van den Heuvel-
Stoet, G., & 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), e57988.
Stajkovic, A. D., & Luthans, F. (1979). Social Cognitive Theory and Self. Efficacy: Implications for Motivation Theory and Practice.
Spearman, J., & Watt, H. M. (2013). Women's Aspirations Towards "STEM" Careers. In *Conceptualizing women's working lives* (pp. 175-191). Sense Publishers.
Strategic Framework for European Cooperation in Education and Training ('ET 2020'), Council Conclusions May 2008, OJL 119, 28.5.2009.

Tseng, K. H., Chang, C. C., Lou, S. J., & Chen, W. P. (2013). Attitudes towards science, technology, engineering and mathematics (STEM) in a project-based learning (PjBL) environment. *International Journal of Technology and Design Education*, *23*(1), 87-102.
White, D. W. (2014). What Is STEM education and why is it important. *Florida Association of Teacher Educators Journal*, *1*(14), 1-9.
Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Comtemporary Educational Psychology*, 25, 68–81. Doi:10.1006/ceps.1999.1015.
van Aalderen-Smeets, S. I., & van der Molen, J. H. W. (2016). Modeling the relation between students' implicit beliefs about their abilities and their educational STEM choices. *International journal of technology and design education*, 1-27.
van den Berg, M., & van Reekum, R. (2011). Parent involvement as professionalization: professionals' struggle for power in Dutch urban deprived areas. *Journal of Education Policy*, *26*(3), 415-430.
van Langen, A., Bosker, R., & Dekkers, H. (2006). Exploring cross-national differences in gender gaps in education. *Educational Research and Evaluation*, *12*(02), 155-177.
Wang, M. T., Eccles, J. S., & Kenny, S. (2013). Not lack of ability but more choice individual and gender differences in choice of careers in science, technology, engineering, and mathematics. *Psychological Science*, *24*(5), 770-775.
Wu, J. Y. (2017). The indirect relationship of media multitasking self-efficacy on learning performance within the personal learning environment: Implications from the mechanism of perceived attention problems and self-regulation strategies. *Computers & Education*, *106*, 56-72.
Yau, H. K., & Leung, Y. F. (2016). Gender Difference of Self-Efficacy and Attitudes Towards The Use of Technology in Learning in Hong Kong Higher Education. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (Vol. 2). http://www.iaeng.org/publication/IMECS2016/IMECS2016_pp819-821.pdf
Y.S. Park, "An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning," Educational Technology & Society, vol. 12, no. 3, 2009, pp. 150-162.
Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25, 82–91.

Appendix A

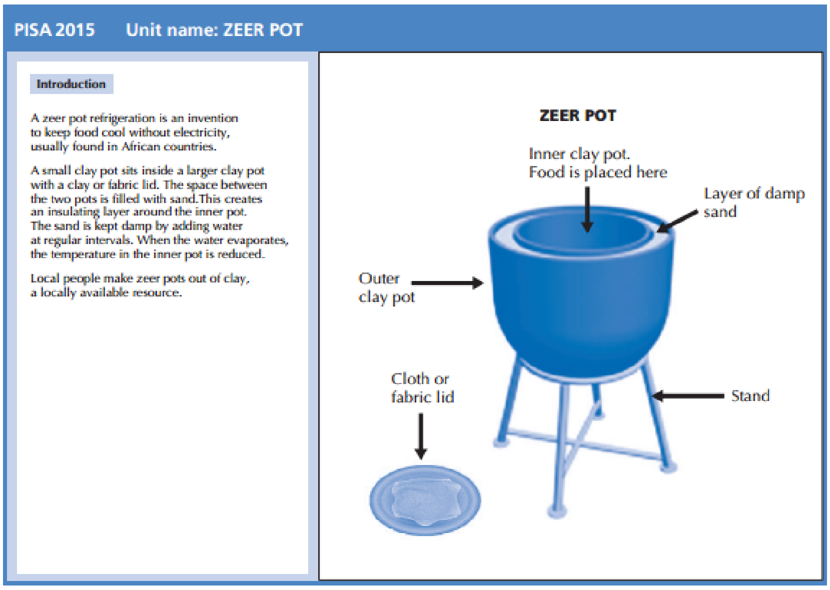
Description of the seven levels of proficiency in science in PISA 2015

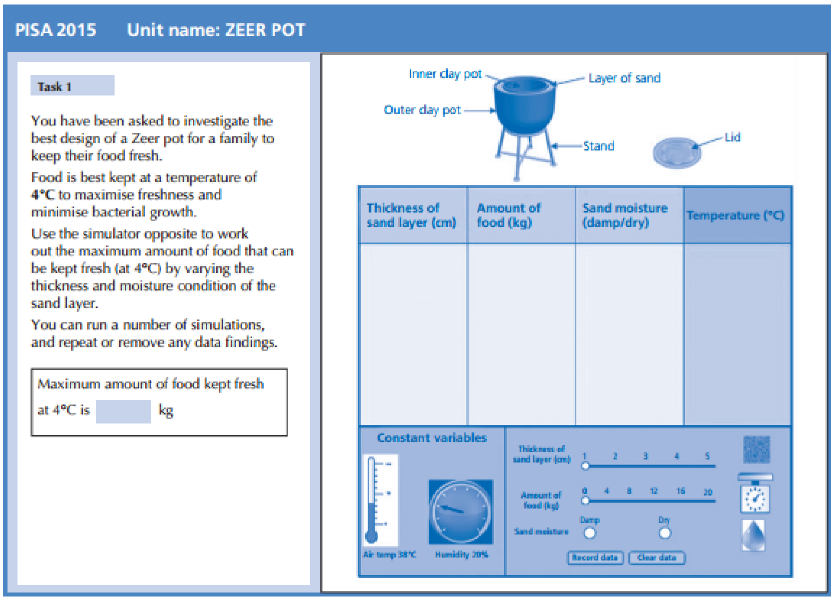
Level	Lower score limit	Characteristics of tasks
6	708	At Level 6, students can draw on a range of interrelated scientific ideas and concepts from the physical, life and earth and space sciences and use content, procedural and epistemic knowledge in order to offer explanatory hypotheses of novel scientific phenomena, events and processes or to make predictions. In interpreting data and evidence, they are able to discriminate between relevant and irrelevant information and can draw on knowledge external to the normal school curriculum. They can distinguish between arguments that are based on scientific evidence and theory and those based on other considerations. Level 6 students can evaluate competing designs of complex experiments, field studies or simulations and justify their choices.
5	633	At Level 5, students can use abstract scientific ideas or concepts to explain unfamiliar and more complex phenomena, events and processes involving multiple causal links. They are able to apply more sophisticated epistemic knowledge to evaluate alternative experimental designs and justify their choices and use theoretical knowledge to interpret information or make predictions. Level 5 students can evaluate ways of exploring a given question scientifically and identify limitations in interpretations of data sets including sources and the effects of uncertainty in scientific data.
4	559	At Level 4, students can use more complex or more abstract content knowledge, which is either provided or recalled, to construct explanations of more complex or less familiar events and processes. They can conduct experiments involving two or more independent variables in a constrained context. They are able to justify an experimental design, drawing on elements of procedural and epistemic knowledge. Level 4 students can interpret data drawn from a moderately complex data set or less familiar context, draw appropriate conclusions that go beyond the data and provide justifications for their choices.
3	484	At Level 3, students can draw upon moderately complex content knowledge to identify or construct explanations of familiar phenomena. In less familiar or more complex situations, they can construct explanations with relevant cueing or support. They can draw on elements of procedural or epistemic knowledge to carry out a simple experiment in a constrained context. Level 3 students are able to distinguish between scientific and non-scientific issues and identify the evidence supporting a scientific claim.
2	410	At Level 2, students are able to draw on everyday content knowledge and basic procedural knowledge to identify an appropriate scientific explanation, interpret data, and identify the question being addressed in a simple experimental design. They can use basic or everyday scientific knowledge to identify a valid conclusion from a simple data set. Level 2 students demonstrate basic epistemic knowledge by being able to identify questions that can be investigated scientifically.
1a	335	At Level 1a, students are able to use basic or everyday content and procedural knowledge to recognise or identify explanations of simple scientific phenomenon. With support, they can undertake structured scientific enquiries with no more than two variables. They are able to identify simple causal or correlational relationships and interpret graphical and visual data that require a low level of cognitive demand. Level 1a students can select the best scientific explanation for given data in familiar personal, local and global contexts.
1b	261	At Level 1b, students can use basic or everyday scientific knowledge to recognise aspects of familiar or simple phenomenon. They are able to identify simple patterns in data, recognise basic scientific terms and follow explicit instructions to carry out a scientific procedure.

Retrieved from http://www.oecd.org/pisa/test/summary-description-seven-levels-of-proficiency-science-pisa-2015.html

Appendix B

A sample of science test item-ZEER POT





(OECD, 2016, p.35-36)

Appendix C

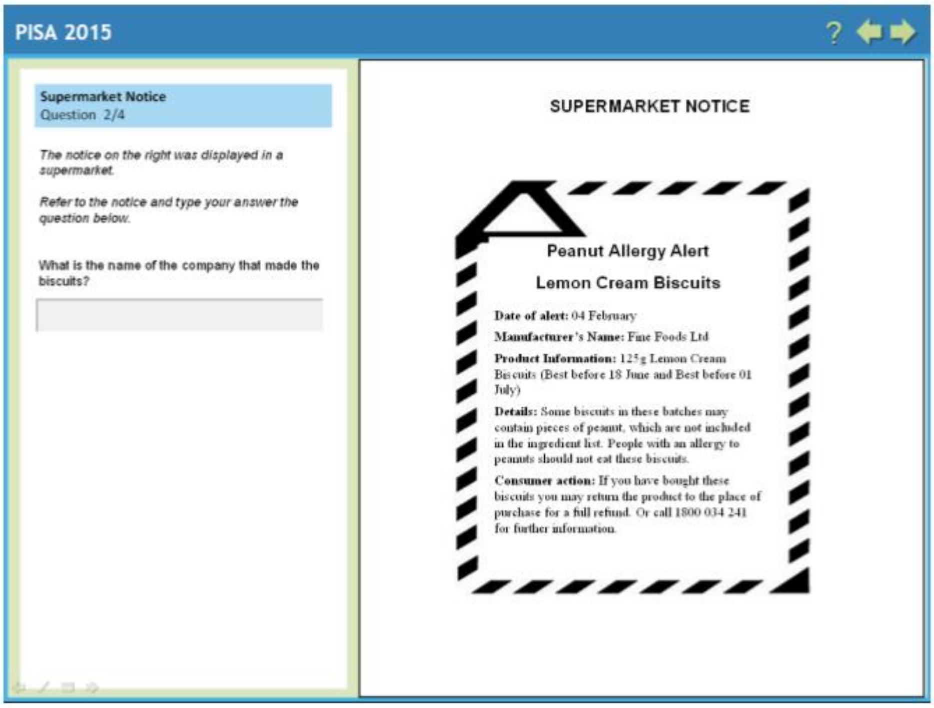
Description of the seven levels of reading proficiency in PISA 2015

Level	Lower score limit	Characteristics of tasks
6	698	Tasks at this level typically require the reader to make multiple inferences, comparisons and contrasts that are both detailed and precise. They require demonstration of a full and detailed understanding of one or more texts and may involve integrating information from more than one text. Tasks may require the reader to deal with unfamiliar ideas, in the presence of prominent competing information, and to generate abstract categories for interpretations. *Reflect and evaluate* tasks may require the reader to hypothesise about or critically evaluate a complex text on an unfamiliar topic, taking into account multiple criteria or perspectives, and applying sophisticated understandings from beyond the text. A salient condition for access and retrieve tasks at this level is precision of analysis and fine attention to detail that is inconspicuous in the texts.
5	626	Tasks at this level that involve retrieving information require the reader to locate and organise several pieces of deeply embedded information, inferring which information in the text is relevant. Reflective tasks require critical evaluation or hypothesis, drawing on specialised knowledge. Both interpretative and reflective tasks require a full and detailed understanding of a text whose content or form is unfamiliar. For all aspects of reading, tasks at this level typically involve dealing with concepts that are contrary to expectations.
4	553	Tasks at this level that involve retrieving information require the reader to locate and organise several pieces of embedded information. Some tasks at this level require interpreting the meaning of nuances of language in a section of text by taking into account the text as a whole. Other interpretative tasks require understanding and applying categories in an unfamiliar context. Reflective tasks at this level require readers to use formal or public knowledge to hypothesise about or critically evaluate a text. Readers must demonstrate an accurate understanding of long or complex texts whose content or form may be unfamiliar.
3	480	Tasks at this level require the reader to locate, and in some cases recognise the relationship between, several pieces of information that must meet multiple conditions. Interpretative tasks at this level require the reader to integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase. They need to take into account many features in comparing, contrasting or categorising. Often the required information is not prominent or there is much competing information; or there are other text obstacles, such as ideas that are contrary to expectation or negatively worded. Reflective tasks at this level may require connections, comparisons and explanations, or they may require the reader to evaluate a feature of the text. Some reflective tasks require readers to demonstrate a fine understanding of the text in relation to familiar, everyday knowledge. Other tasks do not require detailed text comprehension but require the reader to draw on less common knowledge.
2	407	Some tasks at this level require the reader to locate one or more pieces of information, which may need to be inferred and may need to meet several conditions. Others require recognising the main idea in a text, understanding relationships, or construing meaning within a limited part of the text when the information is not prominent and the reader must make low level inferences. Tasks at this level may involve comparisons or contrasts based on a single feature in the text. Typical reflective tasks at this level require readers to make a comparison or several connections between the text and outside knowledge, by drawing on personal experience and attitudes.
1a	335	Tasks at this level require the reader to locate one or more independent pieces of explicitly stated information; to recognise the main theme or author's purpose in a text about a familiar topic, or to make a simple connection between information in the text and common, everyday knowledge. Typically the required information in the text is prominent and there is little, if any, competing information. The reader is explicitly directed to consider relevant factors in the task and in the text.
1b	262	Tasks at this level require the reader to locate a single piece of explicitly stated information in a prominent position in a short, syntactically simple text with a familiar context and text type, such as a narrative or a simple list. The text typically provides support to the reader, such as repetition of information, pictures or familiar symbols. There is minimal competing information. In tasks requiring interpretation the reader may need to make simple connections between adjacent pieces of information.

(OECD, 2016, p. 59)

Appendix D

A sample of computer-based reading item



(Thomson, Hillman & De Bortoli, 2013, p. 36)

Appendix E
Test items for science self-efficacy in PISA 2015

- **A** Recognise the science question that underlies a newspaper report on a health issue
- **B** Explain why earthquakes occur more frequently in some areas than in others
- **C** Describe the role of antibiotics in the treatment of disease
- **D** Identify the science question associated with the disposal of garbage
- **E** Predict how changes to an environment will affect the survival of certain species
- **F** Interpret the scientific information provided on the labelling of food items
- **G** Discuss how new evidence can lead you to change your understanding about the possibility of life on Mars
- **H** Identify the better of two explanations for the formation of acid rain

(OECD, 2015, p. 137)

Appendix F

Gender difference in science and reading performance (PV2-10)

Country	Gender		PV2	PV3	PV4	PV5	PV6	PV7	PV8	PV9	PV10
FL Science	Female	Mean	545.55	544.38	545.85	544.27	544.75	543.77	544.21	545.11	544.65
		N	2707	2707	2707	2707	2707	2707	2707	2707	2707
		SD	87.50	87.09	87.92	87.51	88.10	89.25	88.60	88.77	88.31
	Male	Mean	529.76	529.30	528.83	529.61	529.64	528.84	528.93	528.58	528.72
		N	2763	2763	2763	2763	2763	2763	2763	2763	2763
		SD	97.59	98.36	97.33	96.91	98.21	98.25	97.69	98.27	97.85
FL Reading	Female	Mean	553.65	555.45	555.63	554.09	553.62	555.96	554.60	556.94	554.13
		N	2707	2707	2707	2707	2707	2707	2707	2707	2707
		SD	81.50	82.43	82.68	81.97	81.75	82.27	81.55	80.85	83.79
	Male	Mean	512.69	510.90	510.21	511.99	512.65	510.8	511.25	512.04	511.94
		N	2763	2763	2763	2763	2763	2763	2763	2763	2763
		SD	94.00	94.71	93.74	93.55	93.15	94.47	93.36	93.48	93.94
NL Science	Female	Mean	515.74	515.20	515.22	515.59	515.95	515.04	514.75	515.48	515.88
		N	2570	2570	2570	2570	2570	2570	2570	2570	2570
		SD	94.70	93.69	92.55	92.25	92.724	92.37	93.23	92.90	93.29
	Male	Mean	523.07	522.63	521.62	521.94	520.33	522.10	522.96	522.60	522.11
		N	2486	2486	2486	2486	2486	2486	2486	2486	2486
		SD	100.94	99.33	98.99	99.91	100.27	100.23	99.26	100.52	100.58
NL Reading	Female	Mean	523.31	522.87	523.94	525.89	527.46	525.25	523.56	525.69	525.57
		N	2570	2570	2570	2570	2570	2570	2570	2570	2570
		SD	90.991	90.36	91.22	88.72	90.06	90.28	91.93	91.84	92.72
	Male	Mean	503.57	503.76	504.17	506.01	502.58	501.95	502.36	504.43	501.59
		N	2486	2486	2486	2486	2486	2486	2486	2486	2486
		SD	97.50	94.73	96.41	94.74	96.83	96.78	97.77	97.78	96.63
SP Science	Female	Mean	546.23	545.81	547.35	546.16	545.91	547.01	546.51	545.99	546.85
		N	2921	2921	2921	2921	2921	2921	2921	2921	2921
		SD	98.84	98.25	99.19	98.23	98.51	98.17	98.13	98.67	97.65
	Male	Mean	551.05	550.05	551.19	551.02	550.48	549.75	551.84	551.35	551.56
		N	3078	3078	3078	3078	3078	3078	3078	3078	3078
		SD	107.92	107.99	105.46	107.63	107.13	106.62	106.75	107.86	107.23
SP Reading	Female	Mean	540.07	540.05	540.97	540.67309	538.19	539.98	541.45	541.76	539.54
		N	2921	2921	2921	2921	2921	2921	2921	2921	2921
		SD	96.29	94.49	95.92	95.075	94.60	95.72	94.06	94.44	94.86
	Male	Mean	517.63	518.49	517.13	519.37	516.78	517.29	519.17	518.84	519.63
		N	3078	3078	3078	3078	3078	3078	3078	3078	3078
		SD	101.38	99.51	98.79	100.15	100.16	99.39	99.43	98.21	98.48

Appendix G

Significance of gender difference in science performance (PV2-10)

	Country	Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	T	Sig.
PV2		1	(Constant)	561.081	3.984		140.846	.000
	FL		Gender	-15.631	2.506	-.084	-6.237	.000
	NL	1	(Constant)	506.298	4.352		116.330	.000
			Gender	7.297	2.767	.037	2.637	.008
	SP	1	(Constant)	545.253	4.243		128.507	.000
			Gender	8.027	2.657	.039	3.020	.003
PV3	FL	1	(Constant)	559.231	3.996		139.951	.000
			Gender	-14.949	2.514	-.080	-5.947	.000
	NL	1	(Constant)	505.262	4.295		117.644	.000
			Gender	7.701	2.730	.040	2.821	.005
	SP	1	(Constant)	545.337	4.244		128.487	.000
			Gender	7.525	2.658	.037	2.831	.005
PV4		1	(Constant)	562.484	3.987		141.064	.000
	FL		Gender	-16.857	2.508	-.091	-6.720	.000
	NL	1	(Constant)	506.853	4.267		118.772	.000
			Gender	6.229	2.713	.032	2.296	.022
	SP	1	(Constant)	547.835	4.207		130.221	.000
			Gender	6.693	2.635	.033	2.540	.011
PV5	FL	1	(Constant)	559.089	3.968		140.891	.000
			Gender	-14.803	2.496	-.080	-5.930	.000
	NL	1	(Constant)	506.809	4.271		118.661	.000
			Gender	6.608	2.715	.034	2.434	.015
	SP	1	(Constant)	545.349	4.230		128.939	.000
			Gender	8.013	2.649	.039	3.025	.002
PV6		1	(Constant)	559.820	4.005		139.768	.000
	FL		Gender	-15.091	2.520	-.081	-5.989	.000
	NL	1	(Constant)	509.382	4.287		118.827	.000
			Gender	4.308	2.725	.022	1.581	.114
	SP	1	(Constant)	545.097	4.229		128.882	.000
			Gender	7.708	2.649	.038	2.910	.004
PV7	FL	1	(Constant)	558.436	4.043		138.137	.000
			Gender	-14.749	2.543	-.078	-5.800	.000
	NL	1	(Constant)	505.828	4.286		118.009	.000
			Gender	7.090	2.725	.037	2.602	.009
	SP	1	(Constant)	547.610	4.204		130.247	.000
			Gender	6.154	2.633	.030	2.337	.019
PV8		1	(Constant)	559.117	4.010		139.444	.000
	FL		Gender	-15.061	2.522	-.080	-5.971	.000

	NL	1	(Constant)	504.422	4.281		117.817	.000
			Gender	8.319	2.722	.043	3.056	.002
	SP	1	(Constant)	545.818	4.201		129.925	.000
			Gender	7.955	2.631	.039	3.023	.003
PV9	FL	1	(Constant)	561.616	4.025		139.530	.000
			Gender	-16.423	2.532	-.087	-6.486	.000
	NL	1	(Constant)	506.788	4.296		117.957	.000
			Gender	6.808	2.731	.035	2.493	.013
	SP	1	(Constant)	544.353	4.249		128.112	.000
			Gender	8.594	2.661	.042	3.229	.001
PV10		1	(Constant)	560.194	4.011		139.679	.000
	FL		Gender	-15.716	2.523	-.084	-6.229	.000
	NL	1	(Constant)	507.518	4.309		117.779	.000
			Gender	6.247	2.739	.032	2.281	.023
	SP	1	(Constant)	545.780	4.215		129.482	.000
			Gender	7.931	2.640	.039	3.004	.003

Note. a. Dependent Variable: Plausible Value 2-10 in Science
b. Weighted Least Squares Regression - Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGHT

Appendix H

Significance of gender difference in reading performance (PV2-10)

	Country	Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	T	Sig.
PV2		1	(Constant)	594.152	3.788		156.868	.000
	FL		Gender	-40.811	2.383	-.226	-17.128	.000
	NL	1	(Constant)	540.618	4.182		129.276	.000
			Gender	-19.560	2.659	-.103	-7.357	.000
	SP	1	(Constant)	567.101	4.068		139.417	.000
			Gender	-20.150	2.548	-.102	-7.909	.000
PV3	FL	1	(Constant)	599.528	3.813		157.213	.000
			Gender	-44.417	2.399	-.243	-18.515	.000
	NL	1	(Constant)	539.527	4.114		131.143	.000
			Gender	-18.896	2.615	-.101	-7.225	.000
	SP	1	(Constant)	566.151	4.008		141.244	.000
			Gender	-19.221	2.510	-.098	-7.656	.000
PV4	FL	1	(Constant)	600.962	3.805		157.955	.000
			Gender	-45.458	2.393	-.249	-18.993	.000
	NL	1	(Constant)	541.406	4.173		129.749	.000
			Gender	-19.615	2.653	-.103	-7.394	.000
	SP	1	(Constant)	569.449	4.008		142.075	.000
			Gender	-21.578	2.510	-.110	-8.596	.000
PV5	FL	1	(Constant)	596.111	3.781		157.661	.000
			Gender	-42.172	2.379	-.233	-17.731	.000
	NL	1	(Constant)	543.892	4.079		133.350	.000
			Gender	-19.724	2.593	-.106	-7.607	.000
	SP	1	(Constant)	566.173	4.028		140.562	.000
			Gender	-18.815	2.523	-.096	-7.458	.000
PV6		1	(Constant)	594.036	3.769		157.596	.000
	FL		Gender	-40.791	2.371	-.227	-17.202	.000
	NL	1	(Constant)	550.486	4.148		132.696	.000
			Gender	-25.085	2.637	-.133	-9.512	.000
	SP	1	(Constant)	563.697	4.019		140.272	.000
			Gender	-18.800	2.517	-.096	-7.469	.000
PV7	FL	1	(Constant)	601.261	3.814		157.632	.000
			Gender	-45.209	2.399	-.247	-18.841	.000
	NL	1	(Constant)	545.879	4.148		131.615	.000
			Gender	-23.050	2.637	-.122	-8.742	.000
	SP	1	(Constant)	567.004	4.008		141.453	.000
			Gender	-20.338	2.511	-.104	-8.101	.000
PV8		1	(Constant)	597.890	3.774		158.419	.000
	FL		Gender	-43.335	2.374	-.240	-18.253	.000

	NL	1	(Constant)	542.990	4.209		128.992	.000
			Gender	-21.387	2.676	-.112	-7.992	.000
	SP	1	(Constant)	568.301	3.986		142.589	.000
			Gender	-20.002	2.496	-.103	-8.013	.000
PV9	FL	1	(Constant)	601.783	3.757		160.179	.000
			Gender	-45.081	2.363	-.250	-19.075	.000
	NL	1	(Constant)	544.661	4.206		129.502	.000
			Gender	-21.161	2.674	-.111	-7.914	.000
	SP	1	(Constant)	569.423	3.977		143.162	.000
			Gender	-20.734	2.491	-.107	-8.323	.000
PV10		1	(Constant)	595.946	3.828		155.674	.000
	FL		Gender	-42.090	2.408	-.230	-17.478	.000
	NL	1	(Constant)	547.115	4.203		130.163	.000
			Gender	-23.708	2.672	-.124	-8.872	.000
	SP	1	(Constant)	563.356	3.983		141.439	.000
			Gender	-17.310	2.495	-.089	-6.939	.000

Note. a. Dependent Variable: Plausible Value 2-10 in Reading
b. Weighted Least Squares Regression - Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGHT

Appendix I

Moderator effect of reading proficiency on science performance (PV2-10)

	Country	Model		Unstandardized Coefficients		Standardized Coefficients		
				B	Std. Error	Beta	t	Sig.
	Finland	1	(Constant)	-.135	.009		-15.224	.000
			Zscore: PV2 in Reading	.867	.007	.869	126.184	.000
			Dummy_Gender	.187	.012	.102	15.107	.000
			Science self-efficacy	.069	.005	.089	13.228	.000
		2	(Constant)	-.141	.009		-15.874	.000
			Zscore: PV2 in Reading	.864	.007	.866	126.282	.000
			Dummy_Gender	.182	.012	.099	14.751	.000
			Science self-efficacy	.076	.005	.098	14.376	.000
			Read * Efficacy	.037	.005	.049	7.519	.000
	Netherlands	1	(Constant)	-.166	.008		-20.251	.000
			Zscore: PV2 in Reading	.899	.006	.895	146.888	.000
			Dummy_Gender	.235	.012	.121	20.022	.000
			Science self-efficacy	.056	.005	.073	11.979	.000
PV2		2	(Constant)	-.169	.008		-20.659	.000
			Zscore: PV2 in Reading	.896	.006	.892	146.265	.000
			Dummy_gender	.233	.012	.121	19.947	.000
			Science self-efficacy	.068	.005	.089	13.277	.000
			Read * Efficacy	.028	.005	.037	5.649	.000
	Singapore	1	(Constant)	-.021	.008		-2.565	.010
			Science self-efficacy	.081	.005	.091	16.240	.000
			Zscore: PV2 in Reading	.891	.006	.885	157.270	.000
			Dummy_Gender	.244	.011	.120	21.921	.000
		2	(Constant)	-.025	.008		-3.038	.002
			Science self-efficacy	.081	.005	.092	16.325	.000
			Zscore: PV2 in Reading	.889	.006	.882	155.760	.000
			Dummy_Gender	.242	.011	.120	21.822	.000
			Read * Efficacy	.016	.004	.020	3.649	.000
	Finland	1	(Constant)	-.159	.009		-17.932	.000
			Science self-efficacy	.061	.005	.078	11.719	.000
			Dummy_gender	.230	.012	.125	18.593	.000
			Zscore: PV3 in Reading	.869	.007	.880	128.722	.000
		2	(Constant)	-.165	.009		-18.649	.000
			Science self-efficacy	.068	.005	.087	12.988	.000
			Dummy_gender	.224	.012	.122	18.232	.000
			Zscore: PV3 in Reading	.866	.007	.877	128.939	.000
			Read * Efficacy	.039	.005	.053	8.115	.000
	Netherlands	1	(Constant)	-.162	.008		-19.199	.000

			Science self-efficacy	.056	.005	.074	11.609	.000
			Dummy_gender	.232	.012	.121	19.201	.000
PV3			Zscore: Plausible Value 3 in Reading	.890	.006	.885	139.150	.000
		2	(Constant)	-.166	.008		-19.716	.000
			Science self-efficacy	.071	.005	.094	13.441	.000
			Dummy_gender	.230	.012	.120	19.124	.000
			Zscore: P V3 in Reading	.887	.006	.881	138.605	.000
			Read_Efficacy	.034	.005	.047	6.815	.000
	Singapore	1	(Constant)	-.026	.008		-3.430	.001
			Science self-efficacy	.059	.005	.067	12.423	.000
			Dummy_gender	.242	.011	.119	22.785	.000
			Zscore: PV3 in Reading	.920	.005	.901	167.800	.000
		2	(Constant)	-.028	.008		-3.633	.000
			Science self-efficacy	.059	.005	.067	12.461	.000
			Dummy_gender	.241	.011	.119	22.724	.000
			Zscore: PV3 in Reading	.919	.006	.900	166.228	.000
			Read * Efficacy	.007	.004	.009	1.737	.082
	Finland	1	(Constant)	-.152	.009		-17.001	.000
			Science self-efficacy	.066	.005	.084	12.573	.000
			Dummy_gender	.218	.012	.118	17.491	.000
			Zscore: PV4 in Reading	.873	.007	.876	127.108	.000
		2	(Constant)	-.158	.009		-17.702	.000
			Science self-efficacy	.073	.005	.094	13.860	.000
			Dummy_gender	.212	.012	.115	17.098	.000
			Zscore: PV4 in Reading	.870	.007	.873	127.197	.000
			Read * Efficacy	.040	.005	.053	8.174	.000
	Netherlands	1	(Constant)	-.176	.008		-22.006	.000
			Science self-efficacy (WLE)	.058	.005	.077	12.775	.000
			Dummy_gender	.223	.011	.117	19.488	.000
PV4			Zscore: Plausible Value 4 in Reading	.892	.006	.896	148.846	.000
		2	(Constant)	-.179	.008		-22.349	.000
			Science self-efficacy (WLE)	.068	.005	.091	13.701	.000
			Dummy_gender	.222	.011	.116	19.418	.000
			Zscore: Plausible Value 4 in Reading	.890	.006	.894	148.285	.000
			Read_Efficacy	.024	.005	.032	4.944	.000
	Singapore	1	(Constant)	-.018	.008		-2.272	.023
			Science self-efficacy	.064	.005	.073	12.989	.000

			Dummy_gender	.252	.011	.125	22.875	.000
			Zscore: P V 4 in Reading	.911	.006	.893	159.430	.000
		2	(Constant)	-.021	.008		-2.586	.010
			Science self-efficacy	.065	.005	.073	13.048	.000
			Dummy_gender	.251	.011	.124	22.793	.000
			Zscore: PV4 in Reading	.909	.006	.891	157.914	.000
			Read*Efficacy	.011	.004	.014	2.500	.012
	Finland	1	(Constant)	-.142	.009		-15.779	.000
			Science self-efficacy	.056	.005	.073	10.646	.000
			Dummy_gender	.212	.013	.116	16.871	.000
			Zscore: PV 5 in Reading	.866	.007	.873	125.093	.000
		2	(Constant)	-.147	.009		-16.403	.000
			Science self-efficacy	.063	.005	.082	11.819	.000
			Dummy_gender	.206	.013	.113	16.498	.000
			Zscore: PV5 in Reading	.863	.007	.870	125.021	.000
			Read * Efficacy	.037	.005	.049	7.439	.000
	Netherlands	1	(Constant)	-.181	.008		-21.381	.000
			Science self-efficacy (WLE)	.066	.005	.087	13.703	.000
			Dummy_gender	.227	.012	.119	18.683	.000
PV5			Zscore: Plausible Value 5 in Reading	.890	.006	.882	138.086	.000
		2	(Constant)	-.184	.008		-21.786	.000
			Science self-efficacy (WLE)	.078	.005	.104	14.857	.000
			Dummy_gender	.225	.012	.118	18.608	.000
			Zscore: Plausible Value 5 in Reading	.887	.006	.879	137.597	.000
			Read_Efficacy	.029	.005	.039	5.667	.000
	Singapore	1	(Constant)	-.012	.008		-1.493	.136
			Science self-efficacy	.073	.005	.083	14.612	.000
			Dummy_gender	.236	.011	.116	21.106	.000
			Zscore: PV5 in Reading	.898	.006	.886	156.693	.000
		2	(Constant)	-.018	.008		-2.182	.029
			Science self-efficacy	.074	.005	.083	14.736	.000
			Dummy_gender	.234	.011	.115	20.992	.000
			Zscore: P V5 in Reading	.894	.006	.882	155.334	.000
			Read * Efficacy	.023	.004	.028	5.131	.000
	Finland	1	(Constant)	-.137	.009		-15.114	.000
			Science self-efficacy	.073	.005	.094	13.854	.000
			Dummy_gender	.193	.013	.105	15.390	.000
			Zscore: PV6 in Reading	.869	.007	.866	124.724	.000
		2	(Constant)	-.142	.009		-15.806	.000

			Science self-efficacy	.081	.005	.104	15.085	.000
			Dummy_gender	.188	.013	.102	15.015	.000
			Zscore: PV6 in Reading	.866	.007	.863	124.838	.000
			Read* Efficacy	.040	.005	.053	7.977	.000
	Netherlands	1	(Constant)	-.198	.008		-23.702	.000
			Science self-efficacy (WLE)	.052	.005	.070	11.012	.000
			Dummy_gender	.254	.012	.133	21.099	.000
PV6			Zscore: Plausible Value 6 in Reading	.883	.006	.892	140.706	.000
		2	(Constant)	-.202	.008		-24.167	.000
			Science self-efficacy (WLE)	.066	.005	.088	12.688	.000
			Dummy_gender	.252	.012	.132	21.021	.000
			Zscore: Plausible Value 6 in Reading	.880	.006	.888	140.317	.000
			Read_Efficacy	.032	.005	.043	6.337	.000
	Singapore	1	(Constant)	-.004	.008		-.473	.636
			Science self-efficacy (WLE)	.063	.005	.071	12.558	.000
			Dummy_gender	.235	.011	.116	21.182	.000
			Zscore: Plausible Value 6 in Reading	.902	.006	.889	157.253	.000
		2	(Constant)	-.009	.008		-1.065	.287
			Science self-efficacy (WLE)	.063	.005	.072	12.675	.000
			Dummy_gender	.234	.011	.116	21.070	.000
			Zscore: Plausible Value 6 in Reading	.898	.006	.886	155.646	.000
			Read_Efficacy	.019	.004	.024	4.375	.000
	Finland	1	(Constant)	-.175	.009		-19.248	.000
			Science self-efficacy	.066	.005	.083	12.358	.000
			Dummy_gender	.241	.013	.129	19.025	.000
			Zscore: PV7 in Reading	.883	.007	.877	127.156	.000
		2	(Constant)	-.181	.009		-19.981	.000
			Science self-efficacy	.073	.005	.093	13.683	.000
			Dummy_gender	.235	.013	.126	18.639	.000
			Zscore: PV7 in Reading	.879	.007	.874	127.294	.000
			Read * Efficacy	.042	.005	.055	8.387	.000
	Netherlands		(Constant)	-.181	.008		-21.847	.000
			Science self-efficacy (WLE)	.061	.005	.081	12.987	.000

			Dummy_gender	.262	.012	.137	21.910	.000
PV7			Zscore: Plausible Value 7 in Reading	.892	.006	.890	142.127	.000
		2	(Constant)	-.185	.008		-22.343	.000
			Science self-efficacy (WLE)	.076	.005	.100	14.606	.000
			Dummy_gender	.260	.012	.136	21.844	.000
			Zscore: Plausible Value 7 in Reading	.889	.006	.886	141.809	.000
			Read_Efficacy	.033	.005	.044	6.601	.000
	Singapore	1	(Constant)	-.009	.008		-1.050	.294
			Science self-efficacy (WLE)	.068	.005	.076	13.376	.000
			Dummy_gender	.232	.011	.115	20.650	.000
			Zscore: Plausible Value 7 in Reading	.904	.006	.887	154.840	.000
		2	(Constant)	-.013	.008		-1.573	.116
			Science self-efficacy (WLE)	.068	.005	.077	13.459	.000
			Dummy_gender	.231	.011	.115	20.550	.000
			Zscore: Plausible Value 7 in Reading	.901	.006	.884	153.571	.000
			Read_Efficacy	.017	.004	.022	3.890	.000
	Finland	1	(Constant)	-.164	.009		-18.203	.000
			Science self-efficacy	.061	.005	.077	11.482	.000
			Dummy_gender	.224	.013	.121	17.885	.000
			Zscore: PV8 in Reading	.881	.007	.877	127.023	.000
		2	(Constant)	-.170	.009		-18.932	.000
			Science self-efficacy	.068	.005	.087	12.790	.000
			Dummy_gender	.219	.012	.118	17.507	.000
			Zscore: PV 8 in Reading	.878	.007	.874	127.198	.000
			Read*Efficacy	.041	.005	.054	8.279	.000
	Netherlands		(Constant)	-.176	.008		-21.218	.000
			Science self-efficacy	.059	.005	.078	12.430	.000
			Dummy_gender	.257	.012	.134	21.553	.000
			Zscore: PV8 in Reading	.874	.006	.889	142.353	.000
PV8		2	(Constant)	-.180	.008		-21.657	.000
			Science self-efficacy	.072	.005	.095	13.827	.000
			Dummy_gender	.255	.012	.133	21.489	.000
			Zscore: PV8 in Reading	.871	.006	.887	141.987	.000
			Read * Efficacy	.030	.005	.040	5.973	.000
	Singapore	1	(Constant)	-.031	.008		-3.963	.000

			Science self-efficacy (WLE)	.055	.005	.062	11.228	.000
			Dummy_gender	.253	.011	.126	23.308	.000
			Zscore: Plausible Value 8 in Reading	.914	.006	.897	161.818	.000
		2	(Constant)	-.035	.008		-4.407	.000
			Science self-efficacy (WLE)	.055	.005	.063	11.309	.000
			Dummy_gender	.252	.011	.125	23.212	.000
			Zscore: Plausible Value 8 in Reading	.912	.006	.895	160.411	.000
			Read_Efficacy	.015	.004	.019	3.549	.000
	Finland	1	(Constant)	-.157	.009		-17.182	.000
			Science self-efficacy	.076	.005	.097	14.276	.000
			Dummy_gender	.220	.013	.119	17.343	.000
			Zscore: PV 9 in Reading	.876	.007	.870	124.692	.000
		2	(Constant)	-.164	.009		-18.039	.000
			Science self-efficacy	.084	.005	.108	15.727	.000
			Dummy_gender	.214	.013	.116	16.969	.000
			Zscore: PV9 in Reading	.873	.007	.867	125.188	.000
			Read*Efficacy	.046	.005	.061	9.260	.000
	Netherlands		(Constant)	-.172	.008		-20.916	.000
			Science self-efficacy (WLE)	.063	.005	.083	13.478	.000
			Dummy_gender	.238	.012	.125	20.206	.000
PV9			Zscore: Plausible Value 9 in Reading	.874	.006	.890	143.675	.000
			(Constant)	-.176	.008		-21.451	.000
			Science self-efficacy (WLE)	.078	.005	.103	15.227	.000
			Dummy_gender	.236	.012	.124	20.138	.000
			Zscore: Plausible Value 9 in Reading	.871	.006	.887	143.335	.000
			Read_Efficacy	.034	.005	.047	7.002	.000
	Singapore	1	(Constant)	-.033	.008		-4.193	.000
			Science self-efficacy	.072	.005	.081	14.580	.000
			Dummy_gender	.263	.011	.129	24.005	.000
			Zscore: PV9 in Reading	.916	.006	.891	160.875	.000
		2	(Constant)	-.036	.008		-4.553	.000
			Science self-efficacy	.072	.005	.081	14.652	.000
			Dummy_gender	.261	.011	.129	23.914	.000
			Zscore: PV9 in Reading	.914	.006	.889	159.287	.000
			Read * Efficacy	.013	.004	.016	2.964	.003

PV 10	**Finland**	1	(Constant)	-.151	.009		-17.326	.000
			Science self-efficacy	.059	.005	.075	11.486	.000
			Dummy_Gender	.204	.012	.111	16.808	.000
			Zscore: PV10 in Reading	.877	.007	.882	131.561	.000
		2	(Constant)	-.156	.009		-17.864	.000
			Science self-efficacy	.065	.005	.083	12.465	.000
			Dummy_Gender	.200	.012	.108	16.483	.000
			Zscore: PV10 in Reading	.874	.007	.880	131.501	.000
			Read * Efficacy	.031	.005	.041	6.485	.000
	Netherlands	1	(Constant)	-.181	.008		-21.476	.000
			Science self-efficacy	.059	.005	.078	12.260	.000
			Dummy_Gender	.257	.012	.134	21.209	.000
			Zscore: PV10 in Reading	.879	.006	.888	140.241	.000
		2	(Constant)	-.184	.008		-21.843	.000
			Science self-efficacy	.070	.005	.093	13.363	.000
			Dummy_Gender	.256	.012	.133	21.130	.000
			Zscore: PV10 in Reading	.876	.006	.886	139.641	.000
			Read * Efficacy	.026	.005	.036	5.253	.000
	Singapore	1	(Constant)	-.011	.008		-1.439	.150
			Science self-efficacy	.057	.005	.065	11.763	.000
			Dummy_gender	.227	.011	.113	21.047	.000
			Zscore: PV10 in Reading	.917	.006	.895	162.012	.000
		2	(Constant)	-.016	.008		-1.973	.049
			Science self-efficacy	.058	.005	.066	11.872	.000
			Dummy_gender	.226	.011	.112	20.944	.000
			Zscore: P V10 in Reading	.914	.006	.892	160.400	.000
			Read * Efficacy	.017	.004	.022	4.040	.000

Note. a. Dependent Variable: Zscore: Plausible Value 2 (PV2) in Science to Plausible Value 10 (PV10) in Science
b. Weighted Least Squares Regression - Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGHT

Appendix J

Regression residual for regression (PV1-10)

	Country	Model		Sum of Squares	df	Mean Square	F	Sig.
	Finland	1	Regression	34390.444	3	11463.481	5868.828	.000[c]
			Residual	10676.644	5466	1.953		
			Total	45067.088	5469			
		2	Regression	34515.004	4	8628.751	4468.892	.000[d]
			Residual	10552.084	5465	1.931		
			Total	45067.088	5469			
	Netherlands	1	Regression	132905.723	3	44301.908	7207.488	.000[c]
			Residual	31052.877	5052	6.147		
PV1			Total	163958.599	5055			
		2	Regression	133230.486	4	33307.622	5475.012	.000[d]
			Residual	30728.113	5051	6.084		
			Total	163958.599	5055			
	Singapore	1	Regression	38829.518	3	12943.173	10507.659	.000[c]
			Residual	7384.549	5995	1.232		
			Total	46214.067	5998			
		2	Regression	38839.624	4	9709.906	7892.281	.000[d]
			Residual	7374.443	5994	1.230		
			Total	46214.067	5998			
	Finland	1	Regression	33810.233	3	11270.078	6082.261	.000[c]
			Residual	10128.181	5466	1.853		
			Total	43938.414	5469			
		2	Regression	33913.936	4	8478.484	4622.177	.000[d]
			Residual	10024.478	5465	1.834		
			Total	43938.414	5469			
	Netherlands	1	Regression	137760.591	3	45920.197	7614.629	.000[c]
			Residual	30466.203	5052	6.031		
PV2			Total	168226.795	5055			
		2	Regression	137951.872	4	34487.968	5753.895	.000[d]
			Residual	30274.923	5051	5.994		
			Total	168226.795	5055			
	Singapore	1	Regression	38306.257	3	12768.752	9335.329	.000[c]
			Residual	8199.890	5995	1.368		
			Total	46506.146	5998			
		2	Regression	38324.436	4	9581.109	7019.213	.000[d]
			Residual	8181.711	5994	1.365		
			Total	46506.146	5998			
	Finland	1	Regression	34488.716	3	11496.239	6260.805	.000[c]
			Residual	10036.798	5466	1.836		
			Total	44525.515	5469			
		2	Regression	34608.214	4	8652.054	4767.777	.000[d]

			Residual	9917.300	5465	1.815		
			Total	44525.515	5469			
	Netherlands	1	Regression	132709.001	3	44236.334	6893.676	.000[c]
			Residual	32418.402	5052	6.417		
PV3			Total	165127.403	5055			
		2	Regression	133004.343	4	33251.086	5228.370	.000[d]
			Residual	32123.060	5051	6.360		
			Total	165127.403	5055			
	Singapore	1	Regression	39410.905	3	13136.968	10531.803	.000[c]
			Residual	7477.934	5995	1.247		
			Total	46888.839	5998			
		2	Regression	39414.666	4	9853.667	7902.263	.000[d]
			Residual	7474.173	5994	1.247		
			Total	46888.839	5998			
	Finland	1	Regression	34697.640	3	11565.880	6219.519	.000[c]
			Residual	10164.629	5466	1.860		
			Total	44862.269	5469			
		2	Regression	34820.421	4	8705.105	4737.515	.000[d]
			Residual	10041.848	5465	1.837		
			Total	44862.269	5469			
	Netherlands	1	Regression	135458.081	3	45152.694	7833.538	.000[c]
			Residual	29119.844	5052	5.764		
PV4			Total	164577.925	5055			
		2	Regression	135598.296	4	33899.574	5908.521	.000[d]
			Residual	28979.629	5051	5.737		
			Total	164577.925	5055			
	Singapore	1	Regression	38483.591	3	12827.864	9571.485	.000[c]
			Residual	8034.599	5995	1.340		
			Total	46518.191	5998			
		2	Regression	38491.961	4	9622.990	7186.463	.000[d]
			Residual	8026.230	5994	1.339		
			Total	46518.191	5998			
	Finland	1	Regression	33711.575	3	11237.192	5901.984	.000[c]
			Residual	10407.091	5466	1.904		
			Total	44118.666	5469			
		2	Regression	33815.888	4	8453.972	4484.320	.000[d]
			Residual	10302.779	5465	1.885		
			Total	44118.666	5469			
	Netherlands	1	Regression	131439.685	3	43813.228	6794.783	.000[c]
			Residual	32575.645	5052	6.448		
PV5			Total	164015.330	5055			
		2	Regression	131645.499	4	32911.375	5135.503	.000[d]
			Residual	32369.831	5051	6.409		

			Total	164015.330	5055			
	Singapore	1	Regression	38479.001	3	12826.334	9248.055	.000[c]
			Residual	8314.599	5995	1.387		
			Total	46793.600	5998			
		2	Regression	38515.366	4	9628.841	6971.931	.000[d]
			Residual	8278.234	5994	1.381		
			Total	46793.600	5998			
PV6	**Finland**	1	Regression	34184.447	3	11394.816	5944.510	.000[c]
			Residual	10477.577	5466	1.917		
			Total	44662.024	5469			
		2	Regression	34305.050	4	8576.262	4525.383	.000[d]
			Residual	10356.974	5465	1.895		
			Total	44662.024	5469			
	Netherlands	1	Regression	132181.335	3	44060.445	6986.831	.000[c]
			Residual	31858.987	5052	6.306		
			Total	164040.322	5055			
		2	Regression	132432.610	4	33108.153	5290.775	.000[d]
			Residual	31607.712	5051	6.258		
			Total	164040.322	5055			
	Singapore	1	Regression	38277.089	3	12759.030	9322.127	.000[c]
			Residual	8205.250	5995	1.369		
			Total	46482.340	5998			
		2	Regression	38303.207	4	9575.802	7017.535	.000[d]
			Residual	8179.133	5994	1.365		
			Total	46482.340	5998			
PV7	**Finland**	1	Regression	35222.809	3	11740.936	6101.181	.000[c]
			Residual	10518.612	5466	1.924		
			Total	45741.422	5469			
		2	Regression	35356.483	4	8839.121	4651.524	.000[d]
			Residual	10384.939	5465	1.900		
			Total	45741.422	5469			
	Netherlands	1	Regression	133695.837	3	44565.279	7170.351	.000[c]
			Residual	31399.272	5052	6.215		
			Total	165095.109	5055			
		2	Regression	133964.397	4	33491.099	5433.976	.000[d]
			Residual	31130.712	5051	6.163		
			Total	165095.109	5055			
	Singapore	1	Regression	37766.682	3	12588.894	8974.317	.000[c]
			Residual	8409.601	5995	1.403		
			Total	46176.283	5998			
		2	Regression	37787.863	4	9446.966	6750.391	.000[d]
			Residual	8388.420	5994	1.399		
			Total	46176.283	5998			

PV8	**Finland**	1	Regression	34825.042	3	11608.347	6136.752	.000[c]
			Residual	10339.545	5466	1.892		
			Total	45164.587	5469			
		2	Regression	34953.115	4	8738.279	4676.573	.000[d]
			Residual	10211.472	5465	1.869		
			Total	45164.587	5469			
	Netherlands	1	Regression	133927.672	3	44642.557	7177.662	.000[c]
			Residual	31421.681	5052	6.220		
			Total	165349.353	5055			
		2	Regression	134148.085	4	33537.021	5429.122	.000[d]
			Residual	31201.268	5051	6.177		
			Total	165349.353	5055			
	Singapore	1	Regression	38440.010	3	12813.337	9791.707	.000[c]
			Residual	7845.001	5995	1.309		
			Total	46285.012	5998			
		2	Regression	38456.464	4	9614.116	7361.138	.000[d]
			Residual	7828.547	5994	1.306		
			Total	46285.012	5998			
PV9	**Finland**	1	Regression	34419.175	3	11473.058	5945.601	.000[c]
			Residual	10547.586	5466	1.930		
			Total	44966.761	5469			
		2	Regression	34582.102	4	8645.526	4549.769	.000[d]
			Residual	10384.659	5465	1.900		
			Total	44966.761	5469			
	Netherlands	1	Regression	133510.133	3	44503.378	7321.754	.000[c]
			Residual	30707.268	5052	6.078		
			Total	164217.401	5055			
		2	Regression	133805.357	4	33451.339	5555.783	.000[d]
			Residual	30412.044	5051	6.021		
			Total	164217.401	5055			
	Singapore	1	Regression	38809.928	3	12936.643	9784.515	.000[c]
			Residual	7926.318	5995	1.322		
			Total	46736.245	5998			
		2	Regression	38821.526	4	9705.381	7350.109	.000[d]
			Residual	7914.720	5994	1.320		
			Total	46736.245	5998			
	Finland	1	Regression	35068.614	3	11689.538	6548.285	.000[c]
			Residual	9757.519	5466	1.785		
			Total	44826.132	5469			
		2	Regression	35143.136	4	8785.784	4958.621	.000[d]
			Residual	9682.997	5465	1.772		
			Total	44826.132	5469			

	Netherlands	1	Regression	133567.714	3	44522.571	6951.005	.000[c]
			Residual	32359.068	5052	6.405		
PV10			Total	165926.783	5055			
		2	Regression	133743.542	4	33435.886	5247.596	.000[d]
			Residual	32183.241	5051	6.372		
			Total	165926.783	5055			
	Singapore	1	Regression	38448.722	3	12816.241	9916.018	.000[c]
			Residual	7748.409	5995	1.292		
			Total	46197.131	5998			
		2	Regression	38469.768	4	9617.442	7460.106	.000[d]
			Residual	7727.363	5994	1.289		
			Total	46197.131	5998			

Note. a. Dependent Variable: Zscore: Plausible Value 2-10 in Science

b. Weighted Least Squares Regression - Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGHT

c. Predictors: (Constant), Science self-efficacy, Dummy_gender, Zscore: Plausible Value 1-10 in Reading

d. Predictors: (Constant), Science self-efficacy, Dummy_gender, Zscore: Plausible Value 1-10 in Reading, Read_Efficacy

Appendix K

Model statistics summary of the regression model

	Country	Model	R	R^2	Adjusted R^2	S.E (Estimate)	Change Statistics					Durbin-Watson
							R^2 Change	F Change	df1	df2	Sig. F Change	
PV2		2	.879[b]	.772	.772	1.354365	.002	56.535	1	5465	.000	1.874
	Netherlands	1	.905[a]	.819	.819	2.455712	.819	7614.629	3	5052	.000	
		2	.906[b]	.820	.820	2.448234	.001	31.913	1	5051	.000	1.777
	Singapore	1	.908[a]	.824	.824	1.169525	.824	9335.329	3	5995	.000	
		2	.908[b]	.824	.824	1.168325	.000	13.318	1	5994	.000	1.753
	Finland	1	.880[a]	.775	.774	1.355073	.775	6260.805	3	5466	.000	
PV3		2	.882[b]	.777	.777	1.347106	.003	65.850	1	5465	.000	1.896
	Netherlands	1	.896[a]	.804	.804	2.533169	.804	6893.676	3	5052	.000	
		2	.897[b]	.805	.805	2.521853	.002	46.439	1	5051	.000	1.711
	Singapore	1	.917[a]	.841	.840	1.116854	.841	10531.803	3	5995	.000	
		2	.917[b]	.841	.840	1.116666	.000	3.016	1	5994	.082	1.830
	Finland	1	.879[a]	.773	.773	1.363675	.773	6219.519	3	5466	.000	
PV4		2	.881[b]	.776	.776	1.355538	.003	66.820	1	5465	.000	1.918
	Netherlands	1	.907[a]	.823	.823	2.400838	.823	7833.538	3	5052	.000	
		2	.908[b]	.824	.824	2.395288	.001	24.439	1	5051	.000	1.767
	Singapore	1	.910[a]	.827	.827	1.157677	.827	9571.485	3	5995	.000	
		2	.910[b]	.827	.827	1.157171	.000	6.250	1	5994	.012	1.773
	Finland	1	.874[a]	.764	.764	1.379844	.764	5901.984	3	5466	.000	
PV5		2	.875[b]	.766	.766	1.373037	.002	55.332	1	5465	.000	1.893
	Netherlands	1	.895[a]	.801	.801	2.539305	.801	6794.783	3	5052	.000	
		2	.896[b]	.803	.802	2.531521	.001	32.115	1	5051	.000	1.756
	Singapore	1	.907[a]	.822	.822	1.177677	.822	9248.055	3	5995	.000	
		2	.907[b]	.823	.823	1.175196	.001	26.330	1	5994	.000	1.745
	Finland	1	.875[a]	.765	.765	1.384508	.765	5944.510	3	5466	.000	
PV6		2	.876[b]	.768	.768	1.376643	.003	63.638	1	5465	.000	1.833
	Netherlands	1	.898[a]	.806	.806	2.511217	.806	6986.831	3	5052	.000	
		2	.899[b]	.807	.807	2.501542	.002	40.154	1	5051	.000	1.753
	Singapore	1	.907[a]	.823	.823	1.169907	.823	9322.127	3	5995	.000	
		2	.908[b]	.824	.824	1.168141	.001	19.140	1	5994	.000	1.742
	Finland	1	.878[a]	.770	.770	1.387217	.770	6101.181	3	5466	.000	
PV7		2	.879[b]	.773	.773	1.378500	.003	70.345	1	5465	.000	1.901
	Netherlands	1	.900[a]	.810	.810	2.493034	.810	7170.351	3	5052	.000	
		2	.901[b]	.811	.811	2.482595	.002	43.574	1	5051	.000	1.763
	Singapore	1	.904[a]	.818	.818	1.184386	.818	8974.317	3	5995	.000	
		2	.905[b]	.818	.818	1.182992	.000	15.135	1	5994	.000	1.799
	Finland	1	.878[a]	.771	.771	1.375358	.771	6136.752	3	5466	.000	
PV8		2	.880[b]	.774	.774	1.366939	.003	68.543	1	5465	.000	1.876
	Netherlands	1	.900[a]	.810	.810	2.493923	.810	7177.662	3	5052	.000	

		2	.901[b]	.811	.811	2.485407	.001	35.681	1	5051	.000	1.783
	Singapore	1	.911[a]	.831	.830	1.143936	.831	9791.707	3	5995	.000	
		2	.912[b]	.831	.831	1.142832	.000	12.598	1	5994	.000	1.760
PV9	Finland	1	.875[a]	.765	.765	1.389126	.765	5945.601	3	5466	.000	
		2	.877[b]	.769	.769	1.378482	.004	85.742	1	5465	.000	1.869
	Netherlands	1	.902[a]	.813	.813	2.465409	.813	7321.754	3	5052	.000	
		2	.903[b]	.815	.815	2.453772	.002	49.032	1	5051	.000	1.805
	Singapore	1	.911[a]	.830	.830	1.149850	.830	9784.515	3	5995	.000	
		2	.911[b]	.831	.831	1.149104	.000	8.783	1	5994	.003	1.764
PV 10	Finland	1	.884[a]	.782	.782	1.336087	.782	6548.285	3	5466	.000	
		2	.885[b]	.784	.784	1.331097	.002	42.059	1	5465	.000	1.922
	Netherlands	1	.897[a]	.805	.805	2.530850	.805	6951.005	3	5052	.000	
		2	.898[b]	.806	.806	2.524214	.001	27.595	1	5051	.000	1.758
	Singapore	1	.912[a]	.832	.832	1.136872	.832	9916.018	3	5995	.000	
		2	.913[b]	.833	.833	1.135422	.000	16.325	1	5994	.000	1.763

Note. a. Predictors: (Constant), Science self-efficacy (WLE), Dummy_gender, Zscore: Plausible Value 1-10 in Reading b. Predictors: (Constant), Science self-efficacy (WLE), Dummy_gender, Zscore: Plausible Value 1-10 in Reading, Read * Efficacy c. Dependent Variable: Zscore: Plausible Value 2-10 in Science d. Weighted Least Squares Regression - Weighted by FINAL TRIMMED NONRESPONSE ADJUSTED STUDENT WEIGH