Understanding Discrepancies in Trends Results of PISA, TIMSS, and O-NET and Implications for Education Policies in Thailand

Sirin Tangpornpaiboon

University College London

Institute of Education

PhD Education

Declaration

‘I, Sirin Tangpornpaiboon, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

Signed:

Date: 10/11/2022

Abstract

This research aims to understand determinants of learning outcomes and disentangle sources of discrepancies found in (national and international) assessment result trends in Thailand. It situates in a broader research problem of how learning outcomes can be improved in Thailand. When the trends differ significantly, as in Thailand, there is less confidence in deducing how education quality changed. Hence, the research question is "What could explain the discrepancies in trends of O-NET, PISA, and TIMSS, and what implications do they have on policies?". The time frame of 2011-2015 is selected as the trends differ markedly (PISA shows a significant decrease, O-NET significant increase, and TIMSS no change). The research employs a sequential explanatory mixed methods design, with the quantitative part being conducted first (macro perspectives), followed by qualitative (practitioner or micro perspectives).

The quantitative part employs analysis of descriptive statistics, OLS regressions, and multilevel modelling to investigate the potential sources of discrepancies among PISA, TIMSS, and O-NET[1], using child-level data. Results show general agreement among the assessments on the high- and low- performers, with the largest gaps among school affiliations. At the item level, Thai students perform worse on higher-skilled items over time, which could explain the decline in PISA.

The qualitative part focuses on the perspectives of principals and teachers. Semi-structured interviews were conducted with 27 practitioners of all school affiliations. Data were analysed using thematic and contribution analysis. Results show many systematic factors contributing to the score gap including student background and aspiration, ability to teach beyond the curriculum, and differential performance pressure. Additionally, increase focus on O-NET help explains trend discrepancies.

[1] Respectively Programme for International Student Assessment, Trends in Mathematics and Science Study, and Ordinary National Educational Test

Impact statement

The impact of this research can be considered within and outside academia. Within academia, the thesis contributes to the growing evidence of understanding low learning levels, as well as how assessment results can be understood and used to guide policy. Outside academia, the thesis can potentially contribute to the debate around education policymaking in Thailand.

The thesis contributes to the understanding why learning outcomes are low in Thailand. With this, it contributes to the evidence of the learning crisis across developing countries. Even though Thailand is an upper-middle income country, it still experiences low and unequal learning outcomes beyond basic literacy and numeracy. Additionally, as this research utilises different assessment data, it contributes to understanding the similarities and differences of assessments, as well as how the results can be interpreted. This is especially relevant in a time when many countries participate in more than one assessment, nationally, regionally, or globally.

For Thailand specifically, the research contributes to understanding low learning problems from both quantitative and qualitative evidence. This is useful in both quantifying the nature of the problem and understanding the problem from the perspectives of practitioners. The thesis adds value also by triangulating evidence from three assessments, one of Thailand's under-researched areas.

Outside academia, the thesis has the potential to influence education policy in many areas, especially around assessments. It calls into question the effectiveness of the current policy of what assessment results should be prioritised and how assessments are used for accountability. Further, by triangulating evidence from three assessments, it helps policymakers understand where Thailand is at in terms of learning outcomes, and what can be done to improve them.

To bring about the stated impact, the thesis's findings will be shared with various groups involved with Thai education. Notably, the thesis itself will be shared with

NIETS, the organisation that administered O-NET examinations in Thailand, who kindly granted the request for the researcher to use O-NET microdata. It will also be shared with organisations within the researcher's network. This includes a research think-tank such as the Thailand Development Research Institute, which can build upon the findings of this thesis, as they have already been working with similar data and issues. It also includes non-profit organisations such as Teach For Thailand and Equitable Education Fund, which contribute to improving the quality of Thai education. By sharing the thesis's findings with these influential organisations, it is hoped that the findings can spark a debate regarding how education quality can be improved.

The thesis will also be made available on the UCL Discovery platform for the thesis repository. As such, it can aid future researchers interested in the topic. The researcher also plans to disseminate the findings of the research through an academic journal and blog posts.

Acknowledgements

Completing this thesis has been an incredible learning journey. My supervisors, Prof. Caine Rolleston and Prof. Roy Carr-Hill had been very supportive and patient. Their feedback helped shaped this thesis enormously. I deeply appreciate their understanding and flexibility especially when I was working two jobs and writing up at the same time. The thesis has also benefitted from the feedback from Dr Benjamin Alcott, who kindly read and commented on the final draft. I am grateful to all the participants who agreed to take part in my study and the gatekeepers who kindly introduced me to the schools.

During this journey, there are times when it had not been easy. I realised how lucky I am to be surrounded by people who care for me deeply. From the fabulous people at Fab Inc. to my family, friends, and of course, my partner, thank you for having my back, especially in the most difficult moments.

My PhD would not have been possible without funding from the Anandamahidol Foundation, for which I am deeply grateful.

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

BMA	Bangkok Metropolitan Administration
DLA	Department of Local Administration
EEF	Equitable Education Fund
ESA	Education Service Authority
ESCS	Index of economic, social and cultural status
IEA	International Association for the Evaluation of Educational Achievement
IPST	Institute for the Promotion of Teaching Science and Technology
IRT	Item-response theory
MOE	Ministry of Education
MICS	Multiple Indicator Cluster Survey
MWIT	Mahidol Wittayanusorn school
NCES	National Center for Education Statistics
NESDC	Office of the National Economic and Social Development Council
NIETS	National Institute of Educational Testing Service
NSO	National Statistical Office
OBEC	Office of the Basic Education Commission
OEC	Office of the Education Council
OECD	Organisation for Economic Co-operation and Development

O-NET	Ordinary National Educational Test
ONESQA	Office for National Education Standards and Quality Assessment
OHEC	Office of the Higher Education Commission
OPEC	Office of the Private Education Commission
OTEPC	Office of the Teacher Civil Service and Educational Personnel Commission
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PV	Plausible values
SAR	Self-Assessment Report
STEM	Science, Technology, Engineering, and Mathematics
TDRI	Thailand Development Research Institute
TIMSS	Trends in International Mathematics and Science Study

Chapter One: Introduction

This thesis examines the discrepancies in trends over time of assessment results of PISA[2], TIMSS[3], and O-NET[4], as well as their implications on education policies in Thailand. Many have judged Thailand to have low learning levels (see for example Fry, 2018; OECD and UNESCO, 2016). However, less research has been conducted on how the learning outcomes fare over time. As the results show different trends over the same period, there is less confidence in how the Thai education system performs, whether it has improved, become worse, or remained the same. By investigating why the discrepancies exist, this thesis attempts to provide greater understanding on the salience of low learning outcomes and ascertain that test results are reliable and valid. The results are used to form recommendations for Thai policymakers who want to improve the performance of Thai pupils and reduce inequality. Specifically, it aims to advise how Thailand could design an education system that is fit for the growing economy in the 21st century and resilient from shocks such as the COVID-19 pandemic.

To understand the significance of the research, this chapter starts by introducing the country's context at a macro level, focusing on the Thai economy and the role of education in the country's development. The research problem that the thesis will address is presented, along with more focused research questions and methodology used. It then presents the research's contribution and an outline of the contents of the thesis.

[2] Programme for International Student Assessment
[3] Trends in International Mathematics and Science Study
[4] Ordinary National Educational Test

1.1. Country context

The chapter starts by discussing the economic background of Thailand. Education is arguably one of the most significant precursors of economic growth. From the policymakers' perspective, education can provide the country with a skilled workforce. This labour force is necessary to generate economic growth. At a macro level, education policies and priorities are shaped with these goals in mind. Hence, understanding Thailand's growth strategies and goals relate to understanding the current and future direction of the country's education priorities.

1.1.1. The Thai economy's previous growth strategies

Thailand experienced a high level of economic growth from the 1960s up to the financial crisis in 1999. It was dubbed 'the fifth tiger' (Barber, 1989; Doner, 2009; Suter, 1991) as one of the most promising countries to propel its economy into high income along with Hong Kong, South Korea, Singapore, and Taiwan. The high rates of growth are usually attributed to, amongst other factors, mobilising labour from agriculture to export-oriented manufacturing and capital accumulation (ADB, 2015; World Bank, 2020). The share of the agricultural sector in the GDP had been steadily falling from around 40% in 1960 to 10% in 1999 (ADB, 2015). By moving labour to more productive sectors, the country was able to generate a sustained level of growth. It eventually gained the status of upper-middle-income in 2011, after rebounding from the 1997 financial crisis. However, the growth rate then slowed to an average of only 2.5% between 2011-2017 in contrast to the boom period's growth of 7.8% from 1981-1990 and 4.3% post-financial crisis (2001-2010) (ADB, 2015). Total factor productivity growth had similarly halved to 1.4% between 2009-2017 compared to 3% from 1999-2008 (World Bank, 2020). With this stagnant growth, Thailand's economy appears to be stuck in a 'middle-income trap' (Eichengreen et al., 2011; Jitsuchon, 2012; Kharas & Kohli, 2011; Warr, 2005). It was unable to transform its growth strategies to reach the income of the higher-income economies. Yet at the same time, rising income and wages increase costs of production, which lowers the cost advantage of Thailand in comparison to low-income countries.

Kharas and Kohli (2011) assert that to move past the middle-income trap, the country needs to move up the value chain by becoming more capital-intensive and skilled in manufacturing. As wages increase, global competitiveness declines and the country needs to improve the quality of products offered or find new markets. This could be done by expanding into high-value-added manufacturing and high-skilled service (ADB, 2015). To do this, Thailand needs a supply of qualified workers in productive industries and fosters innovation through research and development (R&D) (World Bank, 2020). However, some structural constraints still exist as the transition away from agriculture in Thailand is not yet complete. Since 1999, the share of agriculture has remained around 10% of GDP and has not fallen further. Even more troubling is that despite a low share of GDP, a significant portion of workers is still employed in agriculture. As many as 40% of workers are in the agricultural sector as of 2015 (ADB, 2015). The majority of these are low- and semi-skilled workers who see agriculture as a ‘safe haven’ when they could not find jobs in the city (Satimanon, 2017). This has two implications. Firstly, the sector could be made more efficient and productive by moving the labour elsewhere. Secondly, the majority of Thai workers remained low-skilled and unequipped for the changes necessary to move the country to high income.

Another challenge is regional inequality in economic development. In 1999, 34.1% of income was held by the top 10% of the population while 2.5% of income belonged to the bottom 10% (World Bank, 2019a). In 2019, the figures are 28.4% and 3% respectively, showing only minimal improvement. Thailand has the most unequal income distribution in relation to other Southeast Asian countries. At the provincial level, those within the top 15% earn four times more than the bottom 15% (NESDC, 2019). The development is also heavily concentrated around Bangkok and its vicinity. The three richest regions account for 68% of GDP despite housing only 36% of the population (ADB, 2015). Additionally, as much as 30% of the gross regional product in poorer regions is from agriculture rather than manufacturing (compared to 3% in the rich regions). This problem is worsened by the unequal share of the budget given to each region. Bangkok and its vicinity receive 72% of the total budget expenditure despite having only 17% of the population (ADB, 2015). This shows that growth and development remain centred around the capital city. To grow further, it is no longer enough to limit the development in Bangkok. Reducing the regional disparity would enable more inclusive growth for the whole nation.

In the latest national strategic plan, Thailand aims to become a high-income economy in 20 years or by 2037. The vision for this period is that "Thailand becomes a developed country with security, prosperity and sustainability in accordance with the Sufficiency Economy Philosophy (National Strategy Secretariat Office, 2018, p. 5)". In line with this, strategies were introduced to support innovation and inclusive growth to propel Thailand to become one of the high-income economies including "Thailand 4.0". Yet, the policy is critiqued for neglecting structural problems such as inequality and social hierarchy (Chiengkul, 2019). The World Bank (2020) is highly critical that the vision is unattainable unless structural changes are made. Even before the pandemic, the Thai economy was lagging behind other countries in the region. Reliance on income from tourism and exports has made Thailand vulnerable to external demands (ILO, 2019). One of the suggestions made in the report is to support the innovation ecosystem with a more skilled workforce. This highlights the importance of having a quality education system that produces a workforce that is equipped with the right skills. Equally essential is identifying what specific skills are needed and evaluating whether the current education system is delivering the desired results or rather, serving as one of the key blockages on the path to development.

1.1.2. Skills shortage and skills gap

Skills shortage and skills gap remain problems for many Thai industries. Satimanon (2017) asserts that in the Thai labour market, there is a mismatch between the skills that employers seek and the skills that the workers possess as acquired from the education system. This happens both in the form of skills shortage (demand for skills is greater than the supply of workers with those skills) and skills gap (current workers do not have the right sets of skills). Additionally, to shift the economy to be on the trajectory of becoming a high-income economy, workers with specific skills are also needed.

A workforce analysis by SCB EIC (2015) identifies two main types of skilled workers that Thailand lacks, based on a survey of Thai companies in key sectors. The first is vocational-technical workers. Another is engineers and scientists. Skill mismatch is one of the leading reasons the firms were unable to fill vacancies. Even though there is a greater demand for Science, Technology, Engineering, and Mathematics (STEM) subject graduates, a large majority of students who graduated from university were enrolled in social sciences programmes. In 2007, 23% of students enrolled are in science and technology subjects (Tangkijvanich & Sasiwuttiwat, 2012). In 2017, the figure was even lower at 21% (ILO, 2019). In the Information and Communication Technology (ICT) industry, in particular, there are shortages of both high- and semi-skilled workers. The estimated shortage is around 45,000 specialists in 2017[5] (ILO, 2019). In short, this mismatch hinders the progress and growth of the industries tremendously as they are unable to fill the positions needed.

Other than shortages in STEM workers, the existing workers also lack certain skills desirable by the employers. A report on the investment climate survey by World Bank (2008) shows that employers are not satisfied with the current workers' skills

---

[5] Parts of the gap could be filled by migrant workers. However, out of 2.3 million registered migrant workers as of September 2021, only around 200,000 are high-skilled or vocational technical workers (FWAO, 2021). This accounts for less than 1% of total working population in Thailand.

in English, IT, numeracy, and creativity. A more recent report on the skills of workers in ICT suggests that Thai workers lack soft and technical skills at work (ILO, 2019). Specifically, critical thinking, analytical skills, and social and cross-cultural competencies are not sufficient. In an interview with the Association of Thai ICT industry in 2016, 90% of workers lack even the basic skills required by the companies. This lack of basic skills significantly hinders firms' ability to innovate and conduct R&D. Some firms adapt by organising training programmes or graduate programmes to ensure the workers have the required skills before starting work (Satimanon, 2017). This strategy is problematic as it does not solve the root cause of the problems and incurs additional costs to firms.

In addition, the 21st century skills are gaining importance in the face of changing global demands. A framework by the Partnership for 21st Century Learning (P21) emphasises on 4Cs (Critical Thinking, Communication, Collaboration, and Creativity), literacy (information, global, financial, etc.), and life skills (P21, 2019). In a comprehensive review, Joynes et al. (2019) added individual autonomy, including flexibility, adaptability and entrepreneurship as part of the core 21st century skills. They also stressed core knowledge areas that serve as a basis for attaining 21st century skills. These are literacy, numeracy, and STEM-associated fields of knowledge. The 21st century skills are essential as globalisation and advances in technology require new skills of workers (Saavedra & Opfer, 2012). Globally, countries are increasingly demanding high-skill workers, especially in the context of rapid development (Joynes et al., 2019). However, results from both workforce surveys (ILO, 2019) and international assessments like PISA showed that Thai students are not sufficiently equipped with the 21st century skills.

To conclude, it seems that Thailand faces both shortages of vocational technicians/high-skilled workers and shortages of current workers not having satisfactory basic skills. This is in part due to students possessing skills less required by the market (graduates of social science degrees for instance). Other reasons cited by ILO (2019) for the shortages are outdated curricula and low quality of schooling, both at basic education and university levels. Graduates seem to not only lack technical skills, but also basic ones such as English and numeracy. These skills are critical foundations of other skills required by employers. The issues have been raised consistently over time by many stakeholders, yet little progress has been made in tackling the problem of skill shortages since the release of the survey by the World Bank (2008). Improving the quality of education has been cited as part of the solution (SCB EIC, 2015) and it is indeed the focus of national reforms and education plans (Fry & Bi, 2013; OEC, 2017). Nonetheless, the problem persists. The failure to provide a sufficient number of workers with the right skills relates directly to education. Here, many suggest that education is a weak link in developing the labour force and the economy. The formal education system needs to better prepare students for the world of work, both by ensuring that students have basic literacy and numeracy skills as well as the higher-level skills desired by the industries.

1.1.3. Education and politics in Thailand

To understand why the education system fails to deliver the desired outcomes, we need to understand the many forces shaping how the education system functions in Thailand. One of the most important ones is politics.

Despite having a stable economy, Thailand is characterised by histories of uncertain and polarised political situations. There is tension among and between the military, monarchy, politicians, and different classes in society (Kanchoochat & Hewison, 2016). The power struggle between the localists (conservatives) and liberalists contributes to frequent changes in government, street protests, and eventually military coups, with the most recent in 2006 and 2014. Both resulted in temporary military rule and new constitutions. In March 2019, Thailand had its first general election in five years, with the junta leader being re-elected as a prime minister. The political instability caused by the frequent changes in governments has led to policy discontinuities, especially in education. Over the course of 20 years, 20 ministers of education have been appointed (Fry, 2018). The shortest period of time an education minister was in a post is two months. This implies that some education policies have been very short-lived. Policies such as extending compulsory education and setting up student loans were carried over while many others (such as the one tablet per child policy) were deprioritised when the minister changed (Fry, 2018; Pongwat & Mounier, 2010).

The power struggle of Thai politics has far-reaching effects on how education policies are formed. Mounier and Tangchuang (2010) argue that many of the educational reforms in Thailand failed to materialise because of conflicts in the ideologies of different groups. Specifically, they analysed the politics behind the first comprehensive nationwide reform of 1999. The reform introduced many forward-looking changes for its time, including a standardised core curriculum, quality assurance, child-centred pedagogies, decentralisation, lifelong learning, increasing share of private provision of education, etc. (Fry, 2018). It was backed by part of ruling elites and the middle class (Mounier & Tangchuang, 2010). However, during the reform implementation from 2000 to 2002, the appointed minister of education faced opposition from high-ranking officials in the ministry and later resigned (Sangnapaboworn, 2018). The ministers appointed since then had varied views on the reforms. Some showed interest, such as in 2007, when the strategies for the second phase of reform were outlined. Others, including the minister during the Yingluck government in 2011, were more interested in pushing forward their own policies and agendas. As the reform was not supported by all stakeholders, it was not fully implemented nor realised.

Mounier and Tangchuang (2010) also argue that the neoliberal ideas pushed forward by the supporters of the reform may be conflicting with perennialist objectives to maintain social hierarchy. In the perennialist view, knowledge stems from beliefs from the past and is passed on from people in a position of power. This knowledge cannot be questioned or changed. With this, conformity is encouraged, and social hierarchy is maintained. This translates to the heavy emphasis on rote learning and memorisation, which is arguably reflected in the curriculum and assessment. By making the learning child-centred, for instance, the power is shifted away from teachers as knowledge is co-created rather than being given from teachers to children. This may lead to some teachers not wanting to implement such pedagogies. At a district level, by decentralising decisions to local authorities, the central government loses its power. This conflict, they argue, is the reason why some aspects of the reforms are adopted, whilst some are disregarded. Increasing the years of compulsory education, for example, was implemented fully early on as there is less conflict of interest among stakeholders. Others, like decentralisation, are more complex. In 2003, 175 local Education Service Areas (ESAs) were created to oversee the schools instead of the central government (Fry, 2018). The main responsibilities of ESAs are to plan and support educational institutions in the area (OBEC, 2018b). However, in practice, the decision-making powers of ESAs are limited as the major decisions including budgeting and personnel management are still made centrally (OECD & UNESCO, 2016; UNICEF, 2017).

More recently, the Thai government had worked with international organisations such as World Bank and UNESCO to study in-depth about Thai education problems. Extensive reports were produced as a result of the collaborations (Lathapipat & Sondergaard, 2015; OECD & UNESCO, 2016). Nonetheless, the recommendations provided were not adopted by the Ministry of Education and changes in educational processes and outcomes have not taken place. Here, conflict of interest also drives most of the opposition. The World Bank report by Lathapipat and Sondergaard (2015) suggested closing and merging small schools to improve efficiency and quality. However, the suggestion faced strong opposition from local communities, some officials in the Ministry, as well as academics (Phuaphansawat, 2021). Now, small schools remain one of the problems hindering quality education in Thailand.

Hence, Thai education is characterised by these conflicting ideologies, rigid hierarchical system, and the power struggle among different groups. Whilst they had contributed to the 1999 reform, they also led to many policies being only partially implemented and not fully internalised. This partially contributed to the problems of skill mismatch and low learning outcomes. Going forward, the conflicts are likely to be a hurdle to improving learning outcomes in the future. Skills tested in PISA such as critical thinking and problem-solving are fundamentally at odds with many values upheld by the conservative sections of the society such as social conformity and hierarchies of power. This is especially so when the military junta sits in the parliament.

1.2. Research problem and research questions

It is impressive that the Thai economy grew as much as it did despite the chronic and structural problems of inequality, skill shortage, and uncertain political situations. However, to remain competitive in the global arena and to move away from the middle-income trap, Thailand will need future investment in its human capital through education and skills development. Section 1.1.2 discusses foundational, technical, and higher-level skills as lacking in the Thai workforce. The former (basic literacy and numeracy) are especially important as they form the basis for more advanced skills. These skills are what students are expected to have at the end of compulsory education. They are mentioned in the national education plan as part of the indicators measuring the quality of the education system (OEC, 2017).

However, based on results of the national and international assessments, recent snapshots of data demonstrate that Thai students are still not performing up to the standards of basic proficiency in literacy and numeracy. Other than having lower ranking relative to other countries, absolute scores of Thai students are low by many benchmarks. PISA results show large proportion of students performing less than level 2 (Lathapipat & Sondergaard, 2015; OECD & UNESCO, 2016), which is deemed by the OECD as the minimum level that students should attain to be able to function and contribute to the society (OECD, 2016a). Similarly, TIMSS results show the majority of Thai students performing below Intermediate level, or the minimum level of proficiency determined by a Sustainable Development Goals

(SDG) indicator (UNESCO-UIS, 2022). Average O-NET scores are also below 50%, a passing mark, in all subjects (Fry & Bi, 2013; OECD & UNESCO, 2016).

The low learning levels can hamper socio-economic development in Thailand going forward. Low learning levels contributes to lower quality of the workforce that can hinder the competitiveness of the Thai economy and its ability to propel itself into a high-income economy. With this background, the first objective of the thesis is to investigate factors that are determinants of learning outcomes. Knowing this contributes to understanding what and how improvements can be made. Additionally, as the claim of low learning levels is based on various assessment results, it is equally important to investigate whether these measures of learning outcomes are reliable and valid or not, as there can be biases due to sampling or test contents. With this, the second objective of this thesis is to understand the discrepancies among the different assessments. Knowing how relevant and useful those tests are in the context of Thailand would allow Thailand to create clearer aims of what (and which tests) should be prioritised and improved. Understanding factors influencing achievement (first objective) is crucial before investigating sources of biases and discrepancies (second objective) as it can identify factors of focus for the second objective.

At present, there is more than one assessment that measure basic literacy and numeracy in Thailand, including the national assessment – O-NET[6], and the international assessments of PISA[7], and TIMSS[8]. These assessments serve different purposes. O-NET is a national assessment used to evaluate schools (for quality assurance) and students (used in university admission). PISA and TIMSS, by contrast, are international assessments. The results of Thailand are compared against other participating countries. They can be used to benchmark the country's position in terms of academic achievement. Despite different objectives, there are some similarities. The students targeted for the assessments are all at secondary

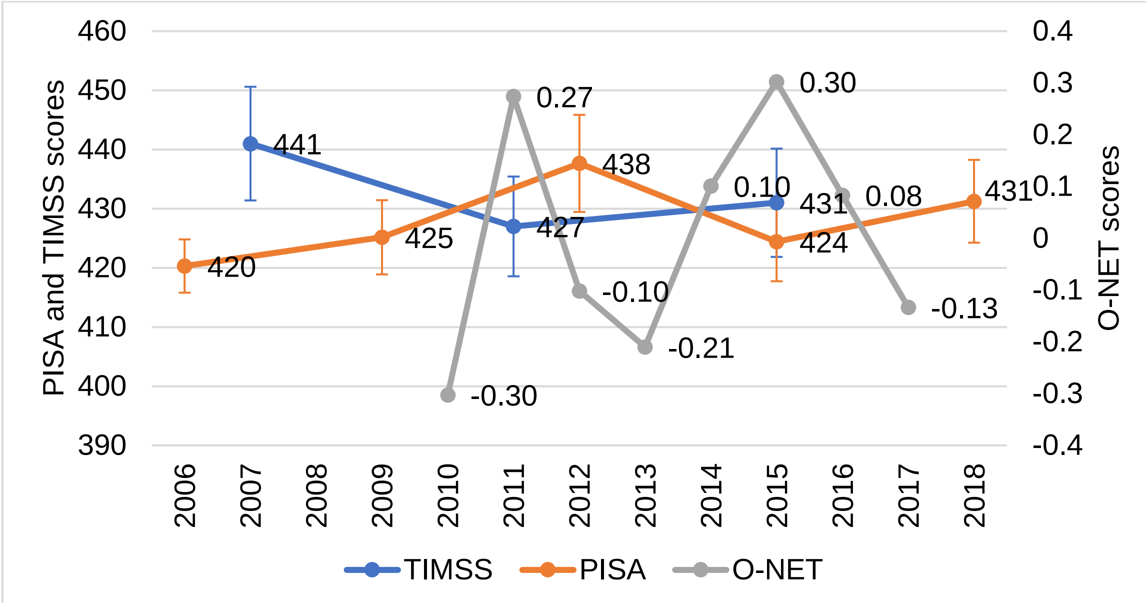
[6] Ordinary National Educational Test
[7] Programme for International Student Assessment
[8] Trends in International Mathematics and Science Study

levels, with Mathematics and Science as key overlapping subjects[9]. Since these assessments similarly assess the cognitive ability of similar groups of students, a demonstration of similar trends can strengthen the robustness of the findings regarding the cognitive performance of Thai students. Yet, if the three trends do not point to the same conclusion, it raises doubt as to what the performance over time is, and what may be driving the differences.

Initial investigations of performance trends of Thai students in Mathematics and Science during 2006-2018 are shown in Figure 1-1 and Figure 1-2[10]. The scores are all on different scales in each assessment and cannot be directly compared in terms of magnitude[11]. Nonetheless, overall trends over time appear different among PISA, TIMSS, and O-NET. O-NET shows more volatility in score change than the other two tests. Depending on the starting point, the trend can be either increasing or decreasing. TIMSS results show a slight decline while PISA results show different levels of improvement (slight improvement in Mathematics and large improvement in Science[12]), with an unusual spike in 2012. While Mathematics show statistically insignificant change over time in PISA and TIMSS, Science shows large and statistically significant changes.

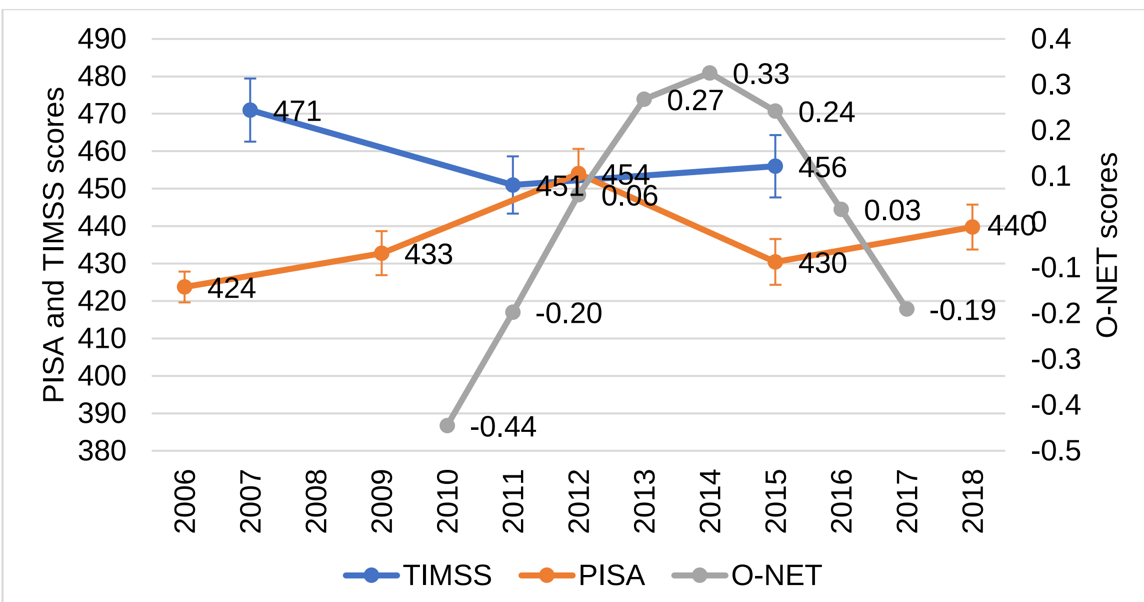
[9] See section 2.3 for more details on similarities and differences of the three assessments.
[10] Trends are shown from 2006 onwards because Science only became a major domain in PISA 2006. Hence, Science scores from the previous rounds of PISA are not comparable with the later rounds. Additionally, PISA scores differ slightly from the official figures reported by the OECD as they exclude vocational schools, so that the sample would be more comparable with TIMSS and O-NET. See section 4.3.3.3. for further details.
[11] O-NET scores are standardised for ease of interpretation. A score of zero is equal to the overall mean score of the subject.
[12] Large or meaningful improvements are changes above 0.2 standard deviation threshold. Rationale for using this is discussed in section 4.3.3.

Figure 1-1 Trends in Mathematics performance between 2006-2018



Source: Own calculation using PISA[13], TIMSS, and O-NET microdata

Figure 1-2 Trends in Science performance between 2006-2018



Source: Own calculation using PISA, TIMSS, and O-NET microdata

[13] PISA scores here and in

Figure 1-2 exclude vocational schools, for reasons that will be discussed in section 4.3.3.3.

The discrepancy is not experienced by a country with a similar context such as Indonesia. Trends over a longer period are presented in Table 1-1[14]. Here, Indonesia has more similar trends between TIMSS and PISA when compared to Thailand.

Table 1-1 Trend in PISA and TIMSS scores for Thailand and Indonesia[15]

	PISA			TIMSS			PISA-TIMSS rough agreement
	2006	**2012**	**Change**	**2007**	**2011**	**Change**	
Thailand							
Mathematics	417	427	10	441	427	-14	No
Science	421	444	**23**	471	451	**-20**	No
Indonesia							
Mathematics	391	375	-16	397	386	-11	Yes
Science	393	382	-12	427	406	**-21**	Yes

Sources: Official PISA and TIMSS reports (Mullis et al., 2016; OECD, 2014)

With this, the research aims to unpack and discuss potential sources of discrepancies in trend results with the overarching research question of **"What could explain the discrepancies in trends during 2011/2012-2015 between the three different assessments of O-NET, PISA, and TIMSS, and what implications do they have on education policy in Thailand?".** Answering this question helps shed light on not only the nature of learning outcomes, but also a broader research problem of how to improve learning outcomes in the Thai education system. The research aims to deconstruct the reasons that Thai

[14] Trends for 2012 and 2015 cannot be compared for Indonesia as the country did not participate in TIMSS 2015. Malaysia is another country that has similar context to Thailand and participated in both TIMSS and PISA. Unfortunately, it cannot be compared with Thailand as PISA data for Malaysia is only available in 2012, as technical problem prevented 2015 results from being reported.
[15] In bold are statistically significant and meaningful changes. See Chapter Four for more details.

education fails to have higher learning outcomes by quantifying the nature of the problem as well as investigating how outcomes and gaps change over time. The question can be broken down further as follows. Firstly, the data is explored quantitatively and qualitatively to answer the question "What are the factors that could explain variations in scores?" to establish determinants of learning outcomes. Then, discrepancies in trends are explored in the two subsequent questions, "In what ways are the trends from PISA/TIMSS/O-NET similar and different?" and "What are the factors contributing to the discrepancies of trends over time?". These sub-questions and methods of analysis are brought together in Table 1-2.

Table 1-2 Research questions

Research question 1: What are the factors that could explain variations in scores? (Chapter Five)	
This sub-question provides a more in-depth analysis on factors that affect learning outcomes. Understanding what factors contribute to variations in scores can provide a better understanding of how the education system functions, and what should be changed to improve it. This aspect can be understood from both analyses of quantitative and qualitative data. Descriptive statistics and regressions can help identify factors that are predictive of scores, while interviews with practitioners can gain insights into what they experience, how practices differ across high- and low-performing schools, and what had changed over time.	
Sub-questions	**Method**
1.1. How are the high-performers different from the low-performers in terms of school and student characteristics? 1.2. What school and student-level factors are predictive of student achievement?	**Quantitative** – Descriptive statistics, OLS regression and

○ Do high-performing school types retain their advantage after prior scores or socioeconomic measures are controlled for? 1.3. How do the regression results differ among the three assessments?	multilevel modelling
1.4. What could explain score gaps among high- and low-performing schools? ○ In what ways do student and school characteristics differ among school types? ○ How do different schools experience pressure to perform on different assessments? ○ How have the three assessments affected schools' priorities and practices? 1.5. How have the three assessments affected the teachers' practices in lesson planning and assessments? To what extent does the existence of PISA and TIMSS distort what is taught in classrooms? 1.6. What are the changes that happened between the period of 2012 and now in terms of curriculum, exam focus, student background, principal policy, teacher training, attitude etc.?	**Qualitative** – Thematic analysis
Research question 2: In what ways are the trends from PISA/TIMSS/O-NET similar and different? (Chapter Six)	
This sub-question provides an overview of how the score trends are similar and different. Other than serving as a benchmark for Chapter Seven, this chapter aims to shed light on quality and inequality in the Thai education system. Looking at both levels and trends over time of learning outcomes, we can determine the implications of how education quality had changed over time. Additionally, knowing where the gaps in learning outcomes are and who are the high- and low-performers would help the policymakers in resource allocation decisions. The chapter starts with exploring differences in the assessment design, before	

investigating the gaps and trends. When the tests are assessing different things, it is natural that there can be variation in scores. Knowing what information can be obtained from the assessments help guide policies in making informed decisions.	
Sub-questions	**Method**
2.1. What are the differences in assessment design that can drive different trends and variations in scores? 2.2. How are the overall trends different among the assessments? 2.3. What is the magnitude of the gaps in learning outcomes among the different subgroups? 2.4. How does the gap differ among the assessments (For example, do some school types consistently perform well in all tests)? 2.5. How are the trends similar or different over time for each subgroup?	**Quantitative** – Descriptive statistics
Research question 3: What are the factors contributing to the discrepancies of trends over time? (Chapter Seven)	
This sub-question dives into potential reasons why the discrepancy exists. Understanding why there are discrepancies in trends could aid Thai policymakers in interpreting results from both national and international assessments. As test results are used widely to justify policy decisions, it is important to be cautious on what are the implications and the conclusions that can be drawn from the results when there are discrepancies. Additionally, the results from this chapter could reveal potential problems with comparing different international assessments in general.	
Sub-questions	**Method**
3.1. Can change in subgroup composition over time account for the score trends? 3.2. How does the effect of school and student-level factors on scores change over time?	**Quantitative** – Descriptive statistics, OLS regression,

	and multilevel modelling

The three assessments, PISA, TIMSS, and O-NET, were selected as they similarly test students around the age and grade of the end of compulsory education (grade nine or 15-year-olds). It is important to know what students, after finishing school, know and can do since they will become the country's future labour force. These would be measured through their test performance. It is also acknowledged, however, that the quality of education is multifaceted and could be defined in many ways. Outcomes of schooling encompass not only cognitive skills but also character, morals, etc. (Mounier & Tangchuang, 2010). However, this research will only focus on the cognitive aspects of educational outcomes as measured by test scores. This is due to data availability and consistent linkage found between achievement and labour market outcomes and sometimes economic growth (Hanushek & Kimko, 2000; Hanushek & Woessmann, 2007). The period of 2011-2015 was selected as it is the longest overlapping period of trends that could be compared among the three assessments (see Table 1-3 for data availability). Additionally, this period also shows the largest contrast in trends among the assessments. In PISA, 2012 shows the largest change in scores, with significant improvement from 2009, then a decrease by a similar magnitude. By contrast, TIMSS shows no change in scores and O-NET shows an increase in both subjects. Mathematics and Science were selected as they share the greatest similarities in contents across the assessments. Literacy, for instance, is tested in PISA but not in TIMSS. O-NET has a Thai subject, but what are tested differ markedly from literacy.

Table 1-3 Years that the assessment took place in Thailand

Year	PISA	TIMSS	O-NET	Year	PISA	TIMSS	O-NET
1995		/		2008			
1996				2009	/		
1997				2010			/
1998				**2011**		**/**	**/**
1999		/		**2012**	**/**		**/**
2000	/			**2013**			**/**
2001				**2014**			**/**
2002				**2015**	**/**	**/**	**/**
2003	/			2016			/
2004				2017			/
2005				2018	/		/
2006	/			2019			/
2007		/		2020			

Sources: PISA, TIMSS, and O-NET official websites

2.6. Mixed methods approach

As the research problem of low learning outcomes in Thailand is multi-faceted and could not be understood from only one perspective, the research adopts a mixed methods design. On one hand, quantitative data can be used to explore and quantify the nature of the problem, identify which groups perform better or worse than others, and provide some explanations for the discrepancies. On the other hand, quantitative methods could not be used to obtain insights into the different practices schools use to prepare for the assessments, the perception of the stakeholders, and the pressure they face. Arguably, both perspectives are equally valid and provide different ways of looking at the same problem. Hence, by adopting a mixed methods design, i.e., using both quantitative and qualitative methods, the research offers a practical and comprehensive approach to the research problem. This rationale is coined by Greene et al. (1989) as 'complementarity', where two methods are used to assess different aspects of the studied phenomenon, to gain a holistic understanding of the issue.

The research adopts a sequential explanatory mixed methods design. This means the quantitative part is conducted first, then a qualitative analysis is done as a follow-up on results found from the first phase (Creswell & Plano Clark, 2011). This is so that the quantitative results could be used to inform the qualitative research. Starting from the quantitative phase can offer determinants of learning outcomes in Thailand and an overview of the discrepancies. Next, the findings were used to inform sampling and research questions in the qualitative phase. With this, the qualitative phase is targeted and built upon the quantitative findings, which offer practical recommendations for not only policymakers but also practitioners.

In mixed methods research, the findings from the quantitative and qualitative methods have to be mixed or integrated (Johnson & Onwuegbuzie, 2004). Even though this research follows a sequential process, the results are grouped by chapters, with the results of both methods corresponding to the same research questions presented together. Interpretation and analysis were done in light of results from both the quantitative and qualitative methods to reconcile the differences and form a coherent narrative. Overall, the question of factors contributing to scores is explored both from the macro perspective using quantitative analyses of nationally representative data, and from the micro school-level perspectives. Additionally, findings from both methods are analysed to recommend actions for Thai policymakers who are interested in improving the quality of Thai education.

2.7. Research contribution

This thesis is situated in multidisciplinary strands of research. It situates primarily in the field of education, but also brings in theories and frameworks from economics and statistics to help answer the research questions. The thesis relates closely to the literature studying the broad research problem of why learning levels are low in Thailand and how can they be improved. Additionally, the thesis is also related to studies investigating why trends over time of different assessments differ, and what they imply to the reliability of assessments and policies.

The thesis fills the research gaps in several ways. First, it is the first mixed methods study to look at the problem of low learning outcomes in Thailand. Other studies are generally purely quantitative or qualitative. There are few studies with multi-methods (see (Fry & Bi, 2013; OECD & UNESCO, 2016)). However, their findings are not integrated to form a coherent narrative. Second, it utilises data from assessment results of PISA, TIMSS, and O-NET to evaluate learning outcomes, which had not been done before. A study similar to this one by World Bank (2012a) compares only PISA and TIMSS. Third, it focuses on trends over time of assessment results, which is one of the under-researched areas in Thai education.

It also contributes to expanding knowledge in the field of education. By investigating the nature and reasons for low learning outcomes in Thailand, the thesis contributes to understanding the learning crisis across developing countries. A World Bank report (2018) demonstrated that many low- and middle-income countries have low learning levels. Even though Thailand managed to move beyond ensuring basic literacy and numeracy, learning levels remain below international standards. The thesis contributes to a better understanding of the nature of learning outcomes in Thailand, including the magnitude of the gaps and what can be done to improve them. Additionally, as this research utilises different assessment data, it contributes to understanding the similarities and differences of assessments, as well as how the results can be interpreted. It contributes to a growing literature critiquing the reliability of assessment results (see Brown et al., 2007; Jerrim, 2013; Wu, 2010). This is especially relevant in a time when many countries participate in more than one assessment, nationally, regionally, or globally.

Beyond academia, knowing why the three assessments show different trends would yield important implications for Thai education. Trends over time of these assessment results can be used to evaluate past and ongoing educational policies or reforms as well as guiding directions of new policies. Additionally, they can be used as an indicator of system's health, and how that changed over time. For policymakers, it would be useful to know if the implemented policies yield any significant results in learning outcomes. An in-depth analysis of where improvements have been made allows us to better target resources to help schools or groups of students that lag behind, and to gain insights on what strategies or policies could improve learning outcomes. With reliable data, relevant policy recommendations could be made so that learning outcomes can be improved, and the future workforce will have the skills that would help with the country's economic growth. For practitioners, knowing which groups of students improved and which groups are vulnerable would be an important diagnostic tool to plan teaching.

2.8. Outline

The thesis is arranged as follows; Chapter Two presents the background, Chapter Three presents the literature review, Chapter Four presents the overall methodology of the thesis, then, the findings for the three research questions are presented in Chapter Five to Chapter Seven, and Chapter Eight concludes.

Chapter Two: Background

2.1. Introduction

The chapter introduces the background of the study in two main parts. The first part provides a comprehensive overview of the assessments of PISA, TIMSS, and O-NET, with a focus on Thailand. Before comparing the results of the three assessments, it is important to understand how the assessments differ and how those differences may potentially affect results. Equally important to the assessment is the understanding of the country's education context. Hence, the second part is an overview of the Thai education system. The basic education system structure is described, key features of Thai education such as school types and decentralisation are reviewed, as well as how curriculum, pedagogy, and assessment system function. The background chapter provides contextual knowledge as a backdrop to the thesis, and most importantly to contextualise the findings in Chapters 5 – 7.

2.2. General overview of the assessments

This section provides an overview of O-NET, PISA, and TIMSS. Understanding the similarities and differences serves as the foundation for assessing whether the assessments are comparable or not, and to what extent. Additionally, it is important to note where the differences in the assessments may have potentially driven changes in learning outcomes and score trends. Some factors may affect scores relative to another assessment, but do not affect trend comparison. For example, students may be more motivated to score well on some tests than others, and hence perform better on those tests relative to other tests. However, this may not change the score trends. If the motivation does not change over time, score trends would still be comparable. However, some aspects of the assessment may drive change over time, such as changing the mode of the test or changing test difficulty.

2.2.1. Objectives

PISA and TIMSS are developed and run by the Organisation for Economic Co-operation and Development (OECD) and the International Association for the Evaluation of Educational Achievement (IEA) respectively. The purpose of PISA is to "measure how well 15-year-old students[16] approaching the end of compulsory schooling are prepared to meet the challenges of today's knowledge societies" (OECD, 2017a, p. 22). This objective highlights that PISA assessments go beyond the school curriculum. It is not enough for the students to excel in the school curriculum defined by their own countries, they need to also be able to apply the knowledge to real-life problems. By contrast, TIMSS has the goal of "helping countries make informed decisions about how to improve teaching and learning in mathematics and science" (Mullis et al., 2016, p. 2). Hence, TIMSS has a strong curricular focus, and the assessment contents are based on the curricula of the participating countries. This does not mean that TIMSS exams do not involve applications of knowledge at all as they are also parts of the countries' curricula. They, however, may be represented in a lesser proportion compared to knowing and applying basic facts and formulas in known contexts. This can affect the comparability of the two assessments as one focuses more on higher-level skills. Hence, it is important to analyse performance by content as well as cognitive domain. Another objective of the two assessments is to provide internationally comparable measures of achievement over time (Wu, 2010). Both tests, therefore, make use of trend items to ensure that the scores are comparable over time (OECD, 2017b). This is helpful when making comparisons of scores over time.

[16] In Thailand, they would either be at grade 9 (end of lower secondary) or 10 (start of upper secondary).

In contrast to the two international assessments, O-NET is a national test, developed by a public organisation, the National Institute of Educational Testing Service (NIETS). O-NET aims to assess the knowledge of students in grades 6, 9, and 12 so that the results can be used as part of assessing pass/fail for the students, school improvement, and evaluating education quality at a national level (NIETS, 2017). With this, O-NET is more high-stake to the students and schools than the other two international assessments. Teachers and schools have little power over how O-NET scores are used to assess them (see section 2.3.4). However, it is still up to the schools how much weight to put on O-NET exams as a part of students' assessments.

Because O-NET is more high stakes for schools and students than PISA and TIMSS, it is possible that students and teachers would put more effort into trying to score well in O-NET compared to PISA and TIMSS. More effort could mean that students may score higher in O-NET than they would in PISA and TIMSS. This can affect the raw scores of the examinations. If the stakes changed over time for stakeholders, it could artificially drive up/down the scores and affect the trends. Therefore, changes in this will be further investigated.

2.2.2. Population and sample

Following the objectives of each assessment, the defined population differs slightly among the tests. PISA's sampling frame is based on in-school students[17] who are 15 years old[18] (OECD, 2017b). TIMSS, on the other hand, aims to assess students' curricular knowledge. Hence, the sample is in-school students who had eight years

[17] Either full-time or part-time, academic or vocational programme, public or private schools. Exclusions include special needs schools, students who are home-schooled, out of the country, or are non-native speaker of the test language. For Thailand, international schools are also excluded as they use different curriculum than other schools (OECD, 2017b).
[18] Aged from 15 years 3 months to 16 years 2 months

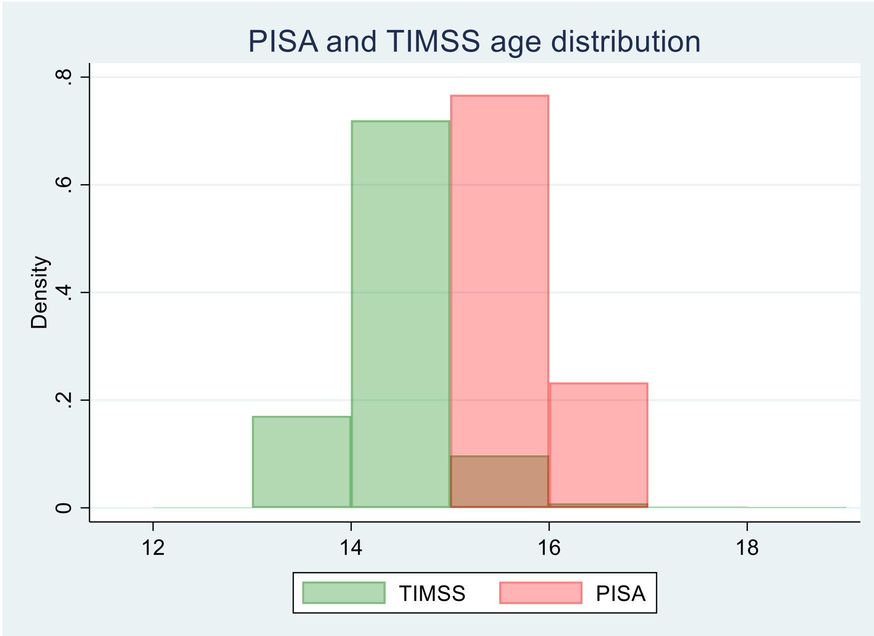
of schooling[19]. In Thailand, this is grade 8[20] (Mullis & Martin, 2013). For O-NET, the population is also grade-based with students in grade 9 as the target population.

Due to differences in the definitions of the populations, the sample collected has different age and grade distributions as shown in Figure 2-1 and Table 2-1. PISA has a slightly narrower grade distribution than TIMSS and has shares of students in different grades. Most of the students taking TIMSS are 14 years old. There are a few numbers of students who are above 16. Table 2-1 shows that the majority of students taking PISA are in grade 10, while students taking TIMSS and O-NET are in grades 8 and 9 respectively. Note that grades 8-10 are in the secondary level of education.

[19] TIMSS tests both grade 4 and grade 8 students. However, this research will only focus on grade 8 TIMSS.

[20] Exclusions include students from vocational track, very small schools with enrolment at Grade 8 less than five, and special needs schools

Figure 2-1 Age distribution of students participating PISA and TIMSS 2015



Source: Own calculation using PISA and TIMSS 2015 microdata

Table 2-1 Grade distributions of students participating PISA, TIMSS, and O-NET

Grade	PISA	TIMSS	O-NET
Grade 7	0%	-	-
Grade 8	1%	100%	-
Grade 9	24%	-	100%
Grade 10	73%	-	-
Grade 11	2%	-	-

Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

The sampling strategy of PISA and TIMSS are similar in many ways. Both have two-stage stratified sampling, with schools as primary sampling units and probability proportional to size (Mullis & Martin, 2013; OECD, 2017a). Both tests' sampling frames have school types as the main stratification variable[21]. The main difference between PISA and TIMSS's sampling strategy is that after the schools are sampled, PISA randomly selects individual students within the schools while TIMSS randomly selects classrooms, and then tests all students in that classroom[22]. With this, within-school correlation may be higher in TIMSS than in PISA, as students from the same classes are likely to be more similar (Mullis & Martin, 2013), especially so in schools that engage in academic tracking[23]. As for O-NET, there is no sampling protocol as all students are required to take the test[24].

2.2.3. Test subjects and item format

PISA tests three subjects; Reading, Mathematics, and Science. In each cycle, all subjects are tested. However, in each cycle, there is a 'main domain' that is given more attention (OECD, 2017b). Operationally, this means the contents are mapped in greater detail and more test items are included. The major domain was Reading in 2000, Mathematics in 2003, Science in 2006, and so on. The Mathematics and Science frameworks had not been developed until they became major domains. Mathematics and Science scores before 2003 and 2006 respectively are therefore not comparable and should not be used to look at trends over time. TIMSS tests two subjects, Mathematics and Science. O-NET tests four core subjects; Thai, English, Mathematics, and Science.

---

[21] The detailed stratification differs slightly. In the PISA sampling frame, the schools are separated using 16 explicit strata by school types and grade levels. Implicit strata include region, urbanisation, and school gender composition (whether same-sex or co-ed). TIMSS's strata are slightly different, with 9 explicit strata by school types and regions within OBEC2 schools.

[22] In 2015, PISA sampled 8,249 students from 273 schools and TIMSS sampled 6,482 students from 204 schools.

[23] With academic tracking, a classroom may have had higher or lower learning outcomes than the school average. It is not possible to reconcile for this difference.

[24] In 2015, a total of 656,817 students from 11,865 schools took part in the assessment.

There are broadly three types of test items: simple multiple-choice[25], complex multiple-choice[26], and constructed response[27]. The proportion of each item type differs in each assessment as well as over time, which can affect results and comparability.

2.2.4. Cognitive domains

Each assessment describes its own cognitive and content domains that the students are tested on[28]. Example items of each domain are provided in Appendix E2. The domains map out skills and knowledge that students are required to know. PISA's domains are developed by experts and agreed upon by the participating countries. TIMSS's content domains are based on the curriculum of participating countries. O-NET does not provide a cognitive domain but has a content domain based on the curriculum[29]. Knowing the domains help with understanding the range of skills tested, as well as the difficulty of each assessment.

In PISA cognitive domains, three mathematical and capabilities (or processes) are separated from the scientific ones (OECD, 2017b). The most basic capabilities are Formulate / Explain[30], requiring students to know and explain simple facts. This is followed by Employ / Evaluate[31], where more application of knowledge is required. The highest level of capability is Interpret[32] requiring students to analyse and evaluate data, as well as solving real-world problems. In TIMSS, the three cognitive domains are Knowing, Applying, and Reasoning. Knowing covers fact

[25] Simple multiple choice requires students to pick one correct answer among the available options.
[26] Complex multiple choice includes multiple questions of true/false, selection of more than one item, selecting drop-down choices to fill in multiple blanks, or "drag-and-drop" responses.
[27] Constructed response requires students to write their own answers.
[28] Full cognitive and content domains are in Appendix B1 and Appendix B2.
[29] From 2010-2019, which is the period of interest, the contents are based on the 2008 curriculum. In 2020, the blueprint changed to reflect changes in curriculum.
[30] Formulate situations mathematically and Explain phenomena scientifically
[31] Employ mathematical concepts, facts, procedures and reasoning and Evaluate and design scientific enquiry
[32] Interpret, apply and evaluate mathematical outcomes and Interpret data and evidence scientifically

recall of concepts. Applying requires students to apply knowledge to solve problems. Reasoning goes beyond Applying and asks students to engage in more complex problem solving (Mullis & Martin, 2013).

Wu (2010) analyses Mathematics items in PISA and TIMSS 2003 and found that PISA may be testing higher level skills than TIMSS. The similarities and differences in each domain as summarised in Table 2-2. There is a high degree of consensus in the first domain, with PISA requiring a slightly higher-skill level than TIMSS, including being able to recognise, offer and evaluate explanations as well as simply being able to recite the facts. For the second and third domains, skills demanded from students seem to be of a higher level in PISA than in TIMSS. For example, in Mathematics, the description of PISA extends beyond TIMSS by adding reflection about real-world context as well as critiquing mathematical models (see Appendix B1). This skill differential can partially account for score differential.

Table 2-2 Degree of agreement in PISA and TIMSS cognitive domains

PISA's cognitive domains	TIMSS's cognitive domains	Degree of agreement
Formulate / Explain	Knowing	High
Employ / Evaluate	Applying	Medium
Interpret	Reasoning	Medium

Source: Summarised from Wu (2010)

In addition to the competencies, PISA 2015 introduced the concept of cognitive demand or the type of mental process required to answer the question (OECD, 2017b). This can be viewed as another aspect of cognitive skills required of the test takers. They are separated into three categories: low, medium, and high. Low cognitive skills are usually questions that require a one-step procedure, such as recalling a fact. Medium skills require two or more steps and involve applications of knowledge. High skills require complex analysis of data (synthesise, evaluate, justify, and reason).

2.2.5. Content domains

The content domains show a high degree of similarities as shown in Table 2-3 and Table 2-4. For the most part, the listed contents overlap significantly among assessments, especially in Science. In some cases, the scope tested in PISA may extend beyond TIMSS's domains. PISA's domain of 'Change and relationships' covers not only algebra but also some statistics. Additionally, PISA's domains are less clear-cut than that of TIMSS and O-NET. 'Space and shape' also includes measurement and algebra, which belongs to another content domain of TIMSS and O-NET. 'Uncertainty and data' includes data representation in addition to data interpretation. This aspect is not covered in TIMSS's 'Data and Chance' or O-NET's 'Data Analysis and Probability'.

Table 2-3 Matching the three assessments' content domains in Mathematics

<table>
<tr><th>TIMSS's content domains</th><th>PISA's content domains</th><th>O-NET's content domains</th><th>Degree of agreement</th></tr>
<tr><td>Numbers</td><td>Quantity</td><td>Numbers and Operations</td><td>High</td></tr>
<tr><td>Algebra</td><td>Change and relationships</td><td>Algebra</td><td>Medium</td></tr>
<tr><td rowspan="2">Geometry</td><td rowspan="2">Shape and space</td><td>Geometry</td><td rowspan="2">Medium</td></tr>
<tr><td>Measurement</td></tr>
<tr><td>Data and chance</td><td>Uncertainty and data</td><td>Data Analysis and Probability</td><td>High</td></tr>
</table>

Source: Own mapping

Table 2-4 Matching PISA to TIMSS's content domains in Science

TIMSS's content domains	PISA's content domains	O-NET's content domains	Degree of agreement
Biology	Living systems	Living Things and Processes of Life	High
		Life and the Environment	
Chemistry	Physical systems	Substances and Properties of Substances	High
Physics		Forces and Motion	
		Energy	
Earth Science	Earth and space systems	Change Process of the Earth	High
		Astronomy and Space	

Source: Own mapping

The national TIMSS report also contrasts content coverage between those tested in TIMSS and those in the curriculum (IPST, 2017). From Table 2-5, there is a high level of consensus between TIMSS contents and reported curriculum, with Science showing perfect coverage. The number showing actual implementation is less than reported, which may offer some explanation for the low performance in some topics. Specifically, there is the highest discrepancy in Data and Chance, with only 1 in 3 students learned the topic despite 67% reported curriculum. Another topic is Algebra, which slightly more than half of the students were taught. Therefore, students might perform worse in these topic areas if they had not learned them.

Table 2-5 Percentage of topics covered in reported and enacted curriculum

	Mathematics	**Science**
Reported curriculum[33]		
% of topics taught to all or almost all students	85%	100%
% of topics taught to only top students	10%	0%
% not included in the curriculum	5%	0%
Enacted curriculum[34]		
% of students that learned the topics	70%	73%

Source: IPST (2017)

The differences in skills and contents can affect comparability. If students score well in one topic but not the others, they are likely to do well in assessments that emphasise those topics as seen in the study of Carnoy et al. (2013). This can affect change over time as well if the focus of test contents changes. Similarly, with skills, students may perform well in lower-level skills but not in the higher ones, and this can affect their scores. As PISA and TIMSS provided content breakdown and O-NET exams are made public, it is possible to investigate and compare the content focus. This will be done in Chapter Seven.

PISA and TIMSS have addressed the issue of comparability over time by keeping some questions the same (trend items) across the cycles. These items are used to assess how students' performance changed over time. O-NET, by contrast, could not have done this as all exam questions were made public after the exam. It has received criticisms for the validity and reliability of trends over time. Specifically, OECD and UNESCO (2016) conducted documentary analyses and interviewed officers from NIETS and found that no attempts at equating the test over time were made.

[33] Reported by the country coordinators based on curriculum contents
[34] Reported by teachers

2.2.6. Test format

PISA was implemented as a paper-based test up to 2015 when the format was shifted to being computer-based. This allows new types of questions to be included, such as questions where students can interact with the items on the screen. TIMSS and O-NET are paper-based during 2011 – 2015[35].

The differences in test format can affect the comparability of the three assessments. If students respond differently to paper-based and computer-based tests, the scores can be affected significantly, particularly as we are comparing PISA results between 2012 and 2015, when the shift happened. The effect of change in format on scores could not be ruled out. Students could be affected differently as well. It is possible that students in more advantageous schools would be more familiar with technology and perform better than students in low-performing schools that lack resources. There is no clear way of testing this effect. Jerrim (2016) utilised data from PISA 2012, where several countries participated in both paper-based and computer-based versions of the assessment. The study found that whereas there is a high correlation in terms of scores and ranking between the two formats, there are some outliers. Shanghai performed 50 points less and Brazil performed 25 more on the computer-based test (0.5 and 0.25 standard deviation). Unfortunately, Thailand did not take part in this. Hence, there may be an effect of test format on scores and trends over time.

2.2.7. Context questionnaire

In addition to the cognitive assessment, PISA and TIMSS also provide questionnaires for schools and students to complete. The student questionnaires contain basic demographics of the student as well as socioeconomic indicators

[35] In 2019, IEA introduced eTIMSS (computer-based) as an option. However, Thailand did not participate in this round of assessment.

and measures of constructs such as motivation, happiness, perception of the tested subjects, etc. The questionnaires for the principals capture basic information about the school, with topics such as autonomy, resources, discipline, teacher development, etc. Since TIMSS samples classrooms, it has additional questionnaires for Mathematics and Science teachers who teach in those classes as well[36]. By contrast, there is no context questionnaire in O-NET. Some school characteristics such as size, affiliations, and regions are available in the EMIS database. This data availability affects what variables can be put into regression models. It would not be possible to control for socioeconomic measures in O-NET. Nonetheless, as O-NET is tested in grade 6 as well as grade 9, it is possible to link students and obtain prior achievements for students, which can offer valuable information about students' backgrounds. How this data would be used is discussed in Chapter Four.

2.2.8. Methods of analysis

The test scores of the three assessments are calculated in different ways. Since both PISA and TIMSS try to measure the cognitive ability of many content domains, it is impossible to include all contents within two hours of testing time. They instead adopted a rotating design, where students only complete parts of the test items (OECD, 2017a). Item-response theory is then used to give scores to students based on the part of the questions that students answer. The scores of other questions are treated as missing values that are missing at random. Then, scaled scores are imputed for each student (called plausible values). There are five plausible values for each student and each subject in TIMSS and PISA 2000-2012. In 2015, there are ten plausible values. The immediate result of this is that the scores are not accurate at a student level as students are not taking the same

[36] There is an optional teacher questionnaire in PISA from 2015 onwards, but Thailand did not take part.

sets of questions (Lopez-Agudo et al., 2017). However, the scores are accurate[37] at a national level or subgroup level (OECD, 2017a).

PISA employs a one-parameter model up until 2012, before switching to a two-parameter model (OECD, 2017a), whereas TIMSS employs a three-parameter model that incorporates a guessing parameter. This model change in PISA is unlikely to affect analysis during 2012 and 2015 as the correlation of scores using one- and two-parameter models is very high (0.99).

The scores of both TIMSS and PISA are calculated using item-response models[38] and scaled to have a mean of 500 and a standard deviation of 100. The scores are not immediately comparable across the assessments as the set of participating countries is different in TIMSS and PISA. For PISA, the score of 500 is set to reflect the OECD mean in the year that the proficiency scale of the subjects was developed (2003 for Mathematics and 2006 for Science). For TIMSS, the score of 500 reflects the mean of the participating countries in 1995. This means that TIMSS scores reflect the pool of countries participating. Therefore, raw scores cannot be compared outright. For example, Thailand scored 424 in PISA and 431 in TIMSS 2015 Mathematics. This does not mean that they perform better in TIMSS than PISA. This thesis would not make a comparison this way. Instead, in-depth analysis can be made about performance in different cognitive or content domains (see Chapters Six and Seven).

In O-NET, as the scores are used to evaluate individual students, they need to be accurate at a student level. Therefore, such methods are not adopted in O-NET. Instead, every student completes the same test questions at the same time. The scores are calculated as the total percent correct. Each question may have a different score, reflecting different difficulties (see Appendix E2 for details).

Other than average scores, PISA and TIMSS also have proficiency levels, which describe the skills of the students, or what they can do if they obtain this level.

[37] Based on model assumptions discussed below.
[38] IRT is not without its criticisms. The model relies heavily on many assumptions, which can affect reliability of scores. Some of the critiques are discussed in Appendix B3.

Each level has a score range that it corresponds to. The full details of the proficiency levels can be found in Appendix B4. Level 2 is considered by the OECD to be a "a baseline level of proficiency that all young adults should be expected to attain in order to take advantage of further learning opportunities and participate fully in the social, economic and civic life of modern societies in a globalised world (OECD, 2016a, p. 64)". Meanwhile, SDG[39] indicator 4.1.1. considers TIMSS's intermediate level (level 2)[40] to be a minimum level of proficiency in Mathematics (UNESCO-UIS, 2022). Hence, level 2 can be a baseline to judge a country's performance against[41].

2.2.9. Reporting

For PISA and TIMSS, there are reports from OECD and IPST[42]. In terms of data, OECD reports provide an overview of the average score of each participating country, scores by subgroups, and summary statistics of items from context questionnaires (See OECD (2016a) for an example of such reports). There are also separate technical reports and assessment framework reports. In addition to the official reports, IPST published their reports in Thai (See OECD and IPST (2018)). The reports include additional information specific to Thailand not found in OECD reports. For instance, scores by school type and by region are reported.

[39] Sustainable Development Goals
[40] Defined as students being able to apply knowledge in various situations and contexts (Mullis & Martin, 2013).
[41] In Mathematics, at Level 2, students can interpret and recognise situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formulae, procedures, or conventions to solve problems involving whole numbers. They are capable of making literal interpretations of the results (OECD, 2016a, p. 191). In Science, 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 (OECD, 2016a, p. 60).
[42] Institute for the Promotion of Teaching Science and Technology, which is the organisation overseeing the implementation of PISA and TIMSS in Thailand

The sampling frame and procedures specific to Thailand are also explained here. Both reports from OECD and IPST also include analysis and policy implications of what the findings mean for the participating countries at a country level.

For O-NET, NIETS reports summary statistics of the scores[43] of each subject and grade to the general public (NIETS, 2017). There are averages by school affiliation, school size, school location (urban or rural), region, and learning module. This is publicly available on their website.

Anonymised student micro-level data is available to the public for PISA and TIMSS. For O-NET, even though microdata is not publicly available, individual-level anonymised data can be requested from NIETS.

2.2.10. Timing and frequency of the test

The three assessments took place at different times of the year. In 2015, PISA was tested in August while TIMSS and O-NET exams were in February.

The assessments are tested in a different timeline. PISA is tested every three years, starting from 2000. TIMSS is tested every four years, starting from 1995. Thailand did not participate in the 2003 round. O-NET is tested every year for grade 9 students, from 2008 onwards. The years that the assessments took place were presented in Table 1-3.

Typically, in Thailand, a semester runs from May to April. Even though schools have autonomy in shifting around the taught contents, students taking the test in February would have been studying for longer than students taking the test in August. This can affect relative scores among the tests but should not affect the comparison of trends over time, assuming that schools do not change their schedules from year to year.

[43] Calculated as percent correct

2.3. The Thai education system

2.3.1. Development of Thai secondary education

Currently, the structure of Thai basic education is 3-6-3-3, with three years in pre-primary, six years in primary, three years in lower secondary, and three years in upper secondary (OECD & UNESCO, 2016). Most schools are mixed sex, with a small exception of some private schools. The school starting age at primary grade one is six. Compulsory education is from the first year of primary to the third year of lower secondary (grade one to grade nine). The Free Education Programme, launched in 2009 promises free education for 15 years, or from pre-primary to upper secondary (UNICEF, 2017). This has not always been the case. Before 1977, secondary education was exclusive to a select few, usually children of government officers. Specifically, in 1971, just 14% of age-appropriate children were enrolled in secondary education (Nakornthap, 2018). However, with the introduction of the 1977 national education plan, massification of education began to take place. The plan, with the purpose to facilitate access and increase educational opportunities (Bualoy, 1998), resulted in changes in education structure and enrolment.

Prior to 1977, primary education was separated into lower (four years) and upper (three years) levels. Fry (2018) argues that this had led to low learning in rural areas, where most students only completed four years of lower education. Furthermore, where there are no nearby secondary schools, poor households often found it more attractive to send their children to work rather than to study (Bualoy, 1998). In order to make children stay in school longer, the reform in 1977 shifted the structure to 6-3-3 (six years of primary, three years of lower and upper secondary education) which is still in use today. To support this new policy, extended primary schools (sometimes called opportunity expanding schools, schools of expanded opportunity, or Rongrian Cayai Ogat in Thai) were established in 1987 (Fry & Bi, 2013; Varavarn, 2006). These are primary schools under the Office of the Basic Education Commission (OBEC) that had lower secondary levels (grades 7-9) added in. The policy was carried out along with a free tuition scheme to attract attendance. With these policies, access to secondary education was extended to more pupils, especially those in rural areas. In 2006, extended primary schools accounted for 21% of total secondary enrolment, comprising of 6,600 schools across Thailand (Varavarn, 2006). More recently, Sripahol (2014) reported that the number of schools has increased to 7,083 as of 2011.

Later on, in 1999, new education reform was initiated. The National Education Act of 1999 is by far one of the most comprehensive system-wide reforms in Thai education (Fry & Bi, 2013). One of the direct changes this reform had on secondary education was the extension of compulsory education from six to nine years, including primary and lower secondary levels, and the commitment to provide free 12 years of education. This led to a further increase in gross enrolment in secondary education. Today, gross enrolment exceeds 100%[44] compared to 61% in 1998 (World Bank, 2019b). Another significant shift brought by the reform is the interest in assessing education quality. Following the reform, the organisation responsible for external quality assurance, the Office for National Education Standards and Quality Assessment (ONESQA), was established in 2000. Then, the new national curriculum was drafted in 2001 (with amendments in 2008 and most recently in 2017 in Mathematics and Science). Finally, the National Institute of Educational Testing Service (NIETS) was founded in 2005 to design and implement national tests. These three major changes define how secondary education should look in terms of measurable outcomes. These changes serve as a background of how Thailand attempted to improve learning outcomes. With the comprehensive reform, the policymakers were optimistic that education quality would be improved.

[44] Gross enrolment rate can be above 100% as it includes overaged and underaged students.

2.3.2. Different school types

Thai secondary schools have a large degree of heterogeneity depending on school affiliation. The different school types have different characteristics that may have affected teaching and learning as well as learning outcomes. Due to this difference, it is used as one of the stratification variables in both PISA and TIMSS assessments. Reports of O-NET also produced summary statistics for the different school types. The summary of all school types can be found in Table 2-6 below.

The majority of students enrolled in public schools, which are under the Office of the Basic Education Commission. These schools get funding from the Ministry of Education (UNICEF, 2017). These can be further separated into extended primary schools (OBEC1), and other schools (OBEC2). The basic governance of the two school types is the same. However, the former is initially schools offering only primary level education. Then, with the secondary education reforms (Fry & Bi, 2013; Varavarn, 2006), they opened up the secondary track to “expand opportunity” to Thai children who live outside the city. As a result, they are more likely to be small and situated in rural area (Lathapipat & Sondergaard, 2015). As school funding is made on a per-head basis (UNICEF, 2017), by being small, they receive less overall funding. By being in rural area, they also have problems attracting qualified personnel (OECD & UNESCO, 2016). With these obstacles, OBEC1 schools tend to perform less well on cognitive assessments. Less OBEC1 school students are participating in PISA comparing to TIMSS and O-NET as OBEC1 schools usually enrol students up to grade 9, the last grade of compulsory education (most PISA-eligible students are in grade 10). In principle, both OBEC1 and OBEC2 schools are free to attend. In practice, however, schools can offer alternative academic programmes and charge fees (such as an English Programme or an academically intensive programme of study).

Private schools are represented in a smaller proportion, with only 14% as of 2015. Despite operating privately, private schools in Thailand face restrictions that may hinder their academic performance. Except for international schools, registered private schools have to follow the basic curriculum from the Ministry of Education. Schools are offered subsidies from the government to help with operating costs. However, if the schools accept these subsidies (called private dependent schools in PISA), they are subjected to an upper limit in tuition fees they can charge students. As of 2012, the majority (76%) of private schools rely on partial government subsidies as their main source of income (government subsidies contribute to 70% of income) (OPEC, 2013). Meanwhile, 14% are charity schools intended to serve disadvantaged students which receive 100% of subsidies[45]. 11% do not receive any subsidies (called private independent schools in PISA). The minister of education reported in 2018 that 70% of Thai private schools received subsidies (Matichon, 2018). Consequently, this limits the ability and flexibility of private schools to invest in infrastructure or teacher development. This arrangement also widens the gap in quality among private schools as some schools can choose to not accept subsidies and charge higher fees to the students (these schools are called private independent schools in PISA). Independent schools are usually famous Catholic schools with strong alumni communities (Nakornthap, 2018). Other than tuition fees, they obtain funding from donations from parents and alumni. The schools are well known for their academic prestige and attract students from advantaged backgrounds, but they account for only 11% of total private school enrolment (OPEC, 2013).

Other public schools are managed by local authorities rather than the Office of the Basic Education Commission. These are BMA schools[46] (Bangkok Metropolitan Administration) and municipal schools (Local schools outside Bangkok). These schools receive funding from their respective local authorities. BMA schools are generally well-resourced, with laboratories and air-conditioning in classrooms, as the tax revenue is high from many businesses operating in Bangkok (Nakornthap, 2018). However, they remain one of the poorest performing types of schools academically. One of the reasons may be because they are perceived as being of

---

[45] Excluded from the PISA and TIMSS sample

[46] Schools in Bangkok and neighbouring areas that are funded and supervised by Bangkok local authorities.

lower quality by parents in relation to other public schools. Therefore, most enrolments are from students of disadvantaged backgrounds (World Bank, 2012a). The municipal schools are funded similarly. However, the funding received is less than that of Bangkok as they depend on the local authorities to source the money themselves and economic activities remain centred in Bangkok. Consequently, there can be large heterogeneity in local schools, with richer provinces receiving more share while others remain poorly sourced. Similar to OBEC schools, in principle, these schools do not charge fees, but they can offer fee-paying alternative programmes.

Other special-purpose schools are Satit and Science schools. Both are selective and require prospective students to sit for examinations and score above a certain threshold to gain admission. Satit schools are affiliated with universities and teacher colleges. They serve as schools for trainee teachers. Despite teaching the same curriculum as public schools, these schools are reputed for their academic rigour and attracted students from well-off families (Nakornthap, 2018). The schools charge fees that are at similar levels to private schools. These schools follow the national curriculum and assessment. However, in addition to teaching the national curriculum, Satit schools extend their curriculum to include subjects such as independent studies or a deeper level of content of basic subjects.

Science schools, on the other hand, were set up to support science education for gifted students nationally. The schools are located in 12 provinces in all regions and have strict catchment areas for admission. Admission criteria are rigorous to identify gifted students. The schools are also more autonomous compared to public schools. Their curriculum extends to include advanced Science and Mathematics subjects. Additionally, class size is capped at 24 students per class. This figure is considerably smaller than an average public school’s class size of 30-50 in large schools. The teachers are also more specialised, with the majority graduated a master’s degree in Science or Mathematics. To support students from poorer background, the schools are all boarding schools with all school fees subsidised (PCSHS, 2018). There are also quotas for local disadvantaged students. Both types of school are different from public schools and scored better on cognitive assessments.

Given the differences in conditions each school type operates in, we would expect the learning outcomes to vary significantly among them. Hence, school type would be one of the main foci of the thesis. It would be important to learn what sets the high-performers apart from the low-performers and what can be learned or adapted to other schools to improve performance.

There are other school types not discussed here, such as international schools, religious schools, and vocational schools. These schools follow a different curricula than the national basic curriculum and may not be comparable to the types of schools presented here. For example, international schools enrol students from the top 1% of high-income families in Thailand with the goal to prepare students to attend universities abroad (Phuaphansawat, 2021). Religious schools enrol Buddhist monks and include religious teaching in the curriculum[47] (Pra Pariyat Schools, 2022). Students enrolled in international schools and religious schools usually do not participate in PISA, TIMSS, and O-NET. Vocational schools are included in PISA assessments. However, as their expectations of learning outcomes are different from students on the general academic track, they were not included in the research. Additionally, they only account for around 1-2% of total enrolment.

[47] These are the schools affiliated with the National Office of Buddhism, and not to be confused with general public schools that include Buddhism as part of their identity.

Table 2-6 School type descriptions

School type	Abbreviation	School Affiliation	Description	Proportion of enrolled[48]
Extended primary schools	OBEC1	Office of the Basic Education Commission	Primary schools (mostly in the rural area) with secondary levels added in following the reform	20%
General public schools	OBEC2	Office of the Basic Education Commission	Other public schools with secondary level not categorised as OBEC1 (schools with both primary and secondary levels before the reform)	55%
Private schools	Private	Office of the Private Education Commission	Private schools, can be independent or dependent (partially sponsored by the government)	14%
Bangkok schools	BMA	Department of Education in Bangkok	Schools in Bangkok and neighbouring areas that are funded and supervised by Bangkok local authorities[49]	1%

[48] Of grade 9 students in 2015
[49] Not all schools in Bangkok are affiliated with BMA schools.

School type	Abbreviation	School Affiliation	Description	Proportion of enrolled[48]
Municipal schools outside Bangkok	Local	Department of Local Administration	Municipal schools outside Bangkok	7%
Satit demonstration schools	Satit	Office of the Higher Education Commission	Elite schools that are used as teacher training centres for universities	Less than 1%
Chulabhorn science-focused schools	Science	Office of the Basic Education Commission	Elite schools with a specific curriculum on STEM subjects	Less than 1%[50]

Source: Enrolment data taken from NESDC (2019)

[50] Remaining 1-2% are enrolments in international schools, religious schools, and vocational schools.

2.3.3. Curriculum

Despite differences in school types, all schools are subjected to use the same curriculum and go through the same national assessment and quality assurance. Understanding the curriculum is key to understanding what knowledge and skills policymakers expect students to have. The first national curriculum of 2001 is standard-based, specifying what students should know and be able to do for each subject at each grade level. Learning units, standards, and indicators were described for each grade level. Specifically, the 2001 curriculum tries to move away from the previous emphasis on memorisation into understanding and applying knowledge (OECD & UNESCO, 2016).

The subsequent curriculum of 2008 builds upon the 2001 curriculum. It mostly has the same contents but more detail on each standard to aid teachers. This curriculum defines quality education in the Thai context and became the basis of how quality is assessed. Particularly, it supports the holistic development of students both in cognitive skills and mindsets (MOE, 2008). The key competencies students are expected to have at the end of compulsory education (grade nine or end of lower secondary) include communication, critical thinking, problem-solving, life skills, and technology application. These are emphasised more at the secondary level and are reflected in eight core subjects. The curriculum had been in use from 2008 to 2017 when it was revised again by IPST[51] in the subjects of Mathematics and Science (IPST & MOE, 2017). The improvement was done as a move toward a competency-based curriculum, with the inclusion of 21st century skills and higher-order thinking (Sangbuaphuen, 2020). The indicators and contents were streamlined, and repetitive indicators were revised (see Appendix A).

Schools follow the national curriculum. In principle, students graduating from compulsory education need to have the knowledge as stated in the curriculum

[51] The institution also oversees PISA and TIMSS in Thailand.

(tested using O-NET, see next section). Schools develop school-level curricula based on the framework of the national curriculum. This can be a joint effort between the school and its local education authority. Guidance is provided on how the curriculum can be adapted to local needs (OBEC, 2010). Note that the written curriculum may be different from the enacted curriculum as schools may lack support in applying them (Gamage & Sooksomchitra, 2004; OECD & UNESCO, 2016; Shaeffer, 2018). As a result, even though the new curriculum states many higher-order skills, it is doubtful whether what is taught in classrooms truly moved from memorisation to the application of knowledge.

There is also some degree of flexibility in what is taught to students. Other than the required subjects, schools have the flexibility to add additional subjects to the timetable based on students' needs or abilities as well as local needs, as long as the minimum number of hours dedicated to basic subjects are met (OBEC, 2010). They only account for around 16% of total teaching hours in primary and 27% in secondary. This means that schools can teach more than what the curriculum requires. In a review by OECD and UNESCO (2016), schools use these hours freely, from arranging subjects more relevant to students to cramming for O-NET exams. In terms of pedagogy, guidance is provided to teachers, yet teachers have full autonomy to design pedagogy or sequence of contents taught to students. This implies that other than the basic requirements, what is taught in classrooms can vary largely from school to school.

2.3.4. Assessment, pedagogy, and quality assurance

Based on the national curriculum, the National Institute of Educational Testing Service (NIETS) and the National Education Standards and Quality Assessment (ONESQA) together make attempts to use assessment results to improve education quality. NIETS provides many testing services to evaluate schools' and students' performance. One of the most widely used national tests is the O-NET, which tests students in core subjects at the end of primary (grade six), lower, and upper secondary (grade nine and twelve). The contents of O-NET are based on the national curriculum. Up until 2019, the contents are based on the curriculum of 2008. From 2020, onwards, the O-NET contents match the 2017 curriculum. O-NET results are given to students, schools, educational service areas, and the general public (NIETS, 2017). Each stakeholder receives different sets of results. For instance, students obtain only his/her score in comparison to the school and national average, whereas schools receive summary statistics of the schools' performance at the school level and compared to the national and local administration's averages. The national averages are also publicised each year.

The results are then used as a part of the school evaluation done by ONESQA. Quality assurance has been conducted every four years since the organisation's inception[52]. In the third round of school assessment, ONESQA placed 20% of the weight on O-NET performance, both in levels of achievement and changes over time (ONESQA, 2012)[53]. For this, schools have to submit Self-Assessment Reports (SARs). All schools – in a non-COVID year, are then inspected on-site. Schools are then given a rating, from Excellent, Very Good, Good, Fair, and Poor, as well as feedback on how improvements can be made. The results from the assessments indicate how well Thai students are performing academically. Yet, it is arguable whether quality assurance eventually helps the schools improve. The inspections rely mostly on paperwork done by the schools and there appears to be no real consequence for failing the quality assurance (Lao, 2017). Schools are also expected to use the results to guide teaching and learning strategies. However, in practice, many schools do not have the capacity to analyse the data and make meaningful adjustments to their policies (OECD & UNESCO, 2016). Hence, monitoring and external quality assurance may not lead to accountability or school improvements. In addition to the external quality insurance, schools are subjected to internal quality assurance by the local authorities.

---

[52] The first round was during 2000 – 2005, the second from 2006 – 2010, the third from 2011 – 2015, and the fourth from 2016 – 2020.
[53] O-NET remains one of the indicators in the 4th round, but it is unclear how much weight is put on this indicator.

Other than results from the two organisations, international tests have also become influential in assessing the Thai education system and guiding public policy. Thailand participated in both PISA and TIMSS from 1995. However, there is considerably greater interest and publicity in PISA results compared to TIMSS (see e.g. Mala, 2018; Tongliemnak, 2018). PISA has shaped Thailand's definition of education quality and how to assess it. For instance, average PISA scores are being used as one of the indicators of quality in the latest 20-years national education plan in addition to O-NET results (OEC, 2017). In 2013, NIETS also developed an item bank containing PISA-like questions for students to practice online (NIETS, 2017). It has been briefly discussed whether these questions should be incorporated into O-NET in later years as part of an attempt to make O-NET more like PISA (Pitiyanuwat et al., 2018). These changes broaden what the students are expected to know since what is assessed by PISA extends beyond the simple acquisition of curriculum-related knowledge.

With O-NET, there are clear guidelines on which results are accountable to the schools. By contrast, PISA and TIMSS results are not explicitly used to assess schools as the assessments are sample-based and the sampled schools are anonymous. Results are reported only at a national level, with no individual results (IPST, 2020). Students are viewed as representatives of the country rather than of individual schools. There is no guideline on whether schools receive individual or school-level results or not. In this sense, these assessments are not high stakes to the schools and teachers directly; performing poorly on them does not yield any negative consequences to the students, teachers, or head teachers.

At a national level, by contrast, results of the PISA and TIMSS tests matter for policymakers as one of the goals listed in the national education plan is the improvement of PISA scores (OEC, 2017). This resonates with a review of PISA's impacts on stakeholders by Hopkins et al. (2008). Participants are from countries that participated in PISA 2006 (including Thailand). They found that policymakers are perceived as the most significant stakeholder when it comes to PISA, as well as being most responsible for the country's results. Overall, there seems to be a disconnection between the answers of policymakers and local actors such as school principals and teachers. The schools appear to perceive PISA as being less trustworthy and results less interpretable.

At a school level, teachers are also assessed for salary and career progression. Salary progression determines increments (0-5% increase) of yearly salary (Punyasavatsut, 2019). Another way salary can progress is through career progression, which involves teachers moving to a higher professional title and status, without necessarily changing day-to-day duties. The evaluation is done by the principal as well as external evaluators. The title[54] comes with a new salary base, allowances, and extra monthly compensation. Additionally, teachers can also move up a career ladder into leadership roles such as vice principal, education supervisor, etc. Notably, learning outcomes account for 1/3 of total points in deciding progression[55] (OTEPC, 2021). This is separated into outcomes based on teachers' own assessments as well as results from O-NET. There are no set criteria on how the scoring is done based on absolute scores (i.e., there is no mandate such as 5% improvement equals five points). This is up to the discretion of the assessors. This implies that O-NET results can be high-stakes for teachers as they relate directly to career progression.

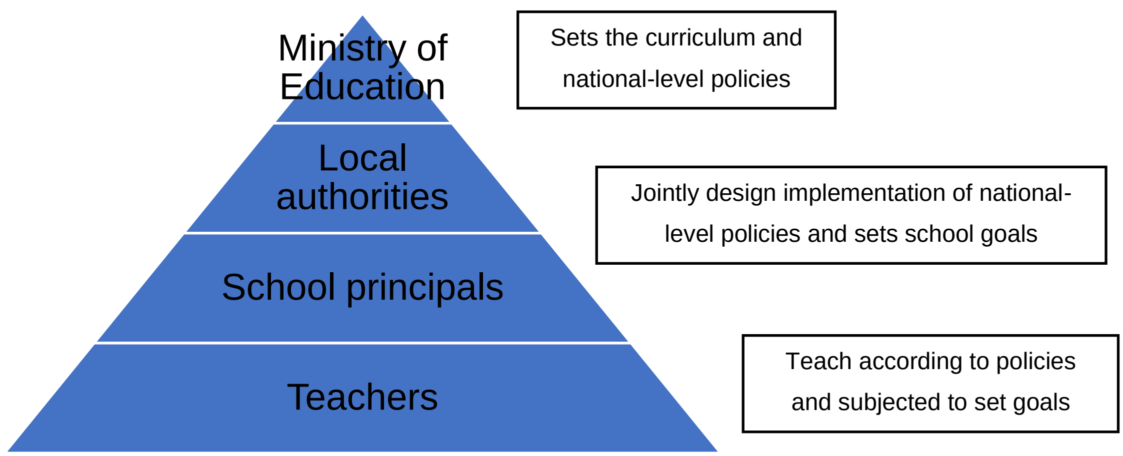
2.3.5. Governance structure

Schools' and teachers' practices and strategies are influenced by many stakeholder groups in a decentralised structure (see Figure 2-2). The structure helps with understanding how the practitioners implement national education policies and provides a framework in understanding potential conflict of interests of different groups.

[54] The titles are Teacher (level 1), Skilled teacher (level 2), Experienced teacher (level 3), Expert teacher (level 4), and Specialist teacher (level 5).
[55] Other criteria in deciding the progression include work experience, ethics, competence, student quality, and evidence-based work, with different levels requiring different standards.

At the topmost level, they have to follow policies set by the Ministry of Education and teach according to the national curriculum (OECD & UNESCO, 2016). This is the same for all school types. Students graduating from compulsory education should have the knowledge as stated in the curriculum. Next, the local authorities have some autonomy to set their policies for schools. Then, at the school level, principals have their visions and plans for the school. Finally, teachers teach contents according to the national and school curriculum but retain some levels of pedagogical freedom. This structure affects who decides what policies to prioritise and which practices to use in each school. Additionally, this dynamic can create tension between different stakeholders, such as schools and local authorities.

Figure 2-2 The decentralised structure of the Thai education



The local authorities differ among school types. The majority of schools (OBEC1, OBEC2, Local, and Science schools) are affiliated with local Education Service Areas (ESAs). In 2003, in support of the decentralisation policy initiated in the 1999 reform, 175[56] local Education Service Areas (ESAs) were created to oversee public schools instead of the central government (Fry, 2018). Since the schools in each area are not equally distributed, each ESA's management covers wide range of

[56] The number of the ESAs is now 225 after several adjustments.

numbers of schools, from 20 to 300 schools, covering 6,000-120,000 students (OBEC, 2018a). The main responsibilities of ESAs are to plan and support educational institutions in the area (OBEC, 2018b). These start from planning local policies that are aligned with both national standards and local needs. At a micro level, school curriculums are developed jointly by schools and ESAs. Following that, ESAs have responsibility over subsidies and resource mobilisation to further support educational institutions in their development. Lastly, internal monitoring and quality assurance are also conducted by ESAs. This could be a source of pressure for the schools as they have to reach target goals set by the ESAs (in terms of assessment results and other goals). To support these roles and responsibilities, each ESA is further segmented into specialised units such as planning, personnel development, monitoring, etc. This means that the ESAs can play a role in influencing school policies.

In practice, however, ESAs face many limitations that may hinder their effectiveness in supporting schools. A review by OECD and UNESCO (2016) suggests that ESAs may be facing a shortage in staff and technical experts. In using assessment results, the central authorities and ESAs themselves abstain from using more sophisticated statistical methods in analysing O-NET data. Specifically, one of the reasons the authorities gave was "local ESAs would have difficulty interpreting it (OECD & UNESCO, 2016, p. 147)". A comprehensive report by UNICEF (2017) supports the low capability of ESAs in terms of implementing the government's free education program. In the fiscal year 2013/2014 when the research had taken place, the representative sample of ESAs only visited schools in their jurisdictions once a year for monitoring purposes. UNICEF asserted that this is due to an inadequate number of staff in comparison to a large number of schools in each ESA. Other than low staff capacity, Fry (2018) also argues that many ESAs do not have enough budget to support their staff in visiting and supervising schools at greater frequencies. In addition to that, even though ESAs are granted funds to support schools, it is suggested that they may not be enough to hire an additional teacher. Therefore, ESAs may not be able to support schools personnel-wise. The limited autonomy may represent a deeper structural problem as the central authorities do not trust the ESAs enough to grant them more autonomy and funding.

Other school types differ in terms of the local authorities. For BMA schools, Bangkok Education Office (BEO) is the local authority with the same function as ESA. Private and Satit[57] schools, however, have no such local authorities. The Office of the Private Education Commission (OPEC) and the Office of the Higher Education Commission (OHEC) oversee the corresponding two school types. However, they do not have the same functioning as ESA or BEO. Rather, they operate at a national level and their duties are to propose policies that are related to the private schools to the MOE (OPEC, 2021). In addition to this, Science schools also have a separate group-level committee with policy autonomy for schools in the Science group. Science, Local, and BMA schools have governance structures similar to academies in the UK and charter schools in the US. They are run by a separate committee and similarly are allowed more autonomy than general public schools (Valant, 2019; A. West & Wolfe, 2018). This implies that they may have different curricula and teaching practices.

[57] Satit schools have the affiliated university that they have to report performance to. However, they are not very involved with day-to-day operations of the school.

Chapter Three: Literature Review

3.1. Introduction

The research is closely related to the literature of different strands and disciplines. The first strand of research of this thesis (section 3.2) is related to empirical literature explaining why the achievement levels of Thailand are low[58] and/or do not show significant improvement over time. This strand is situated in a broader literature of factors that affect pupil learning outcomes. The second strand (section 3.3) is literature investigating discrepancies in different assessment results.

This section introduces firstly the broader literature of the two main strands, with the focus on low- and middle-income countries (where possible), before discussing research specific to Thailand. It then concludes the key debates within the field and discusses how this research can contribute.

3.2. Factors that affect pupil learning outcomes

Generally, there are two main disciplines of research that investigates which factors affect student learning outcomes: economics and education. In the economics literature, the framework of Education Production Function is applied, whereas, in education literature, School Effectiveness Research is what is often referred to. Both strands describe differently what factors influence outcomes. However, they are similar in their methods of analysis. Both are primarily quantitative in nature and often employ variations of regression analyses to analyse the data. In addition to the two quantitative strands, explanations of low learning outcomes have been investigated from a policy point of view, by analysing qualitatively policies, reforms, or contextual features of countries. This strand of

[58] Compared to international standards

literature relates directly to the second sub-research question “What are the factors that could explain variations in scores?”

This section explores the theoretical underpinnings of the three strands of research (section 3.2.1), before presenting evidence from developing countries (section 3.2.2), and then research specific to Thailand (section 3.2.3).

3.2.1. Different research strands

3.2.1.1. Education production function

The education production function framework relates educational inputs to outputs (Hanushek, 2008). Here, the economic concept of production theory can be used to apply to education. Outputs can be defined as various educational outcomes such as enrolment, attainment, achievement, or wages. Meanwhile, inputs are factors that contribute to the creation of those outputs. Studies using this framework attempt to pinpoint which factors are related to achievement. Knowing this, policy recommendations can be made on where best to direct resources to. Glewwe et al. (2011) represented relationships between inputs and output as follows;

$$A = a(S, Q, C, H, I) \qquad (1)$$

where achievement (A) is a function of years of schooling (S), school quality through school and teacher characteristics (Q), child and household characteristics (C and H), and school inputs (I).

To identify which factors are related to achievement, ranges of quantitative methods were used. There is both experimental and non-experimental research under the umbrella of the education production function. For observational research, regression analysis is used to determine the relationship between inputs and learning outcomes. Variations of model specification range from simple regression (Sakellariou, 2016), and Oaxaca-Blinder decomposition (Perera & Asadullah, 2019), to fixed effect models (Lei et al., 2018). Generally, factors related to achievement can be identified with this method. Nonetheless, depending on the methods and research design, the results may be correlational rather than causal.

The education production function is not without its criticisms. Glewwe et al. (2011) noted that endogeneity and measurement error needs to be addressed for the results to be reliable. Additionally, when using non-experimental designs, omitted variable bias should be addressed (for instance, motivation and parents' mindset may not be fully captured). Townsend (2007) asserted that education production function research shows different results depending on the country, grade level, and subject of focus. This implies that the effects of resources can be country specific.

Closely linked to education production function is educational effectiveness research, which is grounded in education rather than economics.

3.2.1.2. Educational effectiveness research

Educational effectiveness research (EER) started as a distinct discipline back in the 1970s (Chapman et al., 2015). It has evolved from a narrow focus on qualitative case studies of specific successful schools to a more rigorous methodology and focus on policy. The main research question of EER is "What factors in teaching, curriculum, and learning environment at different levels such as the classroom, the school, and the above-school levels can directly or indirectly explain the differences in the outcomes of students, taking into account background characteristics, such as ability, [socioeconomic status], and prior attainment (Creemers & Kyriakides, 2008, p. 11)". In other words, studies in the strand try to explain why some schools perform better than others. The line of inquiry is similar to the education production function. In contrast with a narrower input-output model, EER also focuses on process and views education production functions as one of the strands in the discipline (Chapman et al., 2015).

One of the key principles of school effectiveness research is the hierarchical nature of education systems. To estimate the effects of teachers and schools, researchers focus on using multi-level models (Muijs & Brookman, 2015). The rationale is that education data are often clustered, with students nested within classrooms, classrooms nested within schools, and schools nested within local educational authorities. Hence, students in the same classes would be more similar than students in other classes. Ignoring this structure leads to biased estimates and standard errors (Creemers & Kyriakides, 2008). A comprehensive model of EER takes into consideration the effects at a school, class, and student level.

In recent years, there had been a shift within EER to understand the process of why things work rather than just what works. Generally, the field is primarily quantitative. There are some mixed methods research. However, the quantitative part mostly remains the primary component and much research did not fully integrate the findings (Muijs & Brookman, 2015). The recent shift in studies uses qualitative methods such as case studies and observations, this new strand of research seeks to understand how improvements can be made within schools (Hadfield & Chapman, 2015). Practitioners are involved in the research and contextual factors are given a focus. Many qualitative studies explore the variables identified in the quantitative model further to see how that work to improve learning outcomes in practice. School-related -factors such as leadership and specific pedagogies were studied to identify how they relate to learning outcomes.

3.2.1.3. Policy studies

Education Production Function and Educational Effectiveness Research focus mainly on school-, classroom-, and student-level factors. With this, less importance is put on what happens at the system level. However, systemic factors can have a significant impact on students' learning outcomes. Hence, other than the two strands of research, studies that analyse educational policies or reforms with the focus of improving learning outcomes are also of relevance. These studies utilise many tools and theoretical frameworks to analyse why certain policies or reforms work or do not work.

Political economy is one of the lenses that policies or reforms can be viewed. Political economy analysis can help understanding actions or outcomes of economic and education policies by analysing the interests, incentives, values, and norms of relevant stakeholders. DFID's (2009, p. 4) operational definition of political economy analysis is that it is "concerned with the interaction of political and economic processes in a society: the distribution of power and wealth between different groups and individuals, and the processes that create, sustain and transform these relationships over time". Hirosato and Kitamura (2009) asserted that political economy can look at an issue (such as education reform or policy development) in a comprehensive and dynamic way, taking into account various perspectives of politics, economics, institutions, society, and history. Analysis can include how elements such as interests and incentives of different stakeholders, the role of institutions and norms, and values and ideas affect policy (DFID, 2009).

Studies using political economy analyses many features of the system that can affect learning outcomes. When a reform or a policy fail to produce a desired outcome, the political economy framework can help understand why. Equally, knowing why things have gone right is useful. Understanding this complements the studies within Education Production Function and Educational Effectiveness Research. It also helps contextualising the policy recommendations from those studies.

The subsequent sections explore empirical literature that situates in the three research strands, with section 3.2.2 focusing on low- and middle-income countries other than Thailand, and section 3.2.3 focusing on studies done in the Thai context.

3.2.2. Empirical studies in other countries

According to the dynamic model of educational effectiveness, there are four levels of factors that affect student outcomes: context-, school-, classroom-, and student-level (Creemers & Kyriakides, 2008). Generally, research has included elements of these different levels. The quantitative studies are of larger scale and focus on the measurable elements. In the regression, they usually include school- and student-level variables that can explain learning outcomes. The qualitative studies instead mainly focus on high-performing schools and their unique practices. Though they remain rare, there are some mixed methods studies in the field. This section presents the findings of these studies, separating on the methods of quantitative, qualitative, and mixed methods.

3.2.2.1. Quantitative studies

Studies investigating relationships between these factors and achievement are mostly situated in the US and developed countries. Starting in 1966, achievement gap between white and black pupils was found to be from students' socioeconomic backgrounds rather than differences in school resources (Coleman et al., 1966). Other literature also found strong relationships between socioeconomic backgrounds and achievement (Sirin, 2005; Todd & Wolpin, 2007). Despite strong link between background and achievement, other studies focus on whether school-related factors make a difference. Factors studied included per-pupil expenditure (Hanushek, 2003), class size (Krueger & Whitmore, 2000), classroom variables (Creemers & Kyriakides, 2008) etc. Results vary from study to study. A comprehensive review by Townsend (2007) shows that in the UK, studies generally show small but statistically significant effects of additional resources, in terms of class size, pupil-teacher ratio or expenditure per pupil, whereas the results are more mixed in other European countries. Another strand of literature explores teacher quality and its impact on achievement (Everson, 2017; Hanushek & Rivkin, 2012; McCaffrey et al., 2003). Even though the findings generally agree on large variation of teacher's contribution, no specific measurable characteristics are consistently linked with achievement.

Evidence from developing countries increases in number with international tests like PISA, TIMSS, and PIRLS. Hanushek and Kimko (2000) employed an international dataset to show that school resources (measured using level of expenditures) are not significantly related to variations in cross-country test scores. Specifically, for developing countries, it is argued that simply increasing funding or resources may not lead to improved outcomes. After key resources are provided, further increase in spending might not lead to an improvement in achievement (Hanushek & Woessmann, 2007). Glewwe et al. (2011) conducted a literature review on studies involving effects of school resources on achievement in developing countries that were published during 1990-2010. They found most basic school supplies (i.e. textbooks, desks, chairs, electricity, etc.) generally yield positive effect on academic performance. The results are less conclusive for teacher characteristics and school organisation. The review also highlights the scarcity of high-quality studies in developing countries and how this lowers the strength of evidence.

Another strand of work focuses on explaining achievement gap among different groups. They aim to understand why one group performs better than others. Oaxaca-Blinder decomposition[59] is used to separate the score difference into the explained and unexplained portion. This decomposition technique seeks to explain the reasons behind the differences in mean outcomes between groups. The explained portion is from the change in measurable characteristics. For example, students from urban area may be more affluent than rural students, and this can partially account for the score gap. By contrast, the unexplained portion captures the change that the explained portion cannot cover. For instance, even if rural

[59] The method was first used to investigate the sources of wage differentials between population subgroups of male-female and black-white workers (Blinder, 1973; Oaxaca, 1973).

students have the same background as urban students, they may perform better than as expected of them. These are because of unmeasured factors, which could be school quality, motivation, etc. Duncan and Sandy (2013) decomposed score differences between rural and urban students. Others decompose gaps between groups of different ethnicities in Vietnam (Glewwe et al., 2015) and Peru (Sakellariou, 2008). The results are mixed; some found the background characteristics could sufficiently explain the majority of the gap (Duncan & Sandy, 2013; Sakellariou, 2008), whereas others found most part of the change remains unexplained (Glewwe et al., 2015). The findings may suggest that contextual features may significantly influence why the gap occurred, and that in many cases, measurable characteristics alone cannot explain gaps in learning outcomes.

In the field of educational effectiveness research, research focuses more on school- and classroom-level variables in predicting learning outcomes. Some factors show significant but small effects. Scheerens et al. (2013) conducted a meta-analysis of school effectiveness studies from 1984 – 2005. They found curriculum quality, learning time, being achievement-oriented, and school climate to be related to student learning outcomes. However, the effects are small (less than 0.2 standard deviation). Some variables such as teacher cooperation and adaptive teaching practices show no significant effect size. Other variables such as whether schools engage in monitoring, parental involvement, and educational leadership show small but significant effect sizes. Seidel and Shavelson (2007) conducted similar research, but including more outcomes such as learning processes and motivation. Similar to what Scheerens et al. (2013) found, in terms of learning outcomes, all but one factor produce only small effect sizes. Only one factor of domain-specific teaching approach (adapting teaching based on contents taught, such as inquiry-based learning in Science) is found to have large effects (0.3 standard deviation) on learning outcomes. Countries included in the analysis are mostly from middle- to high-income countries.

Other than looking at which factors affect learning outcomes, some studies also question the robustness of the findings. Through analysing longitudinal data of teachers, Kyriakides and Creemers (2008) found that effects of the factors change over time. Value-added of schools and teachers is larger when accounting for longer time periods, as opposed to a single year. Hence, teacher and school effects may be larger or smaller depending on study time frame.

This sub-section summarised the quantitative studies in the field. The next section explores the qualitative studies of factors affecting pupil learning outcomes.

3.2.2.2. Qualitative studies

Qualitative studies are of smaller scale in nature, relative to quantitative studies. There are multiple strands that look into factors that affect learning outcomes. Common methods used are case studies and qualitative interviews of key stakeholders (usually principals and teachers). Findings are similar to that of the quantitative studies. Many studies focus on studying what high-performing schools do differently, particularly on school leadership and teaching practices. Another strand of qualitative studies focuses on teachers and principals' perceptions on assessments.

School leadership and teaching practices

Characteristics of principals in successful schools are similar to those identified in the quantitative studies. They seem to be similar across different context as well. Kondakci and Sivri (2014) used a multiple-case study method to analyse nine high-performing primary schools to identify common characteristics in Turkey. High-performing schools were identified based on results from the national examination. Seven common practices among the schools were identified. They are similar to those identified in the quantitative school effectiveness research. They are achievement orientation (especially on the national examination used to identify the schools), quality of instruction and classroom management, distributive leadership style, positive school climate (collaborative culture, trust, and satisfaction), having monitoring process for students' progress, good relationship with parents, and having adequate educational resources. This resonates with research by Peddell et al. (2020), who conducted interviews with principals from 14 schools that ranked top in the national tests in Australia. Principals were asked to describe their leadership practices. They similarly emphasised on creating a collaborative culture with a shared goal of achievement orientation. They are active leaders, leading in mentoring and classroom observation in addition to the administrative tasks. Additionally, they identified targeted professional development for teachers as critical in achieving their top rank status. Note that this does not demonstrate causal effect of good principal and learning outcomes, as good schools may be able to attract, motivate, and retain good principals better than average schools. Equally, appointing good principals to poor-performing schools may not bring the performance up.

Garza et al. (2014) conducted four case studies of leadership in schools that successfully maintain good learning outcomes over time in the US and Australia. Instead of the practices, they focus on traits and personalities of successful principals. Data was collected from document reviews and interviews with various stakeholders on what they perceive to contribute to the school's success. Inductive thematic analysis was used to analyse the data. All principals were experienced and show commitment to sustaining success in their respective schools. They have similar traits and characteristics such as resiliency, courage, and being ethical. Specific practices vary from schools to schools. Similar practices include investing in creation of collaborative school climate, building trust among teachers, and setting high expectations.

Baars et al. (2018) also focus on factors that characterise high-performing schools. They are more focused on schools that are able to raise learning outcomes of the disadvantaged pupils in England. High-performing schools are identified as schools that have low within-school achievement gap between advantaged and disadvantaged students. Instead of just looking at the high-performers, this research also compares practices of high- and low-performing schools, to identify unique practices of the high-performers. Using a qualitative case study method, they conducted in-depth interviews, focus groups, and observations. Again, they found school culture to be different in high-performing schools. There is an element of achievement orientation, but specific to the disadvantaged pupils. They hold high expectations and have greater conviction that the disadvantaged pupils will succeed. These thinking led to practices of supporting the disadvantaged pupils when they fall behind and giving more challenging work to students.

Based on classroom observations and teacher interviews, DeJaeghere et al. (2021) compared teaching practices between high- and low-performing classrooms in Vietnam. High- and low-performing schools were identified using students' school grades and achievement test results from the RISE project[60]. They found that teachers in high-performing classrooms engage more in teaching metacognitive knowledge (thinking about thinking). This is done by providing explicit instruction, encouraging students to think aloud and verbalise their thoughts, and using well-structured assessment and reflection. Teachers also engage in reflective questions that support students on higher-order thinking. By contrast, in classrooms with low learning outcomes, teachers engage less in these practices. The instructions were less clear and teachers focus more on delivering contents rather than on thinking process and knowledge application.

To conclude, qualitative studies here focus on identifying practices that are unique to high-performing schools. Some studies include only high-performing schools (based on pupil test scores) while some include both high- and low-performing schools to compare the practices. Findings show school culture and leadership relate to learning outcomes. Specific practices are among others, being achievement-oriented, supporting teacher collaboration, and having quality instruction. There are considerably fewer qualitative studies in this area than

[60] Research on Improving Systems of Education

quantitative ones and remain under-researched, especially in the context of developing countries.

Assessment

Another strand of qualitative studies looks at how teacher behaviours are affected by the pressure of assessment, which in turn, can affect learning outcomes. How teachers prepare for the assessments can affect test results, as well as what students learn.

There is evidence that teachers' views about assessments affect their behaviours and practice. Brown and Harris (2016) found that teachers' views on external examinations range from neutral to negative, depending on the context. Their concerns revolve around the usefulness of the test, quality of the test, and practice of teaching to the test. Some perceive that the tests did not provide an accurate or valid representation of students' skills and knowledge or reported mistrust of the external examinations. When the views are negative, the test results are used less in school. In contrast, when the tests are high-stakes, time and resources are allocated to test preparation, scoring, and result analysis, which could lead to teachers narrowing the curriculum and test-centred pedagogies. This is supported by McNamara (2010), who found teachers experience pressure to get their students to do well on the tests and spend more time preparing students for the test. Mons (2009) also found that high-stake tests can lead to a rote learning style of teaching and focusing attention on certain groups of students to ensure score improvement.

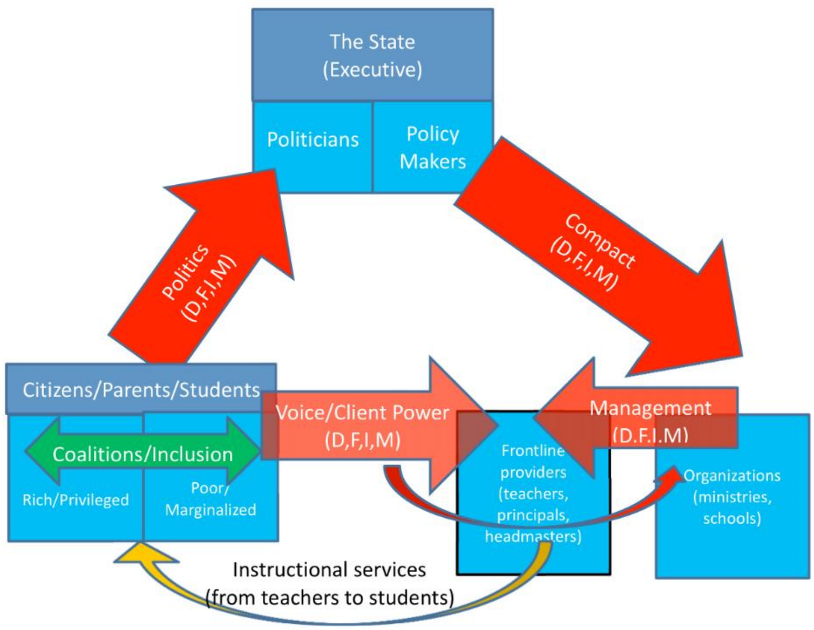
Brown and Harris (2016) also caution that when teachers are subjected to many assessments that are designed for different purposes, it is likely to create cognitive conflict and competition for resources. This draws parallel to the situation in Thailand, where there are many assessments with different skill requirements.

Accountability

Standardised tests were introduced along with greater accountability, an upward shift in powers from local authorities, and the move to quantitatively assess learning (Mons, 2009). In theory, the tests are implemented to improve students' performance through various mechanisms, including hard and soft accountability models. In the hard accountability model, the examinations are high stakes, and both the pupils and schools are expected to work harder to improve the outcomes. Teachers and principals can be held accountable for pupils' results, and hence have to work hard to improve outcomes. By contrast, softer accountability models propose that exam results can be used as feedback for the teaching staff to improve their practices (or formative feedback), instead of being published to the public. Mons (2009) asserted that in practice, the evidence does not support that the testing and accountability system help improve academic outcomes. Instead, it tended to lead to teachers narrowing the curriculum, teaching to the test, or focusing on some groups of students. Booher-Jennings (2005) shows that one school responded to the newly introduced accountability system in Texas by separating students into groups according to their performance and focusing resources on near-pass students or those who almost pass the required standards. There is also evidence of narrowing the curricula and less support provided to low-performing students (Anagnostopoulos, 2006).

System incoherence is another system-level factor that was argued to affect learning outcomes. Pritchett (2015) hypothesised that when elements of the accountability system are not aligned, the education system fails to produce desired outcomes, regardless of policies being implemented. This partially explains incoherence found in quantitative studies of what works in improving learning outcomes, where it was observed that some policies work well in some countries but not others. He argued that when the system is dysfunctional, sound policies such as improving teacher quality fail to bring the desired outcomes. As seen from Figure 3-1, policymakers and politicians influence educational institutions, which in turn influence practitioners or teachers. At the same time, parents influence teachers as well as policymakers. The relationship can be viewed as a principal-agent relationship, in which the principal appoints the agent to act on his behalf. The elements that contribute to the actions of each stakeholder include delegation (what the principal wants the agent to do), financing (money being paid to the agent up front and for completing the task), information (that the principal can monitor from the agent), and motivation (agreement of what the agent would receive if the outcomes are favourable).

Figure 3-1 Depiction of a system accountability triangle



Source: Pritchett (2015, p. 17)

When the elements are not aligned, the system fails to improve learning outcomes. Some of the examples are when there are not enough finance/resources available for the schools to accomplish the goals, when the outcomes are not properly measured, or when the incentives system does not involve desired outcomes. This resonates with many countries including Thailand, where the goal is shifted towards training students to have skills of the 21st century, but the country is still constrained by a rigid system. For instance, teachers are incentivised to perform well on the national test (OECD & UNESCO, 2016), but the test covers little of the 21st century skills.

3.2.2.3. Mixed methods studies

The previous two sub-sections explore studies with purely quantitative or qualitative methods. This section presents studies within the educational effectiveness research that are mixed methods. However, as noted by Hadfield and Chapman (2015), pure mixed methods studies remain rare. Many studies are instead multi-method with an emphasis on the quantitative components. Two studies are presented here are examples of high-quality research that combines quantitative and qualitative methods in the field of educational effectiveness research.

The first study is by Sammons et al. (2014), where the effect of school leadership on student outcomes is investigated, utilising both quantitative and qualitative methods of analysis. Firstly, value-added analysis was used with secondary data to identify effective schools and schools that made significant improvements in learning outcomes over three years. After the improving schools were identified, they were further grouped into typologies based on factors such as whether they started out as low-, medium-, or high-performing, and the percentage of students receiving free school meals. These schools were sent rounds of questionnaires to pinpoint changes in practice. This data was analysed using the simultaneous equations model (SEM) to model the relationship between school and principal characteristics and outcomes. Out of these schools, to be inclusive, 20 schools with different contexts were selected as case studies. Data was collected from both observations and interviews with key stakeholders about the school context and potential enabling factors for school improvement. Qualitative data were analysed using thematic analysis. Results show that one of the main areas of focus of all schools was changing school culture, and that leadership is seen as very important in bringing about improvement. There are differences between initially low-performing schools and medium to high-performing ones, with the former making more changes in school culture and teaching and learning. The quantitative findings are aligned with the quantitative survey findings, which show leadership affects school processes, although there is only a weak link to learning outcomes. With qualitative results, insights as to how the schools brought about those changes were discussed as well, which adds value compared to purely quantitative research, where such issues are rarely discussed.

Day et al. (2008) conducted a mixed methods study on teacher effectiveness. They remarked on the lack of purely mixed methods study in the field. Similar to Sammons et al. (2014), a large-scale quantitative survey was conducted to aid the selection of case studies of schools and teachers to be followed over time. Teachers and schools were selected to ensure a diverse sample (in terms of factors such as experience, age, school location, socioeconomic status, etc.) and to reflect a national population. After the selection, data were collected using semi-structured interviews and analysed based on the grounded theory approach. Teachers were asked about their perceived effectiveness over time and what they thought shaped their effectiveness. Interview questions and focus were adjusted based on the findings. The qualitative data was complemented with quantitative analysis of contextual teacher value-added using data on student achievement. The researchers categorised teachers based on their value-added and analysed their qualitative profiles to identify patterns of teachers with high and low effectiveness (both self-perceived and actual effectiveness based on value-added). Findings show teacher value-added varies over teachers' professional life phase, and teachers with more commitment and who anticipate an upward career trajectory are more effective. In the study, mixed methods is used from the design phase and both quantitative and qualitative data were integrated to assess teacher effectiveness.

To summarise, there is mixed evidence on which factors affect learning outcomes. Generally, physical resources show high effects in contexts where few resources are available. However, as the basic standards are reached, adding more resources does not help with learning outcomes. Teacher and classroom dynamics are shown to be important. Nonetheless, quantifiable teacher characteristics show mixed results, and many classroom variables only show small effect sizes. Qualitative teacher characteristics such as aspects of leadership and teaching strategies were found to be similar across schools that perform well in cognitive assessments. Some studies show that school effects can change over time. Additionally, results are likely to be context specific. What works in one context may not work in others. Other than the work by Glewwe et al. (2011), the majority of literature is conducted mainly in high-income countries. Hence, it is important to contrast the findings with the ones specifically in the Thai context, which will be explored next.

3.2.3. Empirical studies specific to Thailand

This section dives further into research in Thailand that explores the reasons behind low learning outcomes., which are most related to this thesis. The studies are primarily qualitative and seek to explain system-level factors that contribute to low learning outcomes.

3.2.3.1. Literature overview

Research points to different hypothesised sources of underperformance and different implications and suggestions for policymakers. The issue is multi-faceted and seems to have many explanations from the different studies. Both quantitative and qualitative studies are pointing to problems at the system level as well as what happens in schools and classrooms. Some are stressing on system level issues of political economy and implementation shortfalls, and others attributing to inequalities and inefficient resource allocation. All studies are very critical of the systemic factors that contribute to low learning outcomes. The majority of quantitative studies focus on single-year learning outcomes, with only two (Lathapipat & Sondergaard, 2015; World Bank, 2012a) attempting to comment on trends in the results.

Many empirical works seek to explain why the level of achievement is low in Thailand. There are some local studies, written in the Thai language and published in Thai journals, that explore student and school-level factors that are related to achievement (see e.g. Chansri & Kittichotipanit, 2014; Ratchapat, et al., 2012). However, the study quality may be lower than those published in international journals. For instance, critical procedures such as sample selection or which variables to include in the model are usually not discussed. They also did not focus on the context of Thailand or the communities they collected the data. Therefore, these studies were read in light of these weaknesses, and this section would focus instead on the results of research published in international journals/books or reports from organisations such as the World Bank and UNICEF.

3.2.3.2. Low learning levels

Many studies discuss learning outcomes as part of describing the quality of the Thai education. Inferences and judgements were made from the results of national and international cognitive assessments regarding the quality of students and the whole system. The benchmark used to evaluate the learning outcomes are both in absolute (number of students scoring above/below certain thresholds) and relative terms (compared with other countries). The prevailing narrative is that the learning outcomes are low. However, there is disagreement on whether the outcomes have improved over time, depending on the assessments used and the year of comparison.

Quality is one of the main rationales and goals of the 1999 system-wide reform. Indicators of learning outcomes were presented as evidence of Thailand's low-quality education. One of the international assessment results that were cited was TIMSS 1995, where Thailand ranked below many countries, including Singapore, Japan, Korea and Hong Kong (Atagi, 2002). Other than TIMSS, internally conducted assessments show low skills in literacy, mathematics, and science. As a reflection of the reform, ten years later, Fry and Bi (2013) noted that the rankings in TIMSS and PISA remain low and fail to improve over time (as of the 2009 round of testing). Additional measures of student outcomes were introduced to assess the success of the reform. O-NET scores were deemed 'disappointing' as the average is below 50% in all subjects. A report by World Bank (2012a) noted a slight improvement from 2006 – 2009 in PISA and speculated that this could be a turning point in Thai education.

Other, more recent studies share the consensus that the learning levels are not satisfactory. A study by OECD and UNESCO (2016) presented PISA results in comparison to other Asian countries. Parts of the results are concerning, with a large proportion of students performing below basic level (PISA baseline proficiency level – level 2) and large inequality among students going to schools in different regions and school sizes. However, they also note that Thailand performs better than Indonesia and Malaysia, which have similar income levels, and that the results seem to be showing improvement over time. Similarly, Lathapipat and Sondergaard (2015) noted the improvement from PISA 2003 to 2012 but raised a similar concern about the proportion of students performing below PISA functional literacy level (level 2). Fry (2018) presents a comprehensive overview of the education system, including all levels of education and nonformal education. One of the chapters describing the learning outcomes of Thai primary pupils is Shaeffer (2018). He drew on the national assessment of O-NET, PISA, TIMSS, and TOEFL[61] tests to assert that the outcomes are below average. For O-NET, the average scores were below 50%. For PISA and TIMSS, scores were below overall

[61] Test of English as a Foreign Language

average and low ranking was mentioned. Pongwat and Mounier (2010) similarly cited learning outcomes of PISA and O-NET to comment on quality. However, they also advocate that quality measurement should go beyond learning outcomes. Studies such as Atagi (2002) discuss qualitative indicators of low education quality such as teacher practice emphasising rote learning and memorisation and lack of quality teachers.

Learning outcomes have also been portrayed largely negatively in the media. Bangkok Post reported PISA 2015 scores as an indication of the failure of Thai education to improve (Mala, 2016). Most concerning was that PISA 2015 scores showed a significant drop from 2012. The scores were presented to be below the OECD average and many high-performing Asian countries such as Singapore. Other news reporting PISA 2015 results share a similar outlook such as Thairath (2018) and Mala (2018). One media reported a slight increase in grade 6 O-NET scores from 2017 to 2018 (Manager Online, 2019). Nonetheless, the average O-NET score remains below 50%.

The low learning levels are often contrasted against the favourable conditions of the Thai economy and Thai education. Nakornthap (2018) identified several paradoxes in the secondary level of education, namely low student-teacher ratio yet large average class size, girls completing schools in greater percentage than boys, highly qualified teacher force yet low test scores, high investment, and low contribution of private sector compared to global trends. Fry and Bi (2013) contrast the low learning levels with good and visionary educational leaders and policymakers, high level of spending on education, improvement in infrastructure, and presence of Thai students in international academic competitions such as the Science Olympiads.

To conclude, there seems to be an agreement in the literature that the learning outcomes of Thai pupils are not at a satisfactory level. Many students are performing below national and international standards. The ranking of Thailand in the international assessments is viewed as low. Nonetheless, Thailand’s performance is not significantly lower than Asian countries of similar income levels (except for Viet Nam). Different studies suggest differently whether the learning outcomes have improved over time. This depends on the tests and the year the study referenced. For example, as PISA 2012 shows an improvement, studies comparing trends up to 2012 such as Lathapipat and Sondergaard (2015) and OECD and UNESCO (2016) concluded that performance had improved. Studies using PISA 2009 (Fry & Bi, 2013) or from 2015 onwards (Shaeffer, 2018) instead shows stagnation/slight decline. Hence, it is crucial to observe the long-term trends and be aware that the starting point can affect the trends. Next, we turn to explore the rationales behind why the learning levels are the way they are.

3.2.3.3. System-level political economy hinders learning

Low learning outcomes can be attributed to systemic features of Thai education, from the top level of the ministry of education to classroom levels.

Thai culture and the concern with public image

It is important to step back and examine the nature of Thai culture and how this may have affected the stakeholders' perceptions and actions. One of the key features of Thai culture is its concern with public image. This led to actions to preserve and maintain the image, including having a strictly hierarchical society, favouring conformity, and avoiding losing face and discussing inappropriate topics (Jackson, 2004).

The Thai culture has significant concern over "public image" or appearance/reputation/face. Actions are contextualised within "time and place" and public/private domains of life. Additionally, there is an emphasis on social order and conformity to cultural norms. In public, it is favourable to respect social order and appear in agreement with the people of higher social status. People who do not conform to this are viewed as behaving inappropriately. This applied mainly to Thai people, rather than outsiders. There is a culture of public denial of facts that do not fit with the pleasant outward appearance (Jackson, 2004). An example of this is denying the existence of male prostitutes in Thailand. However, this does not mean they believe male prostitutes do not exist, but rather signalling that such issues should not be discussed in public.

At the core of what constitutes a good image is the power structure. A person's title or ranking signifies if he or she is the person of power. Bolotta's (2021) application of Jackson's theory emphasised the power play between big people (phu-yai) and small people (phu-noi). The former includes people in positions of power, which could be state officials, bureaucrats, military, etc. The latter ranks lower in power, in Bolotta's case, the migrant workers. In education, teachers and students could also be viewed as "small people" in relation to the principal, ESA officers, or the Ministry of Education.

The concern with face and strict power structure can have implications on how the education system functions and how improvements are made. For example, the inappropriateness of discussing problems that might upset others could be viewed as "sweeping things under the rug". This mentality in the context of Thai education could mean that the problems go undiscussed starting from the school level to the Ministry of Education. Implications of these cultural facets are discussed next.

System-level factors

At the top-most level, politics is what Mounier and Tangchuang (2010) focus on in their work. They argued that in the 1999 education reform, new neo-liberal ideas could be at odds with the desire to maintain social hierarchy. When the reform was introduced, Thailand's politics had been fairly stable and the political climate had been positive to changes, and the law was passed for the reform to happen. However, as new politicians come into power, not all wanted to further the reform

(Sangnapaboworn, 2018). Particularly, stakeholders were reluctant to carry out changes that they perceive as making them lose power. Mounier and Tangchuang (2010) contrast changes that do not change the power dynamics such as increasing years of compulsory education with changes that affect the power structure such as decentralisation and child-centred learning. They concluded that because these changes were not fully implemented, along with policy discontinuity, had contributed to Thailand's failure to improve quality and persistent low learning outcomes.

At a national level, curriculum and assessment were examined to see whether they had contributed to the low performance. An OECD and UNESCO team reviewed education policies and practices in Thailand (OECD & UNESCO, 2016). Drawing from both official documents and interviews with key personnel and stakeholders, they summarised challenges in the Thai education system. The curriculum (of 2008) was found to lack clarity. While claiming to be standard-based, there are no clear expected standards or progression of skills. Competencies were identified, but they did not explicitly correspond with learning outcomes, and there is no guidance for teachers on how to teach and assess competencies. Assessments conducted through national tests (particularly O-NET) have low validity and comparability. NIETS, the organisation responsible for designing and implementing these assessments in Thailand, seemed to have a low capacity to develop quality assessments, with only five psychometrics experts employed at the time the report was written. Additionally, there is no attempt to equate the scores over time. The study concluded that these key problems in the curriculum and assessment contribute to the stagnant quality of education. However, it is important to note that there had been some changes in a positive direction since the report had come out. IPST now is writing the Mathematics and Science curriculum and designing O-NET exams in those subjects since 2017 (Sangbuaphuen, 2020). So far, there has been no research formally analysing them.

Another theme coming out of the OECD and UNESCO report (2016) is related to the lack of support given to schools. Findings show that both pre- and in-service teacher training and support are inadequate in helping teachers and head teachers excel. Many administrative duties also reduce teachers' time inside classrooms. Additionally, ICT usage is low in classrooms because many schools lack the infrastructure and knowledge to successfully operate them. The policies may have good intentions, but the lack of support makes implementation fall short. Similarly, Fry and Bi (2013) and Pongwat and Mounier (2010) attribute the failure to internalise the reform to the practitioners. This is supported by Hallinger and Lee (2011), who asserted that only one-time training was given to teachers and principals. Less emphasis on making everyone on board and lack of support given in the educational reform makes the quality of education remains low in Thailand, despite the high level of expenditure allocated to improve the quality of education. Specifically, Pongwat and Mounier (2010) analysed that one of the reasons the 1999 reforms failed to materialise is that most teachers do not believe in the child-centred pedagogy presented in the reform. Hence, they continued to use a teacher-centred style of teaching[62].

Phuaphansawat (2021) utilises O-NET and PISA data to explain sources of inequalities in learning outcomes. She argued that wealth is the main driver of learning outcome inequality in Thailand. Descriptive statistics from both O-NET and PISA show large gaps between affluent areas (Bangkok) and the rest of the country, socioeconomic status (wealth quintile), and school types. Additionally, student background determines how well the school performs, rather than the school's geographical location. BMA schools, where students are from poorer backgrounds, perform worse despite being in the urban area of Bangkok. By contrast, Satit schools, which are selective and have better student backgrounds, perform better regardless of the province in which they are located. Gaps in learning outcomes are also from the lack of good quality schools in remote areas. Many good schools are in Bangkok, and many middle-class parents send students

[62] Note that teacher-centered style can be prevalent in many parts of Asian countries, and is not inherently related to poor performance.

to study in Bangkok. However, those who are poorer are left with no choice but to study at local schools, which usually do not have the same quality. Students are further segregated with the presence of special-purpose schools or elite schools such as Science and Satit schools, where learning experiences can be very different. Elite parents send children to international schools, and the middle class send their children to private or competitive public schools, whereas the poor are left with normal public schools.

These system-level factors do not only contribute directly to overall average low learning outcomes, but also indirectly in influencing practitioners' views and actions. It is important to keep this backdrop in mind when analysing practitioners' actions and practices. The next section explores how school leadership and school culture could affect poor learning outcomes.

Limited support to foster school leadership

Hallinger is one of the top researchers on the topic of school leadership in Thailand. He conducted various quantitative and qualitative studies to understand the nature of school leadership and school improvement in Thailand. He asserts that school leadership is of vital importance in improving educational processes and outcomes. To successfully implement a system-wide reform, school principals need to develop the capacity and the will to lead change (Hallinger, 2018). Hallinger's works concentrate largely on the notable education reform of 1999, where many system-level changes were attempted, including school-based management and child-centred pedagogy. From the start of the reform, he explored the sociocultural context of schools that may support or hinder change. For instance, in Hallinger and Kantamara (2000, 2001), the culture of Thai schools and leadership were analysed using Hofstede (1991)'s framework. Many cultural influences prevent school improvement. One example includes the concept of losing face or 'sia naa'. Sia naa refers to "the need to avoid making someone lose face or embarrassing them" (Hallinger & Kantamara, 2000, p. 50). School personnel's behaviours guided by this concept may undermine the reform in many ways as the authors illustrated below:

> *"… in a hierarchical system, persons of higher rank are supposed to know more and to perform better. But when they do not, others may go to great lengths to pretend that they do. The goal is to avoid having their superordinate* ***sia naa*** *(Hallinger & Kantamara, 2000, p. 50)."*

Another interesting cultural difference between Thailand and Western countries include the idea of responsibility. This is of importance since a significant portion of the 1999 reform includes school autonomy and principals taking responsibility for the students' outcomes. Hallinger and Kantamara (2001, p. 399) argue that, whereas responsibility brings 'recognition and respect' in the West, it is vice versa in Thailand. It is an individual's rank and status, rather than responsibility, that bring respect. By contrast, the responsibility that an individual has is merely a signal showing his/her rank. Therefore, school leaders are not encouraged to do more to improve outcomes or take responsibility for them.

During the reform, Hallinger focused on identifying factors that contributed to the successful implementation of the reform in Thailand through both quantitative and qualitative methods. Kantamara, Hallinger, and Jatiket (2006) analysed a case study involving a curriculum reform that had scaled up successfully. Taking both cultural and political circumstances of Thai society into account, the key success factors listed are unique to the Thai (or more broadly, Asian) context. Some of these include the involvement of the officials and politicians from the initial stages to ensure ongoing financial support, gaining support through publicity, and supporting principals throughout the programme. Again, the role of school leadership re-emerges as an important factor of change. However, the government may not have provided adequate support for school principals. Hallinger and Lee (2011) presented principals' perspectives of reform progress after a decade of implementation collected through a survey. The principals identified school leadership as a factor that strongly supports successful implementation. Yet, they felt the support they received was insufficient to lead change. Note that the regional training that Hallinger was involved in (and from which he gathered this data) is one of the very first nationwide support the ministry had given to the principals. The training was conducted in 2008 or nine years after the reform bill was passed. Bunnag (in Hallinger & Lee, 2011) further asserts that during Thai reforms in general, support was usually provided after there is evidence that the reform has failed rather than before. Gamage and Sooksomchitra (2004) similarly stressed the lack of support the schools were given to carry out the reform. This tendency was proved to have serious implications as many aspects of the reform still fail to materialise until today.

<u>Inefficiencies of the education system</u>

Another aspect stressed by many studies is the inefficiency of the Thai education system. Specifically, money put into education is not being spent optimally nor targeted to improve learning outcomes.

Lathapipat and Sondergaard (2015) identified lack of autonomy, underutilisation of information, and inefficiencies in spending as the main causes of low performance. The report focuses on small schools, which seem to perform the worst in the tests. In 2011, 47% of OBEC schools have less than 120 students. The number has been increasing due to lower birth rates. Simple descriptive statistics show a low student-teacher ratio (13.4 compared to 21.4 in other OBEC schools). Yet, this does not translate into better quality education as teacher shortage remains one of the biggest problems of these schools. The average teacher per class is less than one, meaning multi-grade teachings are common. Hence, it is the results of poor planning of the education system that is the source of the low performance.

The findings resonate with Fry and Bi (2013). Drawing from document analysis and participant observation, they similarly attribute the source of low education quality to the inefficiency of the system. They similarly point to a large number of small schools in the system as showing inefficiency. Not only that, but there are also too many higher education institutions as well as staff employed by the Ministry of Education. Funding was allocated to infrastructure building rather than supporting learning. The public expenditure tracking survey (PETS) collected by UNICEF (2017) also came to a similar conclusion. Using regression, it found a negative relationship between expenditure per student and O-NET scores because of inefficient resource allocation. Small schools in rural areas have significantly higher average per-head expenditure as the cost is spread out to a smaller number of students. Yet, despite higher spending in rural schools, teacher shortages in core subjects (such as mathematics and science) are more prominent in rural areas. Hence, gaps in human resources may be causing gaps in outcomes. In the same line of argument, Nakornthap (2018) asserts there are fewer resources allocated to the secondary level of education. This resulted in many underfunded schools such as the opportunity expansion schools (OBEC1) which tend to lack qualified teachers.

Studies here point to systemic and political economy factors as to why the reform had not worked and the learning outcomes remain low. These studies are primarily qualitative. To complement this, we turn to look at quantitative studies next.

3.2.3.4. Quantitative studies explaining achievement gap

Studies here investigate achievement gaps between groups of students, using data from the same year to dissect why one group performs better than other groups. This provides insights into why the high performers are the way they are, and how can learning outcomes be improved.

Lounkaew (2013) used Oaxaca-Blinder decomposition to decompose the PISA score gap between rural and urban students in Thailand. In addition to the standard decomposition method, unconditional quantile regression, developed by Firpo et al. (2009) was used, which allows marginal effects to be estimated for students at different score percentiles. This explains the gap beyond the mean, hence showing how the gap differs between high- and low-performing students. He found that a large portion (45-48%) of the urban-rural gap is unexplained. The gap is greater for the high-performing students (61-69%) compared to low-performing ones (12-15%). The unexplained portion is also higher for girls than boys. The large percentage of unexplained part may indicate that increasing physical inputs may not be enough to close the gap. As this study is primarily quantitative, it did not dive into what the unexplained part may involve but suggested areas such as school governance, teacher quality, parental involvement, etc.

Fry et al. (2018) conducted a mixed methods study to quantify and explain the disparity in education quality among the different provinces in Thailand. Quantitative method is the primary method of the study. Various indicators of education quality were combined to form an index of education quality by province. Some of the indicators used are learning outcomes of O-NET average scores and percentage of schools scoring above 50% on O-NET, and also other indicators such as average years of schooling, percentage of adults with college degrees, and percentage of students successful in gaining admission in prestigious universities. Descriptive results show a large disparity, with the top provinces being Bangkok and tourist hubs, while the bottom-performing provinces are in the Northeastern region and border provinces. Not surprisingly, the low-ranking provinces are also among the poorest. Of the variables included, the percentage of schools scoring above 50% on O-NET and admission to prestigious universities are the most unequal among the provinces and contributed the most to the gap. A qualitative method was used as a follow-up to the quantitative study. Educators in one of the most deprived areas were interviewed regarding factors that may contribute to inequality. Poverty, remoteness, and lack of resources are among of the recurring theme. In remote areas, the schools lack support both in terms of physical education resources and personnel (lack of personal development and support).

Other related Thai literature identifies specific factors and their relations to achievement. Pholphirul (2017) studied the long-term effect of pre-primary education on educational outcomes of 15-year-olds using PISA data. Using standard regression methods, pre-primary education enrolment is found to be significantly related to PISA scores in all three subjects. The effect is largest in low- and middle-income households. Other variables such as the father's education, school location, school resources, and whether a school is private also significantly predict achievement. Patrinos, Arcia, and Macdonald (2015) investigate how school autonomy and accountability affect PISA performance in Thailand. Students in the schools with more flexible principals[63] tend to do better on PISA, with a 6-8.6 points difference. Similarly, using PISA data from the earlier cycles, Lathapipat (2011) also found autonomy and accountability to be associated with high performance.

Despite the link between autonomy and accountability and learning outcomes, the evaluation shows mixed evidence of the level of autonomy and accountability in Thailand. SABER[64] survey (World Bank, 2012b) found that the country has a high level of autonomy and accountability in the participation of the school council in school finance, assessment of school and student performance, and school accountability. However, it has a lower level of autonomy and accountability in budget and personnel management. Salary is set at a central level, but Thai schools have autonomy over the management of non-salary expenditure and can raise funds from alternative sources. In personnel management, ESAs decide on hiring new teachers and schools themselves do not have autonomy over this.

PISA's context questionnaire also has a measure of school autonomy and accountability similar to SABER. They include additional aspects such as autonomy in students' admission to school and curricular autonomy (OECD, 2016b). Indices from PISA 2015 show Thai schools have higher than average

[63] Measured by the ability to go beyond the normal level of autonomy. For example, in the Thai education system, the decision to hire and fire teachers is not up to the principals, but rather the ESAs and the Ministry. However, some principals reported being able to lobby and select teachers.
[64] World Bank's Systems Approach for Better Education Results

(compared to OECD countries) autonomy in terms of school resources, curriculum, establishing students' assessment policies, and approving students for admission to the school (OECD, 2016b). This means that most decisions were made by principals and teachers, rather than the school board, and local or national authorities.

However, in practice, challenges may prevent teachers and principals to exercise full autonomy. In a joint report by OECD and UNESCO (2016), interviews with principals and teachers show that both suffer from workload. Teachers reported they cannot refuse non-teaching responsibilities being delegated to them (such as preparation for school assessments, overseeing school grounds, etc.). Similarly, principals said they spent most of time in administrative tasks rather than being involved with teaching and learning. This may hinders achievement levels.

3.2.3.5. Other reasons for low learning levels

Results from the national and international assessments have attracted attention from the media as well as the research communities. In the media, different rationales were offered as explanations for score improvement/decline. Bangkok Post reported PISA 2015 scores as an indication of the failure of Thai education to improve (Mala, 2016). An academic from a leading Thai university commented that the improvement in 2012 is from the inclusion of elite schools, and the decline is from those schools failing to participate in 2015. The Thairath newspaper (2018) recorded an interview with a former Thai minister of education, who claimed Thai students are not performing poorly, as students from elite schools perform as well as OECD countries. He also claimed that the Thai sample may not be representative as younger and disadvantaged students are included in greater proportion than actual enrolment. By contrast, the OBEC secretary attributed increase in O-NET scores as the success of policies of reducing teachers' administrative tasks and changing teaching style (Manager Online, 2019). These sources provide different views and implications for the Thai education. Hence, it is important to fact-check and reconcile these explanations to correctly portray the system health of Thai education.

Section 3.2 provides an overview of the first research strand of factors that affect pupil outcomes, for both international literature and literature specific to Thailand. Next, in section 3.3, we turn to look at research in the second strand, which covers discrepancies of assessment results.

3.3. Discrepancies in assessment results

This strand of literature is situated in education literature but can span across disciplines such as education statistics, psychometrics, and sociology. It relates to the second and third research question of “In what ways are the trends from PISA/TIMSS/O-NET similar and different?” and “What are the factors contributing to the discrepancies of trends over time?” respectively.

Research attempted to disentangle the sources of the discrepancies can be separated into three categories. The first (section 3.3.1) compares results of different countries that participate in the same assessment and investigates why some countries perform better or worse than other countries. The second (section 3.3.2) focuses on countries that participate in more than one assessment. They provide explanations for why some countries perform well in one assessment but not the other. The third (section 3.3.3), and closely related to the thesis, is research in discrepancies of trends over time within the same country. They attempt to pinpoint what changes over time had contributed to the different trends. Lastly, section 3.3.4 presents studies that focus on Thailand.

All research here situates in a broader theme of limitations in using and interpreting large-scale international assessments. They raise doubt on whether the learning outcomes could be trusted to reflect true ability of students, or whether they are products of biased design and implementation. When results are not reliable, it makes the assessments less useful for policymakers and practitioners.

3.3.1. Discrepancies among different countries

Research here focuses on explaining why some countries perform better or worse than other countries at one point in time. This strand is closely related to the education production function literature presented previously and points out key differences of students from high- and low-performing countries. Additionally, they pinpoint factors related to test design and implementation, which can raise doubt in comparing results among countries.

Research has shown that student characteristics matter in learning outcomes, albeit mostly in explaining within-country gaps. PISA results consistently show that students of higher socioeconomic status performs better than those from poorer background (OECD, 2016a). Studies have found that the differences in student characteristics can drive learning outcomes. Carnoy and Rothstein (2013) compares scores of PISA 2009 between the US and top-scoring countries and similar countries in terms of economics. Social class distribution (proxied using number of books in the home) has been presented as one of the explanations of the low scores. Looking at the distribution, the low average PISA scores of the US could be partially attributed to the composition effect. That is, there are more disadvantaged schools in the US sample compared to the actual population statistics as well as other similar post-industrial countries. As the pupils of lower social class tend to score lower than their upper-class peers, the country average became lower. When the proportion of disadvantaged students are adjusted to be the same in all countries, the US improved in ranking. This is supported by a study in the UK context (Gill & Benton, 2013). They found that student characteristics in other countries significantly differ from the UK. When these are controlled for, using multilevel modelling, the ranking of countries changed significantly, with the UK performing better than before the adjustment. Aloisi and Tymms (2017) also found that socioeconomic status increase is associated with an increase in PISA scores. Hence, the countries' rankings and performance are largely driven by student background.

With student background being highly predictive of scores, some studies investigate whether sampling in international tests are representative of the countries' population or not, and how that may have affected the scores. Hanushek and Woessmann (2011) argued that sample selection can have effects on scores of international assessments. Using PISA and TIMSS during 1995 – 2003, both simple correlations and regressions controlling for student background show that non-response rate and sample exclusion are related positively to the scores, with varying extent by year. This means that countries that exclude more schools and have less response rates at student-level perform better than countries that include more schools and students.

They also note that in developing countries that have not achieved universal primary and secondary education, students that participate in school-based assessments like PISA or TIMSS are likely to be more well-off than 15-year-olds in the country (Hanushek & Woessmann, 2011). This rang true in the case of Vietnam and Turkey. Glewwe et al. (2020) found that student background also contributed to Vietnam's high performance. They combined household surveys with PISA outcomes to adjust Vietnam's PISA sample to be more representative of its 15-year-olds population. They seek to investigate why Vietnam had exceeded all expectations and scored impressively high relative to its income level. They found that PISA sampled relatively well-off students in comparison to the household surveys and this could partially explain why Vietnam attained higher scores. However, even after adjusting down the socioeconomic status to make PISA sample more representative, Vietnam remains an outlier in terms of scores. Regressions and Oaxaca-Blinder decomposition are then performed to identify specific factors that could explain the high performance. The effectiveness of the dummy variable 'being in grade 10' is found to explain most of the gap, which may imply selectivity of passing student from grade nine to ten in the Vietnamese system. This resonates with the case of Thailand, as there is a mix of lower and upper secondary students participating in PISA as well. Similarly, Spaull (2017) showed that for Turkey, students who participate in PISA are of much higher socioeconomic background than normal 15-year-olds in the country who may be out-of-school or in-school but overaged and are not included in the PISA sample. They have also shown how this affect the interpretation of progress the country had made over time (see section 3.3.3).

Some studies found fewer clear-cut answers of why some countries perform better than others. Perera and Asadullah (2019) attempted to explain the underperformance of Malaysia in PISA. They compared Malaysia to Singapore and South Korea, the higher-scoring counterparts. Using Oaxaca-Blinder decomposition, a significant portion of the differences in score lies in the unexplained part, or the effects that unknown contributing variables may have on the outcome. This means that even though Malaysia has similarly advantaged student background as those high-performing countries, it still would not achieve the same scores, and more research is needed to find out why the country is underperforming.

Goldstein (2017) cautions that translation may affect scores and hinder comparability of test scores among countries. Even with translation protocols, the resulting test items may be more difficult or easier in some languages than others due to cultural differences or word choice used in the translation. El Masri et al. (2016) compared Science items of PISA 2006 in three different languages: English, French, and Arabic. They found that language idiosyncrasies could make items in different languages have different levels of difficulties. For example, specific words may be more common in one language more than others and using those words could make the question potentially easier for students who speak the language. Ercikan and Koh (2005) show that the English and French version of TIMSS are quite different. There is evidence of differential item functioning analysis, which makes some items easier or more difficult for different groups of students, in this case, for the English and French speakers. This limits cross-country analysis as the test features can make some countries perform better than others.

3.3.2. Discrepancies among different tests

Instead of analysing one assessment result in isolation, there are studies comparing the results of a country or a group of countries that participate in more than one national or international assessments. It is acknowledged that each test has its own focus and characteristics such as sampling and contents tested. Hence, the results should be compared to form a comprehensive picture of how achievement results are like in a particular country/a group of countries. In comparing across groups of countries, the studies investigate why some countries perform well in one test but not others or why do some countries rank high in one test but low in another.

Wu (2010) compared results of TIMSS and PISA 2003 for 22 countries. The correlation of the country mean scores is 0.84, showing a certain degree of alignment. When the scores are standardised and compared between PISA and TIMSS, some countries perform better in TIMSS than PISA (most Asian and Eastern European countries) and vice versa (Western European countries). The differences in this could be partially explained by accounting for the systemic design of the two tests, such as age of students in the time of the test. Particularly, countries performing better in TIMSS are likely to be older in the time of the test, and countries performing better in PISA are likely to be in grade 10 rather than grade 9. Additionally, since TIMSS and PISA have different percentage coverage of each content domains, countries performing well in dominant domains of TIMSS tends to score higher in TIMSS than PISA and vice versa. Carnoy and Rothstein (2013) reached a similar finding that since US pupils perform poorly on algebra items, the relative weight given to this topic in the assessment affects the relative rank of the US compared to other countries.

G. Brown et al. (2007) compares the results of PISA, TIMSS, PIRLS, and IALS[65] by calculating the average ranking in terms of median and dispersion[66]. Generally, most countries show agreement in different assessments. There is also a clear pattern of higher performing countries having lower dispersion in scores. When comparing whether the results were robust to the item-response models used (between one and three-parameter models), they found that while the median is robust to changes, the dispersion is not. Some countries show much larger dispersion when guessing is taken into account in the three-parameter model, especially lower income countries such as South Africa.

3.3.3. Discrepancies in trends over time within the same countries

Studies presented in this section explore potential sources of change over time in assessment results, ranging from sampling to policy changes.

Sampling has been found to affect score discrepancies when comparing across different countries. When sampling change over time, it can affect the trends as well. Spaull (2017) explored sampling as a factor that could affect Turkey's high rate of improvement during 2003-2012 PISA cycles. Specifically, proportion of eligible students (age-appropriate in-school students) that appear in PISA's sampling frame changed over time, making the results not representative of Turkey's 15-year-olds population. As Turkey had large proportion of out-of-school children and overaged students in 2003, Turkey's scores were significantly overestimated. Demographic and Health Survey data (DHS) was used to derive proportion of PISA-eligible students in comparison to total 15-year-olds and calculated how the scores would have been if these students were included in the sampling. Assuming out-of-school children would not have reached proficiency level 2 in PISA, reported PISA results greatly underestimated the increase in

[65] International Adult Literacy Survey

[66] The dispersion, or the spread, is calculated as (score at the 95th percentile – score at the 5th percentile).

performance over the year for Turkey. This is because during the period of interest, proportion of enrolled students increased significantly. If the ineligible students were accounted for, the score increase would have been much larger (twice for Mathematics and five times larger for Reading), meaning the progress made in education in Turkey is much higher than what PISA estimated. Additionally, inequity has not reduced over time. When socioeconomic status is adjusted for, the true poorest population performs much worse than is reported (i.e., poorest 20% of PISA sample is more affluent than poorest 20% of actual 15-year-olds). This is because the poorest population is not in school and is not tested. Hence, PISA results cannot be said to be representative of the whole 15-year-olds population, in-school and out-of-school. This means scores of international assessments cannot be taken at face value, especially in the context of countries where there are many out-of-school children or overaged students. Furthermore, as representation changed over time, comparing the scores directly may be misleading. Thailand is one of the countries experiencing changes in percentage of eligible students over time, from 54% in 2000 to 71% in 2015 (Spaull, 2017). Hence, due to low coverage rates, the true performance of Thailand's 15-year-olds may have been even lower than the reported scores in 2000.

Carnoy and Rothstein (2013) found that change in social class composition contribute to score change. They looked at trends over time of PISA from 2000 – 2009 focusing on the US By weighting the 2009 scores with social class proportion of 2000, they were able to separate out score change that is from change in group composition and change from educational improvements. Results show that the overall score during the period shows a decline despite educational improvements because there are more lower social class students included in the 2009 sample than in the 2000 sample. A similar study set in Russia (Carnoy et al., 2013) shows that the trends look different when disaggregating by social class group, with the decline coming largely from the advantaged pupils.

In some cases, change in sampling is not sufficient in explaining why scores changed over time. Morsy et al. (2018) investigates why PISA score of Australia fell over time during 2000 – 2015, yet did not arrive at a concrete explanation why this is the case. They interviewed experts and policymakers for plausible explanation of the decline in scores, then test the hypotheses empirically using PISA microdata as well as comparing the scores with TIMSS during the same period. Some of the assertions from the experts were not supported by the data, such as the movement of students from public to private schools lower the quality of public schools. Disaggregating data by socioeconomic status (proxied using books at home) shows that the decline in score is largest in the lowest class of students and in private schools primarily. School type and state differences were also investigated using OLS regression and descriptive statistics. However, both the coefficients and composition did not change substantially over time. Similarly, teaching quality (proxied using salary, class size, pre-service training, etc.) does not seem to change substantially over time.

Social development can also explain the increase in scores. Rowley et al. (2019) identified countries with improved performance on PISA from 2006 – 2012, as well as finding what they had done differently. Significance increase in scores was defined as showing at least 1/3 standard deviation increase. Out of 55 countries, only 10 showed significant improvement, with Qatar showing the largest increase of 58 points. They found countries with improvement are likely to have lower scores in 2006. This might suggest a floor effect, and that it is easier for low-performing countries to improve scores than countries that are already performing well. The improvement is from both high and low performing students, meaning the increase was equitable. Additionally, all countries with improvement had move to a more democratic form of governance or moved to a higher income classification during the period of interest. Indicators such as infant mortality and life expectancy had also improved for those countries. This suggests that educational and social development go hand in hand.

Additionally, Barrera-Osorio et al. (2011) used the Oaxaca-Blinder decomposition to explain increase in PISA scores of Indonesian students during 2003 – 2006. Almost all of change falls into the ‘unexplained’ category, which means that the improvement in student characteristics is not sufficient in explaining the increase, and that there had been improvement in education quality that was not captured in the observable variables. In an attempt to explain where the improvement comes from, the study used regression to identify factors related to student outcomes, as well as looking at changes over time related to teachers and schools. They found many favourable changes: adequate supply of teachers increased significantly during the period, there are less repeaters, and parents of students in the sample are more likely to have attended primary and secondary education (in relation to the parents having had no education).

Instead of focusing on single assessment, some studies investigated why different assessments show different trends over time. Factors such as changes in test procedures can cause discrepancies in scores. Jerrim (2013) investigated multiple sources of England's inconsistent results in PISA and TIMSS; whereas PISA performance declined, TIMSS had improved. To enable comparison, a pool of countries which took both exams during 2000-2009 were selected, and country-level scores were transformed into z-scores separately for each year. Even with this selected set of countries, the results still show conflicting changes over time from the two tests. Some factors that could explain the discrepancy include changes in sampling procedure, the inclusion of Welsh students, changing from age to grade-based samples, changes in testing period, and non-response. Adjusting for these issues gives different trends over time. This raises concern over the use of one test result to inform public policy.

Other than identifying what contributes to score change within the US, Carnoy and Rothstein (2013) also compared trends of TIMSS (1999 – 2007) and PISA (2000 – 2009) in the US. The US shows slight discrepancies in the trends over time between the two assessments (TIMSS increased, and PISA decreased by less than 10 points). The authors could not point concretely what may be the causes but note that the social class composition over time worsens for TIMSS sample in the US too. The increase in students of lower social class is greater than in PISA or NAEP (the National Assessment of Educational Progress), which may imply that TIMSS sampled more disadvantaged students than the other two tests.

3.3.4. Studies specific to Thailand

Two studies attempted to explain score change over time in Thailand. They similarly use descriptive statistics as well as Oaxaca-Blinder decomposition to dissect overall change into change from school and student characteristics and change that are unexplained.

A study commissioned by World Bank (2012a) attempts to explain the slight improvement in PISA Reading results between 2006 and 2009 and the decline in scores between 2000 and 2006 in Thailand using various quantitative methods, from simple descriptive analysis, to regression and Oaxaca-Blinder decomposition. As well as comparing results in Thailand over time, it also makes comparison with other Asian countries they termed ‘regional competitors’. The study dives into possible reasons for the improvement, assess the quality of education, and analyse implications for the Thai education going forward.

The study started by exploring descriptive statistics of the two years of interest. PISA Reading score increased by 4.62 points from 2006 to 2009. The authors recognised that this is a small improvement and not statistically significant. Nonetheless, it is a break from a declining trend from the years 2000 to 2006. From the descriptive analysis, they found more students of 2009 are in higher proficiency levels than 2006 cohorts. Additionally, they are of slightly higher socioeconomic status (more enrolled in urban schools and have higher parental education). Nonetheless, Thailand still scores well below other Asian countries such as Japan, Korea, and Singapore.

To pinpoint further where the improvement is from, Oaxaca-Blinder decomposition (under Education Production framework) was employed to analyse PISA reading scores of the years 2006 and 2009. Change in outcome was separated into explained (change due to improvement in student background) and unexplained part (change due to system improvement). Results of the decomposition show that out of the 4.6 points increase, 2.9 points are due to better student background while 1.7 points are unexplained. The authors used this to assert that system quality did not decrease during the time period.

Next, regressions were used to investigate where these changes are largely driven from. Three variables of focus were rural/urban, wealth, and gender. Regressions were run with data from both years. Interaction term was added to account for change over time (time x variable of interest). Pupils of both rural and urban area show slight insignificant change and there is no evidence that the gap between rural and urban students reduce over time. However, significant change was found for students of poorest 50% and wealthiest 50% (based on PISA's ESCS index) and gender. There is significant decrease in rich-poor and female-male gap[67]. However, the reduction in the gap is mainly driven from higher performers (rich students and female students) scoring worse and lower performers scoring slightly better. The change was about 0.1 standard deviation.

Other than focusing on just the years 2006 and 2009, the research conducted similar analysis comparing results of PISA reading in 2000 and 2006, which show a decline of 16 points. Trends in TIMSS and O-NET were also presented. The trends of O-NET Thai and English language subject and TIMSS Science show similar decline[68]. Descriptively, the results differ from comparing 2006 and 2009. The decline in scores is almost exclusively from poor students (having below median wealth index). There is increase in the gap between the high- and low-performers, with the former performing better and the latter worse. The gap remains the same for big and small schools. For public/private school, the gap reduced over time as public schools (the higher performer) showed greater decrease than private schools. Similarly, the gap widens for girls/boys as the boys, who are the lower performers, scored much less over time than girls.

[67] Female students perform consistently better than male students in PISA.
[68] Other subjects in O-NET did not show a clear decline

Results from Oaxaca-Blinder decomposition shows increase of 6.3 points explained by better student characteristics and large decline of 22.4 points unexplained (due to change in system quality). In 2006 – 2009, the unexplained part has not been this prominent. Diving deeper into the sources of the gap, the decomposition was conducted to see how much of the rural/urban gap can be explained by student characteristics. Here, there are more unexplained part as well. The authors point to the fact that the distribution of scores is very different among students in Bangkok and elsewhere, with Bangkok students having similar score distribution to that of the US (in both levels and dispersion).

School factors were also explored using descriptive analysis and simple regression to test their association with learning. They found practices differ markedly among different school types. For instance, ability grouping is much more common in OBEC1 schools (low-performing) than in Satit schools (high-performing). Simple regression shows some facet of the Thai education such as being in a private school is associated with lower learning outcomes than public schools. However, these regressions are simple comparisons, and the authors did not disclose if or which control variables were used. Results show factors such as availability of educational resources affect the decline in scores (with increase in resources for only bigger schools), lack of autonomy, and lack of out-of-school learning time.

Another study is Lathapipat and Sondergaard (2015), who use Oaxaca-Blinder decomposition to separate changes in reading score from 2003-2012 into explained and unexplained part. The samples were further separated using school location (from village to big city). Including economic, social, and cultural status or ESCS index, gender, and grade level in the model, they decompose changes over time into changes from student background and unobserved system quality (unexplained part). They found both average student background and system effectiveness had increased over time and that a large portion of the changes is unexplained. The results show that unobserved system quality actually increased during the time period. However, it is questionable whether 2012 is an appropriate year to use as a base for comparison since it may be an outlier[69]. Moreover, ESCS index was used over time without adjustments which was advised against by the OECD, since some procedures and scaling had changed (OECD, 2017a). Overall, the authors concluded that despite the small improvement in scores, the Thai education remains low in quality and equity. Policymakers should target to improve small or remote schools and focus on problems such as resources shortage, including personnel and material resources.

Despite being different in their time frame of interest and trends[70], the two studies have higher proportion of score changes that are unexplained. This means that accounting for change in student characteristics alone cannot explain the score change. The unexplained part, termed as changes in the quality of the education system by both studies, account for more than half of the change. In the World Bank (2012a) study, the unexplained part offsets the increase in students' background. The decline seems to be from those more disadvantaged, driving up the gaps in learning outcomes.

To conclude, section 3.3 presented research regarding discrepancies of assessment results. These aspects of the assessments raise doubt in the usefulness of the assessments to infer how well students are doing cognitively, as compared to other countries, and how that changed over time. It is important that policymakers keep this in mind when inferring from the assessment results.

---

[69] See

Figure 1-1

[70] Lathapipat and Sondergaard (2015) showing increase and World Bank (2012a) showing decrease.

Chapter Four: Overall Methodology

4.1. Introduction

To answer the research questions outlined in Chapter One, mixed methods approach is used. This chapter dives into more detail of the methodological approach taken, and how they are used to answer these research questions. The chapter is divided into three main sections. Firstly, the overall mixed methods is explained in section 4.2. Then, quantitative and qualitative strategies are presented in sections 4.3 and 4.4., along with rationale for the empirical strategy and data used.

4.2. Mixed methods

4.2.1. Epistemology

The philosophical stances underpinning the research can influence design and methodology. Creswell and Creswell (2018) identified four major philosophical worldviews (sets of beliefs) in conducting research: post-positivist, constructivist, transformative, and pragmatic paradigms.

Similar to the traditional positivist paradigm, post-positivism is associated with the scientific method. It views knowledge as objective, which can be empirically observed and measured. Post-positivists are less extreme than positivists in a sense that it acknowledges that researchers’ backgrounds and values can influence what is observed. Post-positivists aim to find causal effects among inputs and outputs. This implicitly assumes that inputs and outputs can be measured. With this, it is most often associated with quantitative methods.

By contrast, constructivists believe knowledge is subjective as individuals develop meaning in the world they live in. There can be multiple perspectives to the same phenomenon because experience is subjective (Della Porta & Keating, 2008). This contrasts to an objective view of the world by the positivist. The transformative worldview takes the constructivist view further and focuses on amplifying the voices of the marginalised groups of the society. Both constructivist and transformative worldviews are associated with qualitative research, which emphasises on how participants derive meaning from a phenomenon.

In the 1980s, quantitative and qualitative methods are viewed as incompatible in the 'Incompatibility thesis' (Tashakkori & Teddlie, 2010). However, this had been largely disproved as researchers successfully conducted research that 'mixes' the two methods together. In mixed methods, researchers can use both quantitative and qualitative methods to answer research questions, despite the difference in philosophical stances (postpositivist VS constructivist). Underpinning this is the pragmatic paradigm, which emphasises more on application of methods to solve research problems. Pragmatism puts research questions at the core. Methods and analyses are chosen based on how appropriate they are in answering the research questions (Savin-Baden & Major, 2012). By focusing on this, both quantitative and qualitative methods can be used in the same research. Therefore, this research takes a pragmatic worldview. It focuses on using a variety of methods to understand the research questions and research problems.

4.2.2. Rationale for using mixed methods

Formally, a mixed methods research is defined as a research in which,

> *"a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad*

> *purposes of breadth and depth of understanding and corroboration (Johnson et al., 2007, p. 123)".*

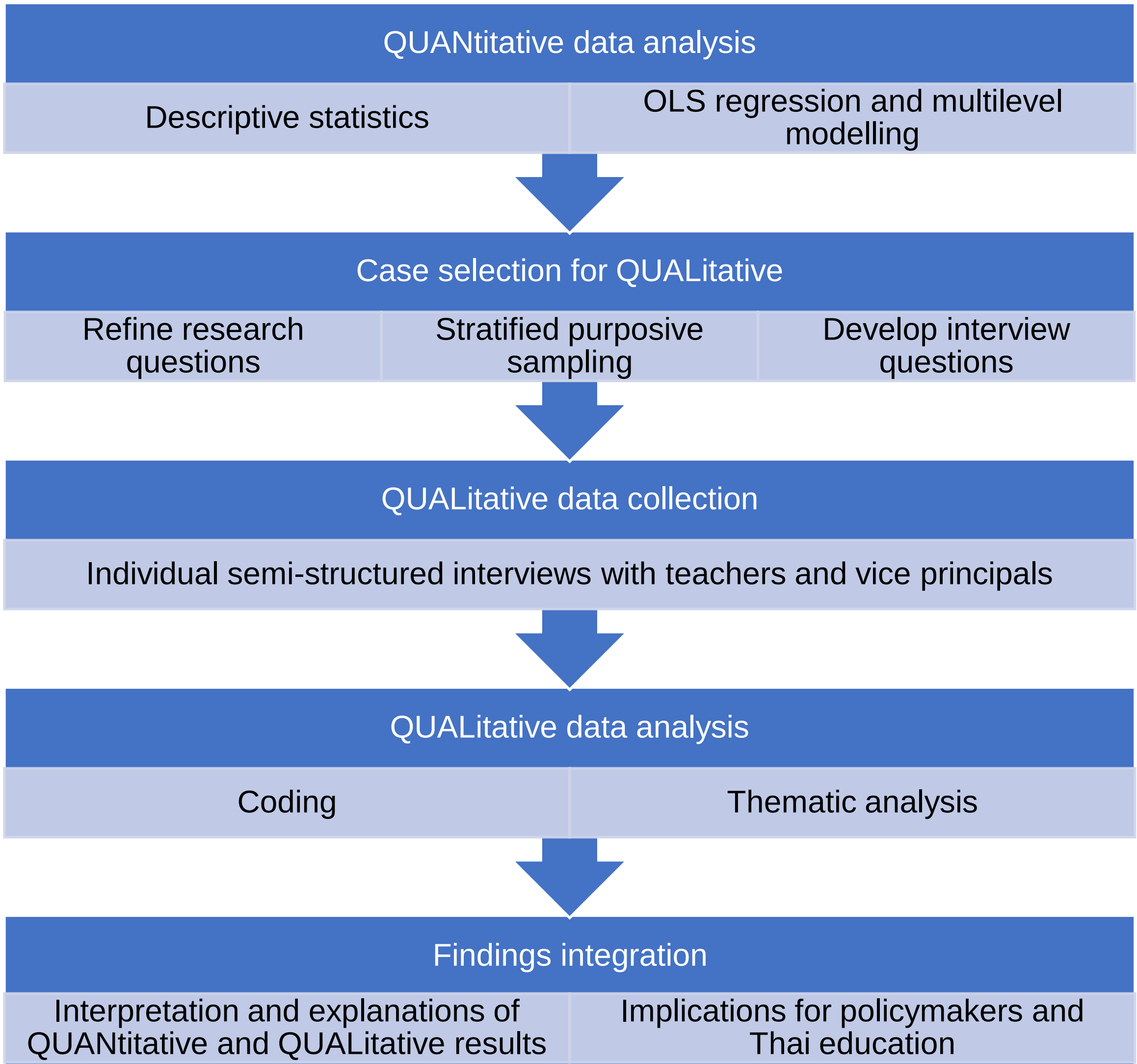
Researchers have used mixed methods for various purposes as summarised by Greene et al. (1989). Purposes that resonate most with the current research include 'complementarity' and 'development'. Complementarity is when greater, elaborated understanding of the issue is achieved when different methods are used to inquire different aspects of the research question, whereas development is when results from one method is used to inform development or design for the second method. As noted by Tashakkori and Teddlie (2010), mixed methods is better than purely quantitative or qualitative methods if they are more well-suited to answer the research questions and provide stronger inferences. Mixed methods studies can have both explanatory and confirmatory aspect. For example, a quantitative method can be used to confirm a hypothesis, and qualitative method to explore further the process involved. This offers richer inferences that cannot be acquired with one method alone. With this, the research can offer a more complete explanation of human behaviour and specific phenomena. Additionally, combining methods allows researchers to utilise strengths of both quantitative and qualitative research. Quantitative method can give greater breadth, while qualitative method can bring in depth into answering the research questions. Together, mixed methods is more equipped in answering multifaceted issues (Greene et al., 1989).

This study's aims are to identify determinants of learning outcomes and disentangle sources of discrepancies found in different trends of national and international assessments. Both conflicting trends and low learning level are multifaceted issues that encompasses many aspects such as the assessment system, teaching practices, policies, etc. Hence, it is more suited to be answered with mixed methods. Quantitative methods can be used to establish the extent and nature of the learning outcomes, as well as pinpoint factors that affect variations in scores and how that change over time. Meanwhile, qualitative methods can be used to explore in-depth the reasons learning outcomes are low and what the potential remedies could be. The two methods complement each other as asserted by Greene et al. (1989). Utilising both quantitative and qualitative methods would offer breadth and depth that one method alone cannot offer. Potential factors of learning outcomes that were not captured in the quantitative survey can be explored in the qualitative phase of the study. At the same time, quantitative phase can offer a macro nationally representative picture lacking in qualitative sample. Therefore, the research opted for mixed methods.

4.2.3. Study design

The research adopts a mixed-method sequential explanatory design. This means that the research is conducted in two phases, starting with quantitative research, then supplemented with qualitative research (Muijs & Brookman, 2015). This design is most useful when researchers attempt to understand not only general trends and relationships in quantitative data but also the underlying causes or mechanisms of the phenomenon (Creswell & Plano Clark, 2011). The overview of steps in conducting the research is in Figure 4-1.

Figure 4-1 Steps in implementing the mixed methods research



The study begins with quantitative data analysis using nationally representative secondary data. The three main research questions[71] are intended to be answered partially quantitatively, using both descriptive statistics (all research questions) and regressions (research questions 1 and 3).

[71] RQ1: What are the factors that could explain variations in scores?, RQ2: In what ways are the trends from PISA/TIMSS/O-NET similar and different?, and RQ3: What are the factors contributing to the discrepancies of trends over time?.

Next, case studies are used as a follow-up from the quantitative analysis to answer research question 1. An extra step of case selection is taken before qualitative data collection. The quantitative results are used to identify the participants[72]. Aspects of the quantitative results are also selected to be further explored in the qualitative phase of the study. The sub- research questions are then adjusted based on the findings[73],and interview questions are tailored to the aspects of focus.

Then, after the design of the qualitative phase is finished, I proceed with qualitative data collection, using individual semi-structured interviews. Due to the COVID pandemic, the data collection is done both online and face-to-face. After the collection, data is coded and analysed using thematic and contribution analysis. Supplementing quantitative results with qualitative findings allows comprehensive understanding of Thailand's underperformance in addition to purely quantitative methods.

After the results of both quantitative and qualitative phase are obtained, further analysis is conducted to integrate the findings of both phases. The analyses draw on results from both methods, providing both macro and micro level accounts of the problem. Firstly, the results are separated and presented according to the research questions. While Chapter Six and Seven (answering research questions 2 and 3) is purely quantitative, Chapters Five (answering research questions 1) present both quantitative and qualitative findings. Next, in each chapter, there is a discussion section that summarises how the findings from the two phases relate, how they are similar and different, and finally how a coherent narrative could be formed.

Next, the subsequent sections describe in detail the quantitative (section 4.3) and qualitative (section 4.4) methods used in the study.

---

[72] In this case, school types, one of the main predictors of test scores, were used as a stratification variable.

[73] For example, a question exploring the differences between high- and low-performing schools was included in the sub-research question, and questions about the skills that teachers focus were included in the interview questions.

4.3. Quantitative methods

This section presents the study's quantitative methods, starting from rationale for the methodology used (section 4.3.1), then data (section 4.3.2), methodological decisions made (section 4.3.3), and finally the empirical strategy (section 4.3.4).

4.3.1. Rationale for the methodology

The first sub-research question, "What are the factors that could explain variations in scores?", can be answered using both descriptive analyses and regression. Descriptive analyses can provide initial understanding regarding score gap. Additionally, to identify factors affecting scores, many literature[74] uses varieties of regression analyses. Regressions can be used to control for factors that might affect achievement. These can be done using both multilevel modelling and OLS regression. The model is estimated for all datasets – PISA, TIMSS, and O-NET. Results can be compared among the different assessments to see which factors affect variation in scores. As different data is available in different assessment (see further in section 4.3.2), the regressions can provide complementary analyses. Results from this chapter help identify factors that relate to achievement and whether the score gaps among subgroups decreased when factors such as socioeconomic status has been controlled for. Knowing this will enable policymakers to design policies that target the sources of the gap.

The second question, "In what ways are the trends from PISA/TIMSS/O-NET similar and different?", is descriptive in nature. The question aims to set a benchmark for the last question by setting out how the trends are like, both overall and within different subgroups. Hence, descriptive statistics is employed as the sole method in this chapter. Scores and other relevant statistics are calculated for groups of interests to see variations by groups such as school types, regions,

[74] See section 3.2 discussing what affects pupil learning outcomes.

socioeconomic groups, etc. This chapter helps establish the foundation of this thesis by answering how learning outcomes had changed over time, where the high- and low-performing schools are, and whether inequality has changed over time. The results of this chapter can aid policymakers in making decisions on resource allocation, as well as assessing changes in system quality over time.

The third question, “What are the factors contributing to the discrepancies of trends over time?”, can be answered with both descriptive analyses and regressions. Change over time of sample characteristics can be analysed from descriptive statistics. In regression analysis, time is used as an independent variable to investigate whether there are changes in the coefficient of variables over time or not. The analyses help identify what changes over time contribute to score change and will allow policymakers to understand better why the score trends differ among the different assessments.

For the first research question, the results can partially answer the research questions as the factors identified are limited to variables collected in the survey. To complement this, qualitative analyses are used to fill in the gap by exploring factors not captured, such as content focus, motivation, etc. and changes over time (further discussed in section 4.4).

4.3.2. Quantitative data[75]

4.3.2.1. PISA and TIMSS data

For descriptive statistics and modelling, PISA data in focus is anonymised student-level data from 2012 and 2015[76]. The test scores of each student are provided in plausible values (PV – see Chapter Two). Other than the scores, background

---

[75] Details of how the data was cleaned before analysis and summary statistics are provided in Appendix C.

[76] Thailand participated in PISA from 2000 – 2018, and other years’ microdata is also available to download from the OECD website.

information of students and schools were collected in the context questionnaire. The information available is similar during 2012 and 2015, with the questions posed in the same way. However, one variable, school type, is only available in 2012[77]. In addition to student-level data, PISA provides information at item-level. These are a brief description of the question, its cognitive and content domain, as well as percentage of students answering the question correctly. This percent correct can be used to compare which parts the students perform well, and whether that is similar across assessments.

Similar to PISA, TIMSS data in focus is student-level data of 2011 and 2015 in Thailand. There are test scores, responses from the context questionnaire and item-level data[78]. One variable, school size is only available in 2011.

Missing data was not extensive for PISA (less than 6% in all variables and in total). The cases with missing data were dropped from the analysis. TIMSS, however, has more extensive missing data problem. Specifically, approximately 17% of participants omitted the question of parental education or had invalid answers. This is unlikely to be missing at random as mean Mathematics and Science scores of those answering “Don’t know” or have missing data are lower than the average in both years. This problem is mitigated by using multiple imputation to predict the missing values (Lee & Simpson, 2014). Variables used for the imputation include number of books at home and number of digital devices in the home[79]. The variables both predict the chance of missing values as well as the underlying values themselves as they represent socioeconomic variables and students who did not answer the questions of parental education are likely to be from lower socioeconomic status.

[77] It was anonymised in 2015 because of concerns that the schools may have been identifiable.

[78] In addition to school questionnaire answered by the principal, there are teacher questionnaires asked to the Mathematics and Science teachers of the selected class. However, these will not be used in this study as there are no similar questions present in PISA and O-NET.

[79] The question is “How many digital information devices are there in your home? Count computers, tablets, smartphones, smart TVs, and e-readers”. Students can answer “None”, “1-3 devices”, “4-6 devices”, “7-10 devices”, or “More than 10 devices”.

4.3.2.2. O-NET data

Of the three assessment data used, O-NET is the least known internationally. O-NET is a national yearly examination that is mandatory for students at grade 6, 9, and 12 from both public and private schools. The grade levels represent the end of primary, lower secondary, and upper secondary levels respectively[80].

Data that used in the research is anonymised student-level data, requested from NIETS, the organisation responsible for conducting O-NET assessments. At student-level, data available includes gender and achievement data (total score or percent correct) in tested O-NET subjects including Mathematics, Science, and Thai from 2010-2017 in three grade levels (grade 6, 9, and 12). A student unique identifier is used to link scores of students across grade levels as shown in Table 4-1. For example, a student who took grade 6 exam in 2007 would have taken grade 9 exam in 2010 and grade 12 in 2013. The first step in data preparation is to use this unique identifier provided to link students in the same cohort together. A small percentage of students (less than 1% in years 2017 and 2018) who do not have a unique identifier were dropped as they could not be linked with scores from other grade levels. Then, students with partially missing test scores (in either grade 9 or 6) were dropped. Number of observations before and after cleaning can be found in Appendix C1.

[80] More information on O-NET is in Chapter Two.

Table 4-1 O-NET longitudinal dataset

Grade 6	Grade 9	Grade 12
2007	2010	2013
2008	2011	2014
2009	2012	2015
2010	2013	2016
2011	2014	2017
2012	2015	
2013	2016	
2014	2017	

In addition to student data, school names and basic characteristics were also provided, including whether the school is in rural or urban areas, what province, district, size, and educational service area the school is located. The schools could also be further grouped using school type subgroup as available in PISA[81]. There is no missing data at school level.

Other than microdata of achievement, O-NET's exam blueprints and exam papers are published every year. The blueprint contains the number of questions asked, total points, and question format[82] for each learning indicator based on the curriculum. These provide item-level data that can be analysed alongside the student-level microdata.

[81] There are additional categories of schools not included in the PISA sample including home-school and Buddhist schools.
[82] For example, multiple choice or short answer.

4.3.3. Methodological decisions

PISA and TIMSS are surveys designed with many characteristics regarding sampling and results calculation that need to be taken into account when analysing the data. The first part of this section, section 4.3.3.1 explores issues that come with complex survey data. These are the methodological decisions that the survey owners have made that can affect the analysis. The second part, sections 4.3.3.2 to 4.3.3.4, explores methodological decisions made by the researcher before conducting the analysis. This includes selecting regression models, sample restriction to make results comparable, and selecting which variables to include in the analysis.

4.3.3.1. Issues arising with using complex survey data

When analysing complex survey data such as PISA or TIMSS, researchers need to be aware of weights and plausible values. When they are not accounted for, the estimates produce different results. The weights and plausible values are features designed by the producers of the surveys. Weights are used for the results to be replicable, and plausible values are estimated as scores following the item-response theory.

In OECD's technical report (2017a), the use of survey weights are required to arrive at correct point estimates and standard errors for inference purposes. Two main weights relevant to this analysis are student weights and replication weights. Firstly, the student weights are used to ensure that the sampled students correctly reflect that of PISA/TIMSS population. The weights account for many design features such as the fact that schools are sampled proportional to size, that some subpopulations are oversampled, and non-response adjustment. By accounting for student weights, the subpopulations are proportional to the population and the coefficients are unbiased. Failure to adjust for this leads to large differences in the coefficients and standard errors, and hence statistical inferences. Secondly, replication weights take into account sampling design such as stratification and use of replacement schools. Adjusting for this does not change the point estimates, but it would correct standard errors. PISA and TIMSS vary in specification of replication weights[83]. Nonetheless, theses weights serve the same purposes in accounting for the complex survey design. Specifying replication weights ensures that statistical significance of each variable, deduced from standard errors, is correct. This is especially important for inferences to population statistics.

Additionally, both TIMSS and PISA utilise 'plausible values' in reporting students' test scores. As students cannot complete all test questions within two hours or less, they are assigned to complete only portions of questions. Based on the students' answers, the scores are generated using item-response theory and imputed for each student as plausible values. There are five plausible values for TIMSS and PISA up to 2015, when there are ten plausible values. The OECD (2017a) recommended that all plausible values be used for the estimations to be accurate. They must be treated as multiple imputations of missing values. In practical terms, this means following Rubin's rules in score estimation (Rubin, 1987) by estimating the parameter of interest five (or ten: depending on the number of plausible values generated) times, and averaging the results. For the estimation of standard errors, imputation error must also be accounted for. This increases the overall standard error.

With this, to produce unbiased estimates, the analysis needs to account for both weights and the use of plausible values. Some studies found that failure to account for weights led to significantly different results compared to when weights are included (B. T. West et al., 2015). In some datasets, however, not including weights does not lead to large changes in inference and implications (Carle, 2009). In the case of PISA data, running a simple OLS regression with and without weights have different point estimates, standard errors, and statistical significance (Lopez-Agudo et al., 2017). In contrast with large effects the weights have on

[83] BRR for PISA and jackknife for TIMSS

inferences, plausible values do not affect the results much. Lopez-Agudo et al. (2017) demonstrated that when researchers use only one value of the plausible values, they affect the inference minimally, as the results differ only slightly.

In this research, analysis will be done according to how the survey designers intended. Weights will be incorporated in the analysis and all plausible values will be used according to Rubin (1987)'s rules. To account for weights and plausible values, this study uses Stata commands, Repest (Avvisati & Keslair, 2020) and PV (Macdonald, 2019). The commands were developed to account for weights and plausible values as intended by the large-scale assessment programmes including PISA and TIMSS. They are used in studies using complex survey data (see Anders et al., 2016; Lopez-Agudo et al., 2017). Additionally, the estimations are cross-checked with the official reports where possible to ensure that the produced estimates are accurate.

4.3.3.2. Model choice

In contrast to the first decision, model choice decision is made by the researcher rather than the survey organisations. Using multilevel modelling would be an improvement over OLS regressions because multilevel models account for the nature of hierarchical data by acknowledging that students clustered in the same schools are more similar than students picked randomly from different schools. Multilevel models correct for the higher standard errors and make rejecting the null hypothesis more difficult. However, there are many complications when fitting multilevel modelling using complex survey data such as PISA and TIMSS (Carle, 2009; Rabe-Hesketh & Skrondal, 2006). This is because PISA and TIMSS requires the application of weights and the use of plausible values for the analysis. These features make applying multilevel modelling not as straightforward as O-NET data, for which no adjustments are required as it is already a population data.

There is no consensus in the literature on how to properly account for weights in multilevel modelling. Rabe-Hesketh and Skrondal (2006) suggest rescaling weights and using different weights at school and student level. They, however, did not take into account replication weights and did not deal with estimations using plausible values. Some suggest directly accounting for stratification and clusters when there is available information on the strata (Carle, 2009; Kreuter & Valliant, 2007). However, this is not possible for PISA and TIMSS as explicit and implicit strata are generally not available. Instead, replication weights are provided. The use of replication weights to correct for standard errors are less discussed and applied in literature. This is perhaps partly due to it not being supported in many statistical software (Kreuter & Valliant, 2007). Another option in analysing complex survey data is to account for design using all weights provided and analyse using single-level regression rather than multilevel modelling (Hahs-Vaughn et al., 2011). This approach allows for weights to be accounted for as intended, with the downside that the multilevel nature of the data is not accounted for.

As seen in section 4.3.3.1, when weights are not applied, results could change significantly. With this, running a simple regression while accounting for all weights is preferrable to running multilevel modelling and disregarding weights. Hence, this research opted for the method described by Hahs-Vaughn et al. (2011). OLS regressions are run for PISA and TIMSS data. At the same time, the multilevel results with applications of student weights (but not replication weights) are also provided in 0.

4.3.3.3. Sample restrictions

Another methodological decision made by the researcher is who to include in the final sample for analysis. As the purpose of the research is to compare the three assessment results, restrictions were made in the data used to ensure they are as similar as possible. This includes restrictions in subjects, years, school types, and scores.

Firstly, only subjects of Mathematics and Science are considered as TIMSS does not have a Reading subject as in PISA. Additionally, even though O-NET has Thai subject, the contents also include grammar as well as literacy. Hence, the differences in contents make it not comparable to PISA's Reading subject.

Secondly, the time frame of 2011-2015 is the focus of the study as it is the longest overlapping period of time the three assessments have been in place. O-NET data of 2012 are used in all analyses.

Thirdly, schools that are not included in other assessments' sampling frame are excluded. This includes vocational schools in PISA and minority school types present in O-NET such as religious schools. For PISA, this can be done by excluding schools based on variable "Unique national study programme code", which classifies whether students go to lower, or upper secondary, or vocational school. As some general track schools offer both vocational and general track, the school are excluded only if the majority of students are in vocational track. This process excludes 20% of the sample in PISA. However, there are some limitations. Vocational schools cannot be excluded from descriptive statistics regarding region and item types, cognitive and content domains. This is because microdata does not have this information and the statistics were taken, instead, from the official OECD or IPST report.

Fourthly, for O-NET, only students with available data for grade 9 and 6 scores are included in the sample. This is because we want to control for prior scores in models using O-NET data. Missing data varies by year. 9% and 4% of data were dropped in 2012 and 2015 respectively. Details can be found in Appendix C1.

With these adjustments, the sample of the three assessments are more similar.

4.3.3.4. Controlling for prior scores and socioeconomic status

Other than model and sample restriction choices, variables to be included in the model are also one of the crucial decisions made. Effect of school type on student outcomes can be from the ability of schools to attract higher ability or higher socioeconomic status students. To disaggregate the effect of this selection, prior scores and socioeconomic measures are included in the model. However, the limitation is that we do not have this information in the same dataset. Prior scores are available in O-NET while socioeconomic measures are available in PISA and TIMSS.

Both cognitive ability (general intelligence) and socioeconomic measures are consistently found to significantly predict test scores in a similar positive direction. There is high correlation between the two factors (Marks, 2013). This implies that students from advantaged background tend to have higher cognitive ability as well, which is likely due to better education received. In predicting achievement, prior scores are shown to have stronger relationship with achievement comparing to socioeconomic status. Marks (2017) shows that prior scores are better at predicting achievement than socioeconomic status in Australia in both year seven and year nine. Similarly, Armor et al. (2018) found that when predicting for achievement, controlling for prior scores reduce the significance of socioeconomic background of students. Most studies involving PISA and TIMSS, however, found strong relationship between scores and socioeconomic measures. It is possible that these measures are capturing contextual effects of prior achievement (Marks, 2013). Controlling for either factors should capture parts of student background and explain parts of the variation in scores.

Specifically, in O-NET, prior scores included in the model are from the subjects of Mathematics, Science, and Thai. These are standardised to have a mean of 1 and standard deviation of 0 for each year. For PISA and TIMSS, various socioeconomic indicators are available. Most often used indicators of socioeconomic status is number of books in the home variable (Carnoy & Rothstein, 2013; Jerrim &

Micklewright, 2012), parental occupation (Jerrim, 2012), parental education (Huang, 2010), home possessions, or PISA's index of economic, social and cultural status (ESCS) (Nieto & Ramos, 2015). As this research is making comparison over time, the socioeconomic variables chosen need to have the same meaning over time as well. This rules out home possessions as there are some country-specific variables which change over time. Additionally, the ESCS index is rescaled each year to have an OECD mean of 0 and standard deviation of 1. Hence, taking the ESCS index at face value might not represent accurate trend over time. Furthermore, Carnoy and Rothstein (2013) cautions that the index arbitrarily gives equal weights to the three socioeconomic variables: parental education, parental occupation, and possessions. This may not truly reflect socioeconomic status in a given economy/country.

In this analysis, instead of using the ESCS index, two socioeconomic variables are used: number of books in the home and parental education. Carnoy and Rothstein (2013) asserted that number of books in the home represents parental literacy, and hence socioeconomic status. In their research, number of books and parental education were considered because the same questions have been asked in TIMSS, which enables comparison between the assessments. In the end, since parental education may reflect different socioeconomic status in different countries[84], the variable may not be appropriate for cross-country analysis. Instead, number of books were chosen. In other literature comparing within-country results, parental education is used to represent socioeconomic status (Jerrim & Micklewright, 2012).

Additionally, for all models, school-level controls are also added such as school size, region, and whether a school is in an urban area. These proxy for resources that the schools have. For instance, a larger school would have more access to certified teachers as well as infrastructure than a small school.

[84] For instance, in countries where there is universal education, higher parental education may not reflect higher social status. This is not the problem for this research as the comparison is done within the one country.

4.3.4. Quantitative empirical strategy

4.3.4.1. Descriptive statistics

Descriptive statistics are used to answer primarily research question 2, “In what ways are the trends from PISA/TIMSS/O-NET similar and different?”. The metric of interest for the continuous variables (such as scores) is the mean and for the categorical variables (such as parental education) is the proportion of each category. These are calculated for the year 2011/2012 and 2015 to see change over time as well as how similar the mean and proportion are among PISA, TIMSS, and O-NET. For PISA and TIMSS, weights are accounted for in the calculation of descriptive statistics to ensure they reflect population values. This is done using Stata command Repest (Avvisati & Keslair, 2020) and PV (Macdonald, 2019) for PISA and TIMSS respectively. The descriptive statistics are cross-checked with the official OECD and IPST numbers when possible. No adjustment was done for O-NET as it is already population data.

Usually, statistical significance is used to assess whether the difference between the means reflect real difference or just variation in sampling. However, due to large sample size of O-NET[85], even small changes will be statistically significant. Yet, small magnitude of changes (or gaps) may not be of interest to stakeholders and policymakers. Carnoy and Rothstein (2013) suggest accounting for only differences that are greater than 0.1 standard deviation and treating smaller differences as not different from zero. However, they cautioned that due to sampling variations and general unreliability of a single-year test score, a difference of 0.2 standard deviation may be more meaningful and relevant to policies (18 points for PISA and 17 points for TIMSS for the US). In this analysis, 0.2 standard deviation is used as a cut point for meaningful changes/gaps. For

[85] 1,320,808 students in 2012 and 2015.

Thailand, the values for 2011/2012-2015 which are used as benchmarks are as shown in Table 4-2.

Table 4-2 Standard deviations of PISA and TIMSS

	PISA	TIMSS
1 standard deviation	80	95[86]
Meaningful difference threshold: 0.2 SD	16	19

Source: Own calculation using PISA and TIMSS microdata for Thailand sample

For scale, it is suggested that a gap of 30 points in PISA is approximately equal to one year of schooling (Jerrim & Shure, 2016)[87]. Therefore, a difference of 0.2 standard deviation is around half a year of schooling and is highly meaningful. Additionally, analysis for PISA and TIMSS are limited to groups that have at least 30 students to ensure sufficient number of observations for statistical analyses.

For ease of interpretation, after restricting the sample, O-NET scores are pooled across years and rescaled to have an overall mean of 0 and standard deviation of 1 for each subject. Hence, the meaningful difference threshold is 0.2. Standardising the scores is helpful as scores of different years have different mean and standard deviation, which may make direct comparison problematic. Additionally, whenever the O-NET scores are compared with PISA and TIMSS, PISA and TIMSS scores are pooled across years of analysis (2012 and 2015) and standardised to have overall mean of 0 and standard deviation of 1, as is done in O-NET. This is so that effect sizes can be easily compared.

Results from the descriptive statistics are presented in Chapter Six. Firstly, overall trends are investigated. Then, analyses are separated into one point in time and

[86] Note that the lower standard deviation of PISA in relation to TIMSS indicates smaller dispersion and thus, higher equity in scores of PISA than TIMSS.
[87] OECD suggested that this be used as a general rule of thumb rather than definitive, as it is based on a few research that shows different magnitude in results.

over time. Using 2015 data, gaps in learning outcomes for each relevant subgroup are investigated, before seeing who the high- and low-performers are, and how they changed over time.

4.3.4.2. Mapping content and cognitive domains

Descriptive analyses described in the previous section include only student-level variables. Item-level data is another source that can potentially explain the discrepancies between assessment results.

Comparison can be made of proportion of each items tested and how it changes over time. This comparison allows us to know whether the assessments have the same emphasis or not, and which tests are potentially easier or more difficult. For PISA and TIMSS, item-level characteristics are available[88] for all items, and item-level analysis can be conducted[89]. For comparison over time, analysis can be done with trend items[90]. For general comparison (such as proportion of item included in the test), all items can be used. For O-NET, by contrast, item-level data is limited. There are no item characteristics[91] nor scores provided for each exam item. Item-level data was requested to NIETS from the researcher but was not granted. Nonetheless, past examination items and exam blueprints are available.

The mapping of O-NET content and cognitive domains can be done manually, so that it can be compared with PISA and TIMSS. Using exam blueprint as a guide, the exam items are mapped to fit with the curricula content domains (see Appendix E2 for example of the mapping). For cognitive domains, the items are mapped with TIMSS's cognitive domains. TIMSS's descriptions are chosen over PISA's as TIMSS's questions are based on curriculum, similar to O-NET. Additionally, O-

[88] Description, content and cognitive domains, and percentage correct.
[89] Note that for many items, actual questions asked are not known as they are not released to the public. PISA and TIMSS only provides basic information such as the domains the questions are in and levels of difficulty.
[90] Items used in both rounds of assessments.
[91] Such as content and cognitive domains.

NET's questions are generally not as advanced as PISA's[92]. With the mapping, O-NET item proportion can be compared with that of PISA's and TIMSS's[93]. Note that there is no validation of whether the mapping is done correctly. To mitigate this, some of the mappings were shown to an educational expert to verify before proceeding with the analysis.

The analyses of content and cognitive domains are presented in Chapters Six and Seven to answer parts of research questions 2 and 3.

4.3.4.3. Multilevel model

To address research questions 1 and 3, regression analysis is used with specifications of multilevel models and OLS regressions.

Research question 1, "What are the factors that could explain variations in scores?", seeks to identify factors that are predictive of student achievement. This fits into the framework of educational effectiveness (Creemers & Kyriakides, 2008). Knowing what student and school factors relate to achievement allows us to have greater understanding of the problem of low learning levels. Here, using regression analysis is appropriate in modelling the relationship. Joint effects that different factors have on achievement can be observed. Additionally, as the regression models are analysed after the descriptive findings, the models are intended expand on some of the findings. Particularly, descriptive statistics show some school types significantly perform better than other schools. To help explain this, it can be analysed if the advantage is retained after ranges of student and school controls were added.

Multilevel models are used for analysing O-NET data. Most education data are hierarchical in nature, with students clustered in classrooms in schools. Hence, students from the same schools tend to be more alike than students from different

[92] For instance, students are not asked to interpret data or give opinion.
[93] Only proportion of items can be compared. There is no information on percent correct.

schools. When this is not accounted for, each sample member is treated as an independent unit; and, given their similarity with their class- and school-mates, true standard errors are significantly underestimated, and some variables may have been found to be statistically significant when they are not. However, the effect on point estimates are negligible (Huang, 2018; McNeish, 2014). Coefficients from OLS and multilevel models are similar in their magnitudes and implications. To correct for the standard errors, Multilevel modelling (MLM) or hierarchical model is used in much social science research. MLM allows residuals at different levels of data to be examined and produces unbiased standard errors.

The analysis starts with a null model as follows, then variables are added step-by-step as recommended by Raudenbush and Bryk (2002).

$$score_{ij} = \beta_0 + u_{0j} + e_{ij} \quad (2)$$

where $score_{ij}$ refers to score of student i in school j, u_{0j} and e_{ij} refer to school- and student-level residuals respectively. This is done separately for Mathematics and Science for 2012 and 2015 to see if there are significant differences in test scores between schools that should be accounted for or not.

When testing the results using likelihood-ratio (LR) test, all results are significant. This implies that the multilevel nature of the data should be accounted for. Intra-class correlations[94] are shown in Table 4-3. This indicates that around 80% of the variance in test scores is within schools, and the rest is between-school. Variations can also be within classes. However, due to data limitation, it is not possible to identify classes students are enrolled in.

[94] ICC shows percentage of variation in the outcome variable that is between-group, as opposed to within-group (Robson & Pevalin, 2016). The group in this case is schools.

Table 4-3 Intra-class correlations (ICC) of grade nine O-NET scores

Year	Mathematics	Science
2012	17.3%	20.1%
2015	16.3%	16.3%

Source: Own calculation using O-NET microdata

After the null models were created, residual plots were also produced to informally investigate the data and identify potential outliers (see Appendix C5). Most schools perform around the mean. There are some schools that perform significantly above the mean. These are almost exclusively Satit and Science schools. As discussed in section 2.3.2, these schools have selective admission policies and school curriculum that teaches higher level skills. These potentially make the schools top-performers.

With this premise, one could argue that Science and Satit schools differ so much from other schools that they should be excluded. Nonetheless, I argue for the inclusion of them as these schools still adopt the same national curriculum as well as being subjected to the same quality assurance system. Additionally, the research problem is concerned with the nature of low learning outcomes and inequalities of learning outcomes in Thailand. Hence, including these schools would show a more complete picture of the Thai education system.

The difference is accounted for in the next model. The reference group for school type is OBEC2, which is the public schools under the ministry of education. This reference is chosen as the coefficients can be interpreted as the advantage (or disadvantage) the different schools have in relation to the general public schools. For example, the coefficient tells us what the score differential is from enrolling in Satit school compared to OBEC2 schools.

$$score_{ij} = \beta_0 + \boldsymbol{\beta_1 Schooltype_{0j}} + u_{0j} + e_{ij} \tag{3}$$

Next, gender and grade 6 scores in Mathematics, Science, and Thai were added as student-level explanatory variables to the fixed part of the model. The model is written as follows.

$$score_{ij} = \beta_0 + \beta_1 Schooltype_{0j} + \boldsymbol{\beta_2 student_{ij}} + u_{0j} + e_{ij} \quad (4)$$

Prior scores of students offer invaluable information of student background and allow us to estimate the effect of schools more precisely using a value-added model. In regression analyses, it can be difficult to separate effect of schooling from effect of what happens in students' homes. This is because students are self-selected into schools with better academic outcomes. With the inclusion of prior scores, the effect of being in a particular school type, net of students' prior achievement, can be estimated. That is, it can be answered if some school types made more progress in terms of student outcomes compared to others. This offers a more precise effect schools make on achievement (Creemers & Kyriakides, 2008).

However, as O-NET is taken as students finish grade six, then at grade nine, there is a lag of three years in which data on schooling is not available. By taking the coefficients as school effect, we are assuming that students stay at the same school for three years (from grade seven to nine or throughout lower secondary). The data does not precisely define whether students were at the same school from grade seven to nine or not. While data is available whether students remain at the same school in at the end of primary and lower secondary (when they took the exam), it tells little about whether they remain in the same school throughout lower secondary. Many students moved schools after they finish primary due to many reasons[95]. Nonetheless, it is still a fair assumption that the majority of students remain at the same physical school throughout their lower secondary levels. There may be exceptional circumstances in which students have to move schools (such as migration), but these are likely to be exceptions rather than the rule.

[95] Some enrolled in schools without secondary levels. Others may move to schools with programmes matching their interests.

Lastly, school-level variables of size and location (regions and urban area dummy) were added in.

$$score_{ij} = \beta_0 + \beta_1 Schooltype_{0j} + \beta_2 student_{ij} + \boldsymbol{\beta_3 school_j} + u_{0j} + e_{ij} \tag{5}$$

For size and location, the lowest-performing groups (from the descriptive statistics) are picked as the reference group (small schools and lower north-eastern region).

The final model shows factors that can explain variations in test scores and factors contributing to school type advantage. Observing how the coefficient of school types change over time as more controls are added to the model would allow us to see the effect the student and school characteristics have on the advantage of being enrolled in certain school types.

The models here are random intercept models. It allows intercepts to vary by school IDs, and thus accounting for the multilevel nature of the data (Robson & Pevalin, 2016). It is possible to fit the data using a random coefficient model, where coefficients are allowed to vary alongside the intercepts. This was decided against as to make the comparison between MLM and OLS models as comparable as possible, since PISA and TIMSS data will be estimated using OLS.

Another part of the research questions that multilevel models can answer is question 3 "What are the factors contributing to the discrepancies of trends over time?". To investigate if the sample composition changes over time, data is combined for both years, and multilevel models are estimated with year as a main dependent variable of interest. Other dependent variables related to school and student characteristics are then added stepwise as presented previously.

$$score_{ij} = \beta_0 + \boldsymbol{\beta_1 time_{0j}} + \beta_2 student_{ij} + \beta_3 school_j + u_{0j} + e_{ij} \tag{6}$$

By observing the year coefficient, it can be investigated whether change over time can be explained by sample composition of specific variables or not. If the effect of time became insignificant as school or student variable is added, it can be interpreted that change over time is influenced by changing sample composition of that specific variable. However, if the time coefficient and statistical significance do not change significantly as more variables are added, this means that sample composition of main subgroups do not change significantly over time. In other words, there is no bias introduced through changing sample composition.

4.3.4.4. OLS regression

For reasons discussed in section 4.3.3, OLS regression is estimated instead of MLM for PISA and TIMSS data. The variables are added to the model stepwise, like in multilevel models with this final model as follows[96].

$$score_i = \beta_0 + \beta_1 Schooltype_i + \beta_2 student_i + \beta_3 school_i + u_i \quad (7)$$

Variables available for PISA and TIMSS differ from O-NET (see section 4.3.4.5). With the absence of prior scores, value-added model cannot be estimated as in O-NET. Nonetheless, the datasets have socioeconomic status variables. This means that part of the advantage of students that stems from better household background can be controlled for. This can then be compared with O-NET results.

Another shortcoming of PISA and TIMSS data is that school type variable is only provided in PISA 2012 microdata. TIMSS does not have a school type variable at all. This means that the effect of school type cannot be estimated as in the case of O-NET. We can only look at factors that are predictive of test scores. As school type is proved to be very important in explaining variations in scores, not including this makes the estimates less precise, as well as less useful. This is one of the limitations of this research.

Results should be interpreted with caution as OLS regressions are run instead of multilevel models. O'Dwyer and Parker (2014) compared estimation coefficients of factors predictive of students' scores. They showed that most coefficients are similar in direction, magnitude, and statistical significance. At the same time, some variables that are not statistically different show large differences in the coefficients

[96] School characteristics are school-level variables but estimated at a child-level.

and one variable changed its statistical significance. Therefore, the authors concluded that in the presence of multilevel structure of data, using OLS can sometimes lead to misleading results. For the present study, OLS regressions may show magnitude of coefficients that are less precise than multilevel models. To mitigate biases, the results will be compared with O-NET estimation results, where multilevel models were run instead of OLS regressions.

It is important to keep in mind that results from the regressions can be interpreted as correlation not causation. It is meant to describe factors that are related to variation in test scores, how this differs over time and across assessment data. These factors can then be inspected more closely in the qualitative phase of the study.

4.3.4.5. Variables

There is different data availability for the three assessments as shown in Table 4-4. Other than the scores of each student, the three datasets have different background variables. Variables that are predictive of achievement can be separated into student and school level. For student-level variables, socioeconomic status and prior scores have both been shown to significantly predict achievement (Jerrim, 2012; McCaffrey & Hamilton, 2007). Hence, they should be accounted for. However, the data is not available in all tests. As PISA and TIMSS's data are cross-sectional, there is no information on prior achievement. They, instead, collect a wide range of socioeconomic measures including parental education, parental occupation, home possessions, and number of books in the home. O-NET, by contrast, has data on prior scores, as students took the exams back in Grade 6. These grade 6 examinations are different tests, designed to test students of different grades, but have common subjects with Grade 9 students' exams. Lastly, item-level data are also available in all tests, but with limited availability in O-NET (see section 4.3.4.2).

Table 4-4 Data availability of the main variables

Variable	PISA	TIMSS	O-NET
Student characteristics			
Gender	/	/	/
Socioeconomic measures	/	/	x
Prior score	x	x	/
School characteristics			
School type	2012 only	x	/
School region	x	x	/
Rural/urban dummy	/	/	/
School size	/	2011 only	/
Item-level data	/	/	limited
Number of students[97]	12,165	12,432	1,320,808

With the available data, variables used in the analyses are presented below.

Dependent variable – test scores

Scores of Mathematics and Science of all assessments are used as an outcome variable for the regression. For PISA and TIMSS, the scores are provided as plausible values[98]. The scores are in the scale as initially set by the OECD/IEA (with mean of 500 points and standard deviation of 100 points). For O-NET, the scores are provided as percent correct according to the scoring scheme of the exam of that year[99]. They were pooled across years and standardised to have an

[97] For the years 2011/2012 and 2015
[98] See section 4.3.3.1 for more information.
[99] Some questions are worth more points than others, such as short answers or questions with more than four choices.

overall mean of 0 and standard deviation of 1 for each subject to ease interpretation.

Independent variables

Table 4-5 summarises the variables used in the quantitative analysis at both the student level and school level. Appendix C4 provides details on exact questions asked in each questionnaire.

Table 4-5 Summary of student-level independent variables used in the quantitative analysis

Variable	Variable type	Description	Availability
Student-level variables			
Gender	Dummy	- Student reported gender - Categories: ‘female’ and ‘male’ - The reference category is ‘male’	PISA, TIMSS, and O-NET
Prior score	Continuous	- Students’ grade 6 scores in Mathematics, Science, and Thai - Values from 0 – 100 (percent correct) were transformed to have a mean of 0 and standard deviation of 1	O-NET
Parental education	Categorical	- Student reported parents’ highest level of education - The highest level of education that either students’ father or mother have	PISA and TIMSS

Variable	Variable type	Description	Availability
		- Categories were grouped to be 'Some primary or lower secondary or did not attend school', 'Lower secondary', and 'Upper secondary or above'[100] - The reference category is 'Some primary or lower secondary or did not attend school'	
Books at home	Categorical	- Student reported number of books at home that they have - Categories were grouped to be '0-10 books', '11-25 books', '26-100 books', '101-200 books', and 'More than 200 books'[101] - The reference category is '0-10 books'	PISA and TIMSS
School-level variables			
School type	Categorical	- Type of school as described in section 2.3.2 - Categories: 'OBEC1', 'OBEC2', 'BMA', 'Local', 'Private', 'Satit', and 'Science' - The reference category is 'OBEC2'	PISA, TIMSS, and O-NET

[100] TIMSS's categories are more granular than PISA and had to be regrouped. See questions asked in Appendix C4.
[101] PISA's categories are more granular than TIMSS and had to be regrouped. See questions asked in Appendix C4.

Variable	**Variable type**	**Description**	**Availability**
School size	Categorical	- Size of school[102] - Categories: 'S', 'M', 'L', and 'XL' - The reference category is 'S'	PISA, TIMSS, and O-NET
Urban area	Dummy	- Whether the school is in rural or urban area[103] - Categories: 'Urban' and 'Rural' - The reference category is 'Rural'	PISA, TIMSS, and O-NET

Some variables are available across all assessments (gender and all school-level variables). Prior score is available for O-NET data only. For PISA and TIMSS, socioeconomic variables of parental education and books at home were used instead. Both are categorical variables asked in a student questionnaire. For parental education, in the regression, parental education is the highest level of education of either students' father or mother. Some categories differ slightly with TIMSS being more granular than PISA in parental education and less granular in books at home. For comparability, the categories of both variables were regrouped.

School size and urban area are derived from other variables. School size is coded based on enrolment to match the definition used typically in schools. Urban area is coded based on school location. A school is categorised as urban if the answers are "A city" or "A large city" in PISA and "Urban–Densely populated" in TIMSS. Otherwise, a school is categorised as rural. For O-NET, this information can be found with O-NET data[104]. Urban schools are schools located within the municipal

[102] Derived from enrolment number, with S: less than or equal to 500 students, M: 501-1,500 students, L: 1,501-2,500 students, and XL: greater than 2,500 students
[103] Derived from school location question (PISA and TIMSS) and based on capital municipalities in O-NET. Schools in Bangkok and its perimeter are all urban by definition. Urban schools also include those located in the capital municipalities of other provinces (amphoe mueang).
[104] In the year 2011 and 2012, there seems to be a mismatch in how the variables were coded by NIETS in other years. For example, a school in urban area in 2013 – 2017 were coded as rural,

area and rural are schools outside the municipality. This definition mirrors what is asked in PISA and TIMSS.

Other variables

Region is a location categorical variable asked to schools. In PISA, there are nine regions: Bangkok and perimeter, Central, Upper North, Lower north, Upper Northeast, Lower Northeast, South, East, and West. The regions are the same with TIMSS data, except that Bangkok and the perimeter provinces were separated. These are grouped together so that it mirrors PISA. For O-NET, province names were provided along with the school. These are categorised into PISA regions as well. Region is not provided as a variable in the student-level data for both TIMSS and PISA. Hence, it cannot be included in the regression for TIMSS and PISA. Descriptive statistics as reported by IPST were presented in Chapter Six. For PISA, the statistics in the report are from the sample that includes both general track and vocational track schools. As this does not appear in the microdata, it cannot be separated out manually.

4.4. Qualitative methods

This section presents the qualitative methods, starting with rationale for the qualitative methodology used in section 4.4.1, before discussing researcher's position in section 4.4.2, and diving into empirical strategies used in section 4.4.3.

---

and vice versa. Hence, the values of urban variables were changed for year 2011 and 2012 based on the value of year 2013 data.

4.4.1. Rationale for the methodology

As part of a sequential mixed methods design, the qualitative part of the research is conducted after the quantitative investigations to answer the first research question of “What are the factors that could explain variations in scores?”. The qualitative analyses are intended to complement, deepen, and expand upon the quantitative results and offer explanations to the research questions not covered in the quantitative data. Hence, the main quantitative findings are used to inform design and research questions of the qualitative phase. Specifically, the quantitative findings for research question 1 show largest score gaps among school types that cannot be explained by school and student characteristics alone. To complement the findings, the gaps among school types are further investigated in the qualitative part of the study. This is done by using school type as a stratification in sampling schools and teachers to be part of the research. Ensuring all school types are included helps in identifying the differences among them and answer what contributes to the achievement gaps.

To answer the research question, data collection is done using semi-structured interviews[105] with teachers and vice principals. Semi-structured interview is often employed in qualitative or mixed methods research in the field of educational effectiveness (Hadfield & Chapman, 2015). Since the research seeks to understand what the interviewees think influence score gaps and changes, the interviews are appropriate as this allows them to reflect on their experience with these issues. Semi-structured nature of the interviews allows participants to go deeper into specific topics, without being too constrained by the question structure. Usually, semi-structured interviews are conducted face-to-face with the participants. However, due to the COVID pandemic, data is instead collected with a hybrid approach – both online and face-to-face where possible.

[105] In Thai, the main language of the participants.

Alternative methods of qualitative data collection were considered such as ethnography and classroom observation. They would offer rich data on practices at the school, and what teachers focus on in class, as opposed to interviews, where participants recall their teaching techniques. When use in conjunction with the interviews, it can offer a complete picture of participants' current practices. However, it is not feasible to conduct ethnography and classroom observation in the time frame of this research (November 2020 – June 2021). With the COVID pandemic, permission to conduct research in schools is extremely difficult to obtain. Both ethnography and classroom observation require researchers to be immersed in schools for certain period of time, and this is not feasible during the pandemic. Therefore, the research opted for semi-structured interview, which does not necessarily require the researcher to be present in schools.

In addition to the design, the quantitative findings informed what questions to be asked to the participants. Particularly, differences in trends may be caused by differential performance in items of different skills (e.g., students perform better on lower-skilled items and worse on high-skilled items). Hence, in the qualitative part, the questions asked to participants include the skills and students that teachers focus on, as well as factors not covered in the PISA and TIMSS surveys. Parts of interview questions also focus on how practices or perceptions have changed over time from the first experience of the participants with PISA to the time of the interview. Topics of focus include the aspects of curriculum, exam focus, student background, principal policy, teacher training, and attitude. Participants' experiences cannot be used directly to answer research question 3 regarding discrepancies in trends, as they are individual experiences from a non-representative sample. Despite this, knowing what changed over time provide important insights in understanding causes of low learning outcomes.

The interview data is then transcribed and analysed using thematic analysis. Thematic analysis is the method most commonly used in the field (Hadfield & Chapman, 2015). It is also appropriate as a method to find connections among the data (Miles et al., 2014). For this research, this is an appropriate method in making sense of the participants' experiences related to the research questions. Thematic analysis allows researchers to arrive at a common narrative that can explain what influences score gaps among schools as well as changes over time. The findings are then analysed alongside the quantitative findings.

4.4.2. Researcher's position

Researchers' position can have effects on the findings. This can be through participants not answering truthfully or act normally in the presence of the researchers (Miles et al., 2014). Data analysis is also affected by my own values and positioning. Hence, it is important to discuss my position as a researcher when conducting the fieldwork and how that may affect data and subsequent analysis.

For the interviews, I was able to contact the participants mainly from my Teach For Thailand alumni network. Teachers and vice principals were recommended to me by the network. With this, the introductions went smoothly, and most participants were eager to participate. They were not pressured to participate in the study. In fact, due to busy schedules of some participants, some refused to participate in the study. In most cases, I was able to find replacements of another teacher or management position personnel who fit the criteria. Since the participants are from the same schools, this is less likely to bias the findings.

During the interviews, participants were encouraged to speak freely. Confidentiality of the interview was emphasised. As a Thai national myself, I was able to conduct the interviews without the need of interpreters, which may have helped with confidentiality. With this, I feel participants were able to speak freely and critically about the assessment system, incentive structure, and pressure they receive. Some even admitted to cheating or gaming the system.

Nonetheless, given that I do not work in the schools the participants are in, I remain an outsider in their perspectives. This may influence some participants to sugar-coat their opinions on controversial matters such as how the school is pressured by different stakeholders of the system. Additionally, as illustrated by one teacher, there is a culture of distrusting the outsider, specifically in discussing aspects that are not favourable to the schools.

> *"[Another BMA school] didn't join the programme that involves bringing junior teachers into the schools temporarily. They said they were afraid that the juniors will talk to the public about what's happening at the school, about the education in BMA schools (Mathematics teacher, BMA school)."*

Nonetheless, I tried to mitigate this by including multiple participants from the same school. This could be used to compare what participants say, in attempt to form a coherent narrative of what it is like at the school. In some instances, some follow-up clarifications were made to the participants via messaging.

4.4.3. Qualitative techniques

4.4.3.1. Sample selection and sample characteristics

The sample selection was designed with the research questions and quantitative findings in mind. To answer research question 1 (What are the factors that could explain variations in scores?), the sample should include both high-performing and low-performing schools. School type, identified as one of the main predictors of test scores, is used as the main strata in school selection to ensure schools at different achievement levels are included in the sample. Consequently, the schools were firstly selected using stratified purposive sampling. Then, the main participants of the research, teachers and vice principals/academic teachers, were selected for each participating school.

The final sample includes different schools of Satit, Science, OBEC1, OBEC2, BMA, Local, and private schools. Brief descriptions of the schools are provided in Table 4-6. Overall, they vary in terms of location (urban/rural), size and region, as well as contextual features. As OBEC2 and Private schools can be quite heterogeneous, two schools were included from these groups to reflect different contexts of the schools. OBEC2 schools account for 55% of enrolment and composed of schools of all levels of resources and performance. To illustrate this difference, OBEC2 school 2 is an extra-large school at the perimeter of Bangkok whereas school 3 is a small border school serving indigenous population. For private schools, independent and government-dependent private schools can be different as the latter receives the majority of funding from the government and are limited in how much they can charge students, while the former is generally more prestigious and can charge higher tuition (see section 2.3.2 for further details). School 6 in the sample is a prestigious independent private school while school 5 is an outlier dependent school which performs well on O-NET despite being dependent on government for funding. By including schools of different achievement levels, the findings can shed light on what high- and low-performing schools do differently, helping to answer research question 1.

The sample is not meant to be representative of schools in Thailand. It is a mix of schools that are typical of schools of that type[106] as well as schools that are outliers in terms of learning outcomes. Here, school 5 and 7 score better than as predicted from their school types. These outliers were included to provide insights on how schools can perform well even with similar constraints with the schools of the same type. Two schools that were contacted refuse to participate (BMA and private schools). Hence, replacement schools with similar characteristics were included instead. A total of nine schools were included in the final sample.

[106] In terms of O-NET performance, enrolment, school size, and other observable characteristics.

Table 4-6 Sampled schools' characteristics

School	Type	Description
1	OBEC1	A medium-sized school in the rural area, serving children of wage workers with unstable income. O-NET scores are below average.
2	OBEC2	Extra-large school located within the perimeter of Bangkok. Students came from both within and outside catchment area. Agriculture is the main parental occupation. O-NET scores are below average.
3	OBEC2	A small rural school located near the border to Laos, with majority of the students being indigenous population. Main occupation of the community is agriculture. O-NET scores are average.
4	BMA	A medium-sized school at the outer part of Bangkok, serving mainly children within the catchment area. Their parents are mostly wage workers. Many students live with their grandparents or with single parents. O-NET scores are below average.
5	Private	Extra-large school with long-standing reputation for academic rigor. The school is partially subsidised, meaning it charges less tuition comparing to unsubsidised schools. O-NET scores are above average.
6	Private	Extra-large unsubsidised school serving students from average to above average socioeconomic status. Parental occupation is diverse. Most parents are university graduates. O-NET scores are above average.
7	Local	A model school of the province. Parents have diverse background and socioeconomic status. The main occupation is civil servant. O-NET scores are average.

School	Type	Description
8	Science	A medium-sized school in the rural area. Relatively low-performing in relation to other Science schools. O-NET scores are above average.
9	Satit	Elite school in Bangkok, with selective admission policy. Most parents are university graduates. O-NET scores are above average.

Sources: Description of schools based on participants' account and school websites. O-NET scores derived from microdata.

After the schools were selected, a vice principal (or principal or academic department teacher) and two teachers were invited to participate in a semi-structured interview. Vice principals and teachers were selected as they are practitioners and hence the best people to answer the research questions, particularly about how assessment affects practices, and changes over time. The separation is made between teachers and vice principals as vice principals are more appropriate in answering questions at a school-level, such as school policies and school vision, while teachers can offer classroom-level perspectives. The accountability system may also be different for management position comparing to teachers. Originally, I planned to interview head teachers. However, since most principals' terms are for four years[107] (MOE, 2021), most principals present during the research's timeframe have transferred schools already. Therefore, vice principals or academic department teachers[108] were contacted instead of principals, as they tend to stay in school for longer. They are also well-suited to answer management-level interview questions. By contrast, teachers are more appropriate to answer questions regarding teaching and assessment practices.

The participants should also have the following characteristics:

[107] It is possible to extend the term of service to eight years.

[108] These teachers work in the academic department of the school. As they work closely with principals and vice principals, they are good replacements in cases where vice principals could not be contacted or refuse to participate.

For teachers

- Is currently a teacher in either Mathematics or Science (or Thai, if no other suitable candidate) at a secondary level
- Has been teaching at this school for at least nine years (since 2011)
- Has experienced a PISA or TIMSS assessment at least once

For vice principals (or principals or academic department teacher)

- Has been working in management role for at least nine years (since 2011)
- Has experienced a PISA or TIMSS assessment at least once

Time criteria are set to aid answering change over time. By being at the school since the timeframe of the study, the participants can provide information on change over time. They also have to experienced PISA or TIMSS at least once, so that they are familiar with the test and can answer questions about test preparation. The participation can be from any test cycle. In total, there are 27 participants from nine schools. In the same schools, I tried to have a mix of male and female teachers, as their views on assessments may differ. This is not always possible since the majority (73% as of 2018) of teachers in Thailand are female (EEF, 2021b).

Some sampling criteria were not met. One of the criteria is for the teachers to teach Mathematics or Science as these are the subjects of focus in the quantitative part of the research. However, in some schools (3 and 7), Thai teachers were assigned the main roles in PISA preparation. Hence, these teachers were included instead as they would provide interesting insights into how these schools engage with and prepare for the assessments. Additionally, five teachers have been in the school for less than nine years. One of the reasons is that in some schools, such as OBEC1, an academic department teacher reported that teacher turnover is typically high for this type of school. These schools are still included in the sample as replacement schools would have had similar problems, and the participants included fulfilled other criteria (teach in relevant subjects and participated in either PISA or TIMSS).

Characteristics of participants can be found in Table 4-7.

Table 4-7 Participant characteristics

No.	School	Position	Sex	Length at school (years)
1	1	Mathematics teacher	Female	4
2		Academic teacher[109]	Female	10
3		Science teacher	Female	4
4	2	Vice principal	Female	10
5		Mathematics teacher	Female	26
6		Science teacher	Male	5
7	3	Academic teacher	Female	13
8		Thai teacher	Female	12
9		Mathematics teacher	Female	8
10	4	Mathematics teacher	Female	11
11		Academic teacher	Female	11
12		Science teacher	Female	11
13	5	Vice principal	Female	7
14		Science teacher	Female	11
15		Mathematics teacher	Female	24
16	6	Academic teacher	Male	20
17		Mathematics teacher	Male	21
18		Thai teacher	Female	28
19	7	Thai teacher	Female	14
20		Mathematics teacher	Female	14
21		Principal	Male	14
22	8	Science teacher	Female	17

[109] Academic teachers are those who work in academic department. They often have academic-related management roles in addition to teaching duties.

No.	School	Position	Sex	Length at school (years)
23		Mathematics teacher	Female	11
24		Vice principal	Male	10
25	9	Mathematics teacher	Female	15
26		Vice principal	Female	23
27		Science teacher	Male	9

Source: Participants' interviews

4.4.3.2. Data collection: Semi-structured interviews

Data was collected via semi-structured interviews. Compared to structured interviews or quantitative interviews, semi-structured interviews offer more flexibility and focus on participants' perspectives (Bryman, 2012). As the study seeks to understand the views of teachers and vice principals regarding assessments, semi-structured interview is fit-for-purpose. Additionally, this method of data collection is more appropriate for information about intangible aspects such as values and beliefs (Cohen et al., 2007), which is also of interest in this research.

The ideal method would be if the interviews were conducted face-to-face. However, due to COVID risk, the face-to-face option is not always possible. Since the COVID-19 risk in Thailand was low to moderate at the time of data collection (according to the FCDO travel advice), I was able to travel to Thailand to conduct fieldwork. Before participating, other than the general information, participants were informed of risks regarding transmission of COVID-19. With this, they can either opt for online or face-to-face interviews. However, due to risk level perceived by the participants, most interviews ended up being online, with some face-to-face interviews. All interviews were conducted in Thai, the native language of the participants and the researcher. The interviews lasted from 50 minutes to two hours, with an average of around 80 minutes. The interviews were recorded and transcribed in Thai, with partial translations quoted in the thesis. Participants' personal data was pseudonymised so that they cannot be identified.

4.4.3.3. Interview questions

Research question 1 is of the focus of the qualitative study. The qualitative sub-question is "What could explain score gaps among high- and low-performing schools?", with the following sub-questions.

- How do the accountability systems differ among school types? How does that affect school autonomy?
- How do different schools experience pressure to perform on different assessments?
- What strategies do schools engage in to increase assessment performance? How does that relate to school characteristics?
- How have the three assessments affected the teachers' practices in lesson planning and assessments? To what extent does the existence of PISA and TIMSS distort what is taught in classrooms?
- How have the three assessments affected schools' priorities and practices?
- What do practitioners see as factors affecting scores?
- What are the changes that happened during the period of 2012 and now in terms of curriculum, exam focus, student background, principal policy, teacher training, attitude etc.?"

These research questions were broken down into smaller questions that form the actual interview questions asked to participants. Examples of this is in Table 4-8. The interview questions were separated into four aspects: personal information, awareness of the assessments, pressure from the assessments, and assessments and practices. There are different questions for participants who are teachers, and those who are in the school management team. Questions asked to teachers focus more on what happened in classrooms (teaching practice and assessment) whereas questions asked to the management team focus on the school-level aspect.

The first sub-question, “How do the accountability systems differ among school types? How does that affect school autonomy?” aims to study the accountability system faced by the school. This can be understood from the two aspects of quality assurance and performance appraisal (career and salary progression). Hence, the participants were asked to describe how these processes happen at the school. The second sub-question, “How do different schools experience pressure to perform on these assessments?” can be broken down further to “awareness of the assessments” aspect, where participants were asked to describe what the three assessments measure, and what do they see as pros and cons of the assessments, and “pressure from the assessments” aspect regarding how these results affect them. Particularly, aspects of interests are what learning outcome indicators are present, how much pressure was put on the school to accomplish the goals, and from where/whom does the pressure originate? This links to the third, fourth, and fifth sub-questions of how the assessment pressure affects practice of teachers and vice principals. These categories include questions asking about details of how participants prepare students for the tests, how teachers teach and conduct both formative and summative assessments, how assessment results are used at the school, etc. The sixth sub-question mirrors the quantitative question of what factors affect the scores, but from the perspectives of the practitioners. Lastly, as the last sub-question focuses on change over time, the participants were asked these questions both for the current practices and past practices (nine years ago and when they first had experience with PISA/TIMSS).

The questions are then translated into Thai and piloted with two teachers and a vice principal before being adjusted based on the answers and participants’ suggestions (some questions were rephrased, others were cut or added). The final list of interview questions is in Appendix C5. Note that as this is a semi-structured interview, the questions are not necessarily asked in this order. Some questions were added, and some topics were further probed depending on the interview.

Table 4-8 Examples of how interview questions are formed

Sub-questions	Example questions
2.1. How do the accountability systems differ among school types? How does that affect school autonomy?	- What is the quality assurance system like at this school? - How are learning outcomes indicators include in quality assurance? - What is the process of career progression and salary progression?
2.2. How do different schools experience pressure to perform on these assessments?	- Can you describe what PISA, TIMSS, and O-NET measure? - What are the pros and cons of these assessments? - How do assessment results affect you?
2.3. What strategies do schools engage in to increase assessment performance? How does that relate to school characteristics?	- Can you give examples of some policies you have implemented to improve student learning? Who sets those policies? How successful are those policies?
2.4. How have the three assessments affected the teachers' practices in lesson planning and assessments? To what extent does the existence of PISA and TIMSS distort what is taught in classrooms?	- What knowledge or skills do you focus on in your assessment? - Can you describe formative and summative assessments of your subjects?
2.5. How have the three assessments affected schools' priorities and practices?	- How have the assessment results been used at school? - What are the schools' values, goals, and policies on learning outcomes? How are

Sub-questions	Example questions
	these developed? How are they measured?
2.6. What do practitioners see as factors affecting scores?	- What are factors that affect learning outcomes at this school? - What are some of the challenges in improving learning outcomes?
3. What are the changes that happened during the period of 2012 and now in terms of curriculum, exam focus, student background, principal policy, teacher training, attitude etc.?	- How has the school's academic focus changed over time? - How have your teaching and assessment practice changed over time?

4.4.3.4. Thematic analysis

Data is analysed using thematic analysis, which is a method for identifying, analysing and reporting patterns (themes) within data (Braun & Clarke, 2006). In the analysis, the data is initially coded following the coding framework. Here, both deductive and inductive coding techniques were employed. In deductive coding, some codes were pre-identified prior to data collection based on literature, research questions, and findings from the quantitative part (Miles et al., 2014). Some of the pre-defined categories include accountability, incentives, test preparation strategies such as tutoring, teaching to the test, etc. These are the themes that came up during literature review of factors that affect pupil learning outcomes. In inductive coding, by contrast, the codes are grounded empirically. Additional contents that the participants mentioned can be coded separately from the pre-defined codes. For example, issues of misconceptions about the assessments and comparison with other schools are important topics that were not pre-defined. Some of the pre-defined codes are adjusted based on data. For example, pressure to perform on assessment was made more granular by separating into pressure to score above a threshold and pressure to show improvement in scores. After the data is coded, they are sorted into potential themes that form coherent narrative that answers the research questions. The themes are reviewed and summarised in Chapter Five.

Chapter Five: What Are the Factors That Could Explain Variations in Scores?

5.1. Introduction

The chapter starts with descriptive statistics for each school type to help explain how they differ from one another. Then, these differences are controlled for using regression analyses, to see if the schools maintain their advantage after controlling for student background and school characteristics. Additionally, the models allow analysis of school and student-level factors that are predictive of student achievement. Next, data from the qualitative study are presented to help explain the ways that the high-performers differ from the low-performers, based on differences not captured in the quantitative data. Then, I investigate schools' pressure to perform on different assessments, and how that affect practices, both in terms of school priorities and classroom practices.

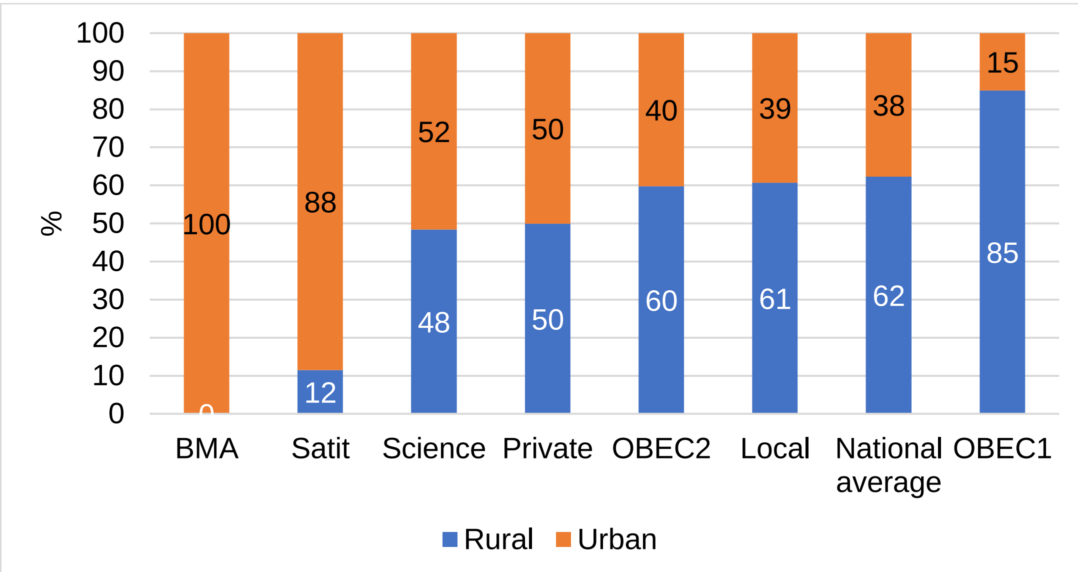
5.2. Quantitative results: Descriptive statistics

This section explores school characteristics of different schools to determine if they potentially drive large gaps observed between school types.

5.2.1. School characteristics of each school type

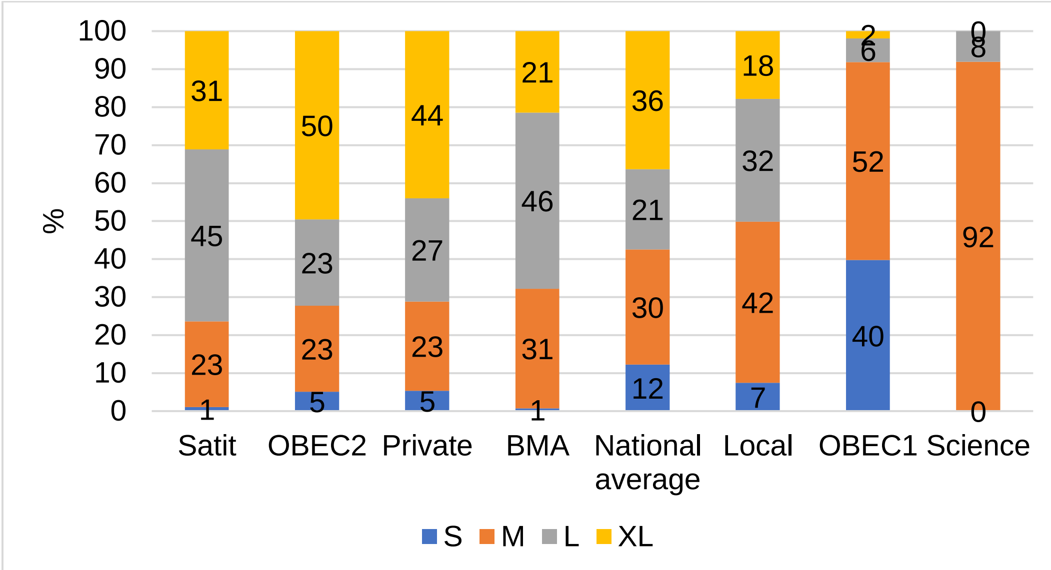
Figure 5-1 and Figure 5-2 show clear differences among schools in percentage of enrolment that are in rural area and percentage of school sizes for each school type. Satit and BMA schools are disproportionately located in urban areas, whereas OBEC1 schools are mostly in the rural area. Most of the schools are medium-sized, which means the total enrolment is around 501-1,500 students. OBEC2 schools have the largest proportion of extra-large schools (greater than 2,500 students). OBEC1 has the largest proportion of small schools (less than 500 students).

Figure 5-1 Percentage of enrolment that are in the rural and urban area by school type



Source: Own calculation using O-NET 2015 microdata

Figure 5-2 Proportion of different school sizes by school type

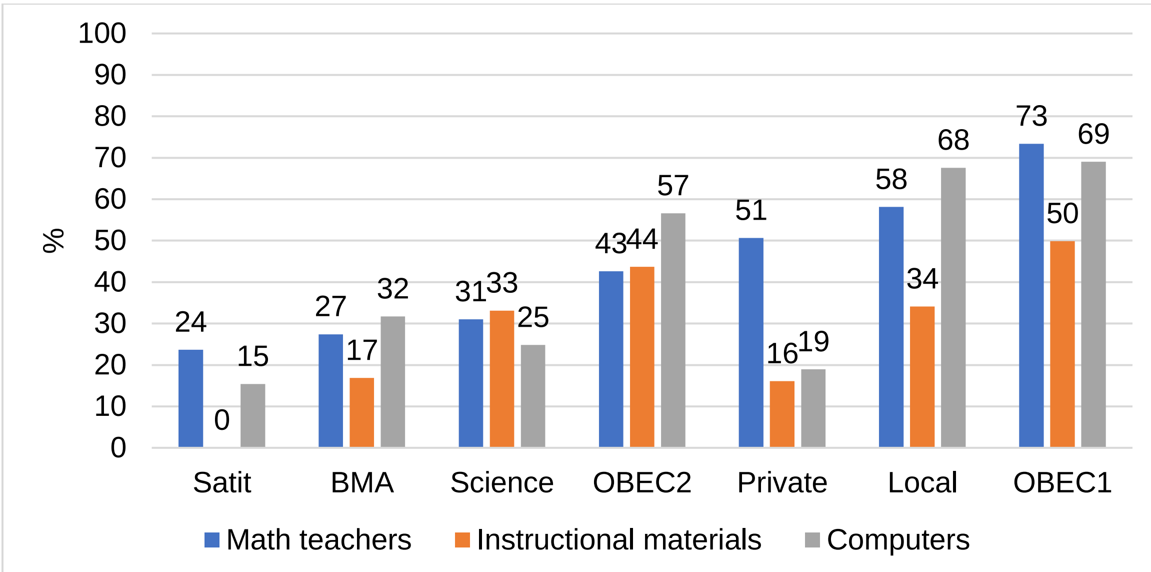


Source: Own calculation using O-NET 2015 microdata

A combination of these school characteristics affect score gaps in various ways. There can be many advantages of schools situating in the urban areas and being larger in enrolment. In terms of location, resources may be more available, and teachers may be more attracted to working in the urban areas than remote rural areas. Lathapipat and Sondergaard (2015) found that small rural schools in Thailand have trouble recruiting and retaining subject-specialised teachers. Nonetheless, there are schools that do not fall into this pattern. BMA schools are in Bangkok by definition and are all located in the urban areas. Yet, they have much lower learning levels. Similarly, Science schools are located more in the rural area, yet perform well academically. As for school size, since funding to schools in Thailand is on a per-head basis, small schools receive lower and usually inadequate funding (UNICEF, 2017). This implies that OBEC1 schools, which are generally small and are in rural areas, are more disadvantaged in terms of resources. Satit schools, by contrast, should have more resources.

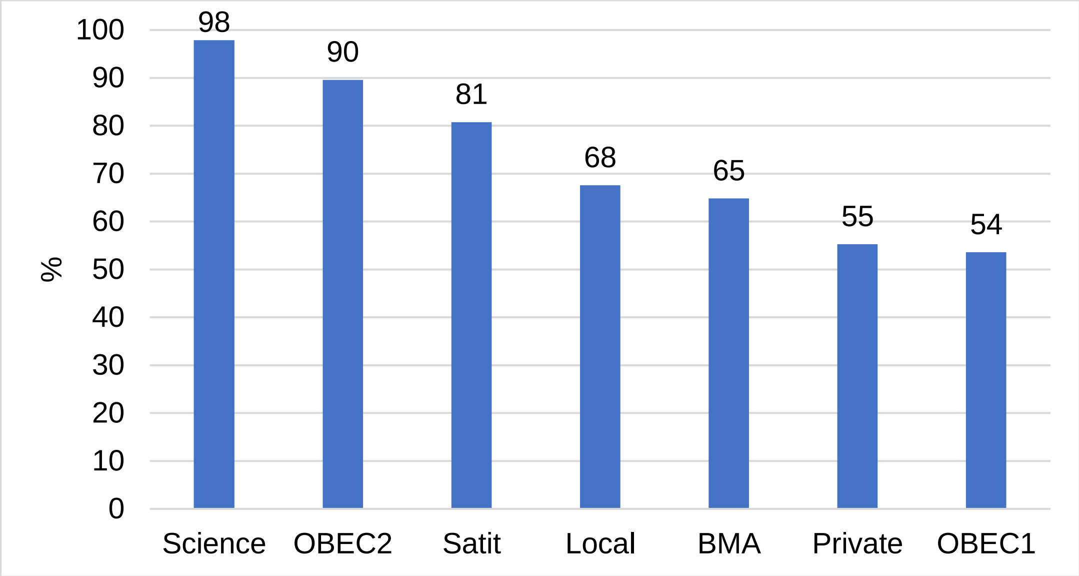
Data on school resources support the gap among school types. Figure 5-3 shows percentage of schools answering “To a certain extent” or “A lot” to whether the school's capacity to provide instruction is hindered by the issues of qualified Mathematics teacher, instructional materials (such as textbooks), and computers for instructions. Figure 5-4 shows proportion of Mathematics teachers in each school that graduated with Mathematics major, as reported by the principals.

Figure 5-3 Shortages of different resources by school type



Source: Own calculation using PISA 2015 microdata

Figure 5-4 Proportion of Mathematics teachers with Mathematics major by school type



Source: Own calculation using PISA 2012 microdata

The figures show that schools report shortages of resources to some degree. There are more shortages in computers, followed by staff shortages, then shortages in instructional materials. The figures show a considerable disparity among schools. At one end, Satit and Science schools reported much less problems on all categories of resources, and have the majority of teachers qualified. This shows that they are more well-resourced than other schools. Particularly, Satit schools have no lack of instructional resources at all. BMA schools are also very well-resourced as they are located in Bangkok, yet have less Mathematics major teachers. By contrast, OBEC1 and Local schools seem to have significantly lower level of resources (both comparing to Satit and OBEC2 schools). It is alarming that 73% of OBEC1 schools reported having shortages in qualified Mathematics teachers. Further, just over half of OBEC1 Mathematics teachers graduated with Mathematics major.

This implies that learning gaps among schools can be partially explained by schools being exposed to different levels of resources. Some schools may lack critical resources to improve learning levels. For schools with low level of resources

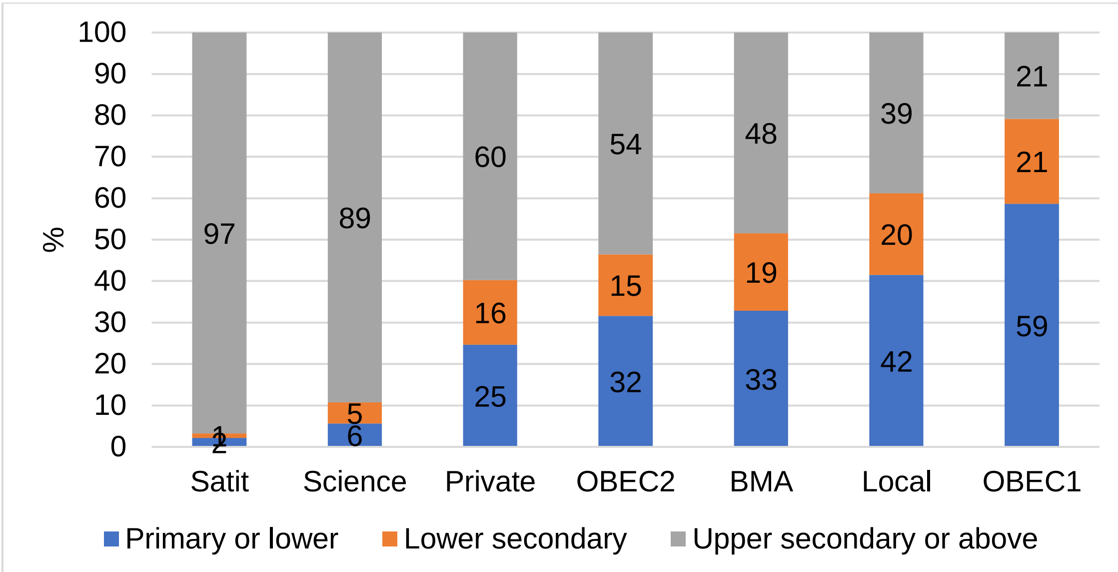
(OBEC1, and to a certain extent Local and Private schools), equalising resources can help lay foundation for the schools to improve learning outcomes. Giving resources alone, however, is not sufficient. BMA schools have high resources, yet low learning outcomes. Other factors may also be at play. A research by the World Bank (2012a) argues that the schools perform badly on assessments because they enrol students from poorer households. This shows that resource is only part of the problem and increasing access to resources alone cannot improve learning outcomes.

These figures also show which school types are more homogeneous. Particularly, OBEC1 and Science schools show more similarities among themselves. It is interesting to note that Science schools do not have as favourable characteristics as Satit schools, another top performer. Most are medium-sized schools and are in the rural area. OBEC2 schools, by contrast, are more heterogeneous, with schools of all sizes and in all regions. This should be kept in mind when discussing OBEC2 results, as learning outcomes can vary largely among different OBEC2 schools.

5.2.2. Student characteristics of each school type

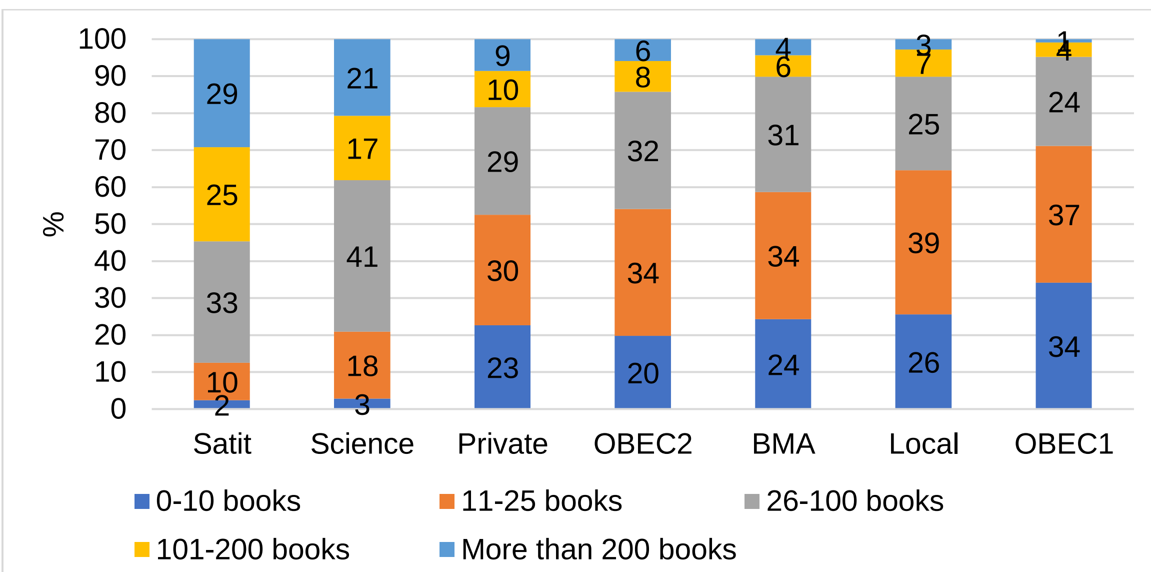
Top schools have better socioeconomic indicators than low performing schools. Parents of higher performing schools are more educated (Figure 5-5). The majority of parents in Satit and Science schools graduated upper secondary or above, whereas most parents of students in OBEC1 schools graduated only primary or lower levels. Another indicator of socioeconomic status, number of books in the home from PISA questionnaire, shows similar results (Figure 5-6). Other than Satit and Science schools, at least half of students have less than 25 books at home.

Figure 5-5 Proportion of highest parental education of students by school type



Source: Own calculation using PISA 2012 microdata

Figure 5-6 Proportion of books at home of students by school type



Source: Own calculation using PISA 2012 microdata

Similar trends are found with prior scores[110] (Figure 5-7) and proportion of students continuing education (Figure 5-8). As expected, the top performing schools have students with the highest prior scores, around one standard deviation higher than other schools. This shows that they enrolled students with greater cognitive ability than other school types, which may indicate better education these students have received prior to entering these schools. This implies that selection is indeed present, and forms parts of the reasons why Satit and Science schools perform so well. Students with better prior scores have better academic foundation than those who did not score well in grade 6 O-NET. This allows them to perform better in assessments at secondary level. By contrast, data supports World Bank's (2012a) research that BMA schools enrol students with poorer characteristics. Worse socioeconomic indicators and prior scores contribute to poor performance. When selection is at play, improving resources alone, or simply copying practices from Satit and Science schools cannot improve learning outcomes.

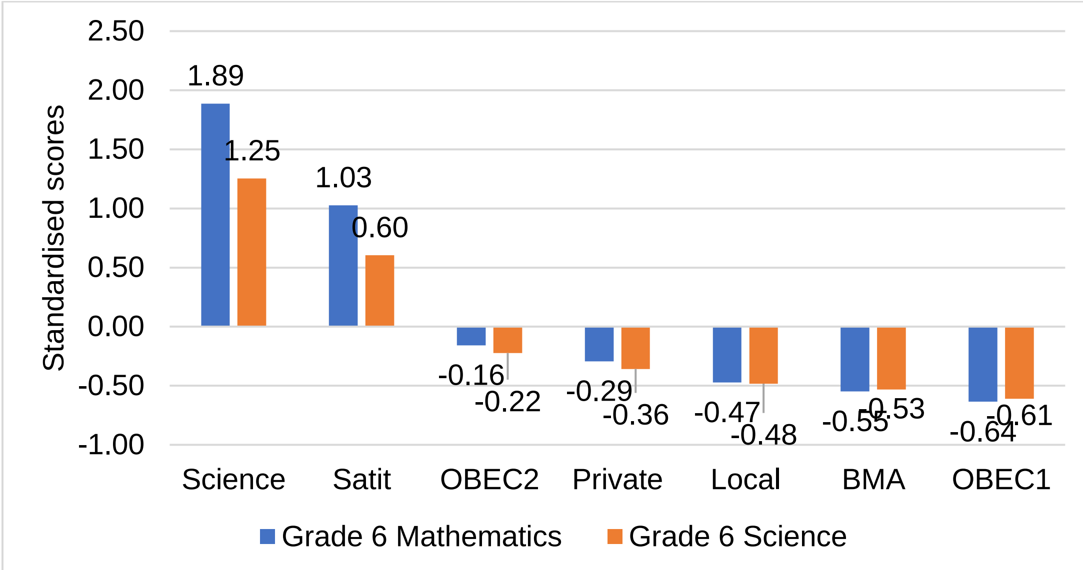
Additionally, Satit and Science schools have markedly higher percentage of students continuing to general track upper secondary. By contrast, the majority of students in OBEC1 and BMA schools did not continue to take the exam in grade 12. This shows that students in different school types have different educational aspirations and pathways[111]. While most of Science and Satit students follow general academic track and enter universities upon graduation, it is only true for some of OBEC2 and Private schools. Students of OBEC1 and BMA either dropped out after completing compulsory education or chose to continue in vocational track. The differences in aspiration can affect learning outcomes. If students do not intend to pursue academic progression, they may pay less attention to what is taught and assessed, resulting in lower learning outcomes.

---

[110] Standardised grade 6 O-NET scores

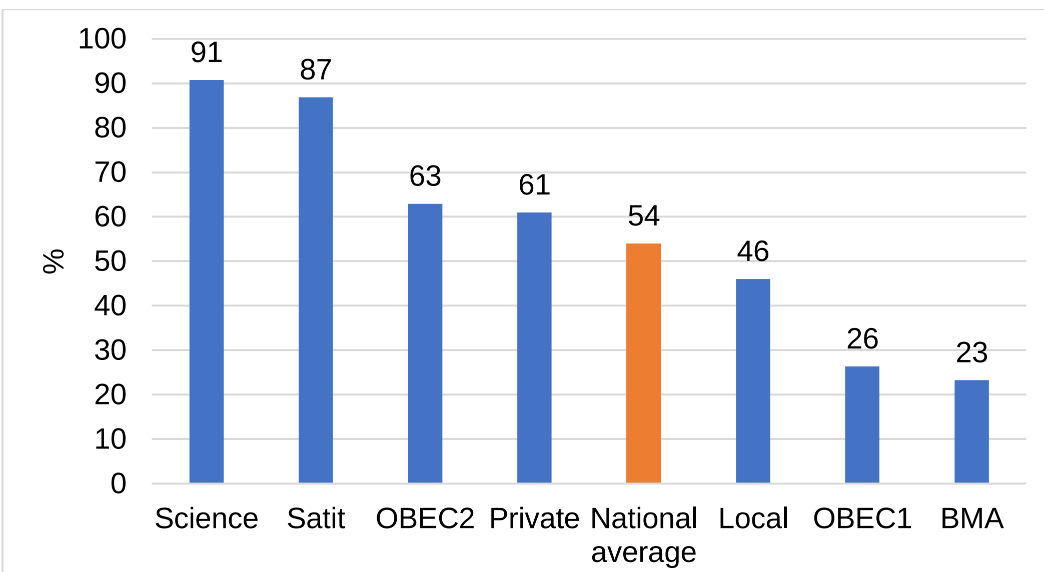
[111] They may not continue their education because they chose to go to work instead. This issue is explored in the qualitative results later on.

Figure 5-7 Average prior scores in Science and Mathematics by school type



Source: Own calculation using O-NET 2015 microdata

Figure 5-8 Percentage of students continuing to academic track upper secondary by school type



Source: Own calculation using O-NET 2014 microdata[112]

[112] 2014 data is used because it is the latest available data that is linked with grade 12 students

To summarise, the characteristics of each school type loosely mirror the achievement results. OBEC1 schools show the most disadvantaged characteristics. By contrast, Satit schools, one of the top performers, have the most favourable characteristics, both in terms of school and student characteristics. Science schools, which also perform well in all assessments, are more in rural areas and medium-sized. Despite this, they are able to attract qualified teachers and have favourable student characteristics similar with Satit schools. These contribute to both school types having an advantage academically. The better characteristics partly reflect their ability of selecting pupils into schools. With this, increasing resources alone, or making every school follow Satit or Science schools curricula and practices cannot help with improving learning outcomes. Further, since OBEC1 schools enrol much more students from disadvantaged background, there needs to be a measure to help students catch up on basic skills, before moving on to higher-level skills.

5.3. Quantitative results: Modelling results

5.3.1. O-NET data: Multilevel model results

In this section, multilevel model results estimated using O-NET data are presented. The models answer two parts of the sub-research questions. Firstly, they help with identifying what factors affect variations in scores. Additionally, by putting school type as the main variable, the models can estimate the effect of being in different school type on academic achievement, as well as explaining partly why some school types perform much better than others.

The models were estimated in a stepwise fashion. The analysis starts with a null model (1), with no explanatory variables. Then, school type variable was added (2), and student and school characteristics were included respectively in (3) and (4). This sequential addition allows us to see how the school type coefficient changed as more variables were included. Such steps were repeated for different subjects and years of data. The results for O-NET Mathematics 2012 were presented in Table 5-1 (see estimation results for other subjects and years in Appendix D1[113]).

Table 5-1 Multilevel model results for O-NET Mathematics 2012

	(1)	**(2)**	**(3)**	**(4)**
Fixed Part				
Constant	-0.21***	-0.21***	-0.12***	-0.20***
	(0.00)	(0.01)	(0.01)	(0.01)
School type (Reference: OBEC2)				
OBEC1		-0.02*	0.07***	0.11***
		(0.01)	(0.01)	(0.01)
Private		0.02**	0.08***	0.08***
		(0.01)	(0.01)	(0.01)
BMA		-0.05	0.01	-0.06*
		(0.03)	(0.03)	(0.03)
Local		0.00	0.07***	0.09***
		(0.01)	(0.01)	(0.01)
Satit		**1.20*** **	**0.79*** **	**0.75*** **
		(0.06)	(0.06)	(0.06)
Science		**0.96*** **	**0.58*** **	**0.59*** **
		(0.09)	(0.09)	(0.09)
Gender-female			0.02***	-0.02***
			(0.00)	(0.00)
Prior scores				
Grade 6 Math			**0.21*** **	**0.21*** **
			(0.00)	(0.00)

113 Four models were estimated for two subjects and two points in time. The results are similar, with slightly different magnitude. Hence, this chapter presents only results for one of the models.

	(1)	**(2)**	**(3)**	**(4)**
Grade 6 Science			0.13***	0.13***
			(0.00)	(0.00)
Grade 6 Thai			0.04***	0.04***
			(0.00)	(0.00)
School size (Reference: S)				
M				0.02***
				(0.01)
L				0.05***
				(0.01)
XL				**0.18*** **
				(0.01)
Urban area				0.02**
				(0.01)
Region (Reference: Lower NE)				
Bangkok and its perimeter				0.06***
				(0.01)
Central				-0.01
				(0.01)
Upper North				0.05***
				(0.01)
Lower North				0.04***
				(0.01)
Upper NE				0.04***
				(0.01)
South				0.01
				(0.01)
East				-0.03*
				(0.01)
West				0.00
				(0.02)

	(1)	(2)	(3)	(4)
ICC	17.24%	16.35%	17.70%	17.20%
Model fit				
Log-likelihood	-7434623	-743496	-695060	-694925
AIC	1487498	1487007	1390142	1389896
BIC	1487509	1487087	1390268	1390160
Observations	692,506	692,506	692,506	692,506
No. of groups	11,383	11,383	11,383	11,383

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using O-NET 2012 microdata

Overall, school type is shown to be one of the most significant predictors of test scores. In the model without covariates (2), being in Science and Satit school is associated with around one standard deviation higher in scores. The coefficients on school type change notably after more controls were added. In model (3), when prior scores were added, the model takes a value-added specification, and the effect of school type is conditional on different starting points of achievement. With this, the advantage of being in Science and Satit schools reduced by around 0.4 standard deviations. Nonetheless, even with prior scores, the top performers, Satit and Science schools still maintain large and significant advantage over public schools and other school types. This suggests that being in these schools is associated with better progress in addition to better scores. In the final model (4), school characteristics were added, but the school type coefficients do not change much.

The results imply that part of why these schools perform better than the rest can be attributed to more favourable student and school characteristics, especially prior scores. The ability to attract students of better academic background helps Satit and Science schools maintain advantage over public schools. Nonetheless, this does not tell the full story, as the gap remains large even after the controls were added. There are other factors that set Satit and Science schools apart from other schools not captured in the model. This could be factors such as teaching practices, parental support, or pupil motivation, etc. This aspect is explored further in the qualitative part of the research.

Another interesting result is that beyond Science and Satit schools, being in other school types does not have significantly different learning outcomes compared to being in OBEC2 schools. This implies that other than the two top-performers, the rest of the school types have uniformly low learning outcomes as measured by O-NET. This describes a different picture of learning problems in Thailand. From the descriptive analyses in the previous section, OBEC1, Local, and BMA schools have much lower learning outcomes than average levels (and OBEC2 schools). When analysing using multilevel models, these effects have decreased and are no longer meaningful by the 0.2 standard deviation threshold. This may imply that by accounting for the multilevel nature of the data, the effect of being in certain school types are overestimated/underestimated. The change in coefficient is similar to the ones described by O'Dwyer and Parker (2014). This reduction in coefficients has implications for when PISA data is analysed using OLS regression, as the coefficients may have been overestimated for low-performing schools.

The results with stepwise addition focus on how the coefficients changed over time. To answer another part of the sub-research question regarding what factors affect variations in scores, Table 5-2 shows combined results of the estimations in different years and subjects (without the stepwise addition).

Table 5-2 Multilevel model results comparison across subjects and years

	Mathematics		**Science**	
	2012	**2015**	**2012**	**2015**
Fixed Part				
Constant	-0.20***	0.35***	0.03**	0.36***
	(0.01)	(0.01)	(0.01)	(0.01)
School type (Reference: OBEC2)				
OBEC1	0.11***	0.15***	0.17***	0.18***
	(0.01)	(0.01)	(0.01)	(0.01)
Private	0.08***	0.12***	0.10***	0.07***
	(0.01)	(0.01)	(0.01)	(0.01)
BMA	-0.06*	-0.04	-0.08**	0.04
	(0.03)	(0.03)	(0.04)	(0.03)
Local	0.09***	0.01	0.15***	-0.04**
	(0.01)	(0.01)	(0.02)	(0.02)
Satit	**0.75*** **	**0.80*** **	**0.82*** **	**0.49*** **
	(0.06)	(0.06)	(0.07)	(0.06)
Science	**0.59*** **	**1.67*** **	**0.82*** **	**0.96*** **
	(0.09)	(0.09)	(0.11)	(0.09)
Gender-female	-0.02***	0.08***	0.02***	0.07***
	(0.00)	(0.00)	(0.00)	(0.00)
Prior scores				
Grade 6 Math	**0.21*** **	**0.44*** **	**0.29*** **	**0.33*** **
	(0.00)	(0.00)	(0.00)	(0.00)
Grade 6 Science	0.13***	**0.22*** **	**0.22*** **	**0.36*** **
	(0.00)	(0.00)	(0.00)	(0.00)
Grade 6 Thai	0.04***	0.13***	0.12***	0.20***
	(0.00)	(0.00)	(0.00)	(0.00)
School size (Reference: S)				
M	0.02***	0.00	0.02**	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)

	Mathematics		Science	
	2012	**2015**	**2012**	**2015**
L	0.05***	0.03***	0.09***	0.02
	(0.01)	(0.01)	(0.02)	(0.01)
XL	**0.18*** **	**0.23*** **	**0.28*** **	0.17***
	(0.01)	(0.01)	(0.02)	(0.02)
Urban area	0.02**	0.03***	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.00)
Region (Reference: Lower NE)				
Bangkok and its perimeter	0.06***	0.08***	0.05***	0.01
	(0.01)	(0.01)	(0.02)	(0.02)
Central	-0.01	0.05***	-0.03	-0.00
	(0.01)	(0.01)	(0.02)	(0.01)
Upper North	0.05***	0.12***	0.15***	0.13***
	(0.01)	(0.01)	(0.01)	(0.01)
Lower North	0.04***	0.06***	0.07***	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Upper NE	0.04***	0.05***	0.01	0.02**
	(0.01)	(0.01)	(0.01)	(0.01)
South	0.01	0.06***	0.02*	0.05***
	(0.01)	(0.01)	(0.01)	(0.01)
East	-0.03*	0.07***	-0.04**	0.03*
	(0.01)	(0.01)	(0.02)	(0.01)
West	0.00	0.10***	-0.00	0.06***
	(0.02)	(0.02)	(0.02)	(0.02)
ICC	17.20%	12.00%	22.06%	14.48%
Model fit				
Log-likelihood	-694925	-743315	-754412	-719497
AIC	1389896	1486677	1508870	1439039

	Mathematics		Science	
	2012	**2015**	**2012**	**2015**
BIC	1390160	1486938	1509133	1439300
Observations	692,506	628,302	692,506	628,302
No. of groups	11,383	11,413	11,383	11,413

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using O-NET microdata

Results show similarities in the direction and magnitude of coefficients across subjects and time frame. For school type, the effect is similar across subject and over time, except for a Science school dummy. Being in a Science school has significantly greater academic advantage in 2015 compared to in 2012, and in Mathematics more than Science. In 2012, Satit schools were the highest performer, with around 0.2 standard deviation score difference from Science schools. In 2015, however, Science schools topped the ranking.

The differential coefficient can be loosely attributed to changes in practices[114] that Science schools underwent during this time regarding school curriculum and admission process. Even though the schools were established during 1994 – 1996, it was not until 2011, that the schools made changes to become the Science schools we know today. They adopted the Science school curriculum and started admitting students using selective admission criteria (PCSHS, 2018). New students were admitted at the start of lower and upper secondary in the first year (grade seven and grade ten). This means that, for O-NET, grade nine students in 2012 were not selectively admitted, nor were they using specialised curriculum[115]. In 2015, grade nine students have been exposed to those changes. Consequently, change over time is higher, as it reflects this shift from public schools to Science schools[116]. Even though this cannot be implied as causal effects, the practices

[114] Adoption of new curricula and selective admission criteria
[115] At 2012, only grades 7, 8, 10, and 11 were taught Science school curriculum. Remaining students were taught the national curriculum
[116] PISA does not have the same problem as it tests mostly grade ten students, who by then would have already been exposed to the Science school curriculum.

contribute partially to the change in coefficients over time as students who took O-NET exams from Science schools in 2012 have not been exposed to the changes like students in 2015. This implies that changing admission policies and curriculum may help increase learning outcomes significantly. Nonetheless, this does not provide a straightforward link to policy. As these changes happened concurrently, it is not possible to isolate changes that are from changes in curriculum and pedagogy as opposed to changes in making schools more selective. Policies that improve academic rigour can be introduced to low-performing schools, but it is unclear how much difference that would make, given that they have much lower initial learning outcome to begin with. Further, making the school more selective cannot raise achievement levels for all schools. As students with higher learning levels move to a few schools, others are left with students with lower achievement.

Other variables: gender, region, and being in an urban area have smaller coefficients than school types and prior scores. Additionally, the size of the coefficients is negligible once the controls were included. It shows that these variables matter less in explaining variations in learning outcomes in the presence of school type. Region variable illustrates this point. In the descriptive statistics, schools in Bangkok and its perimeter shows better learning outcomes than others. However, the regions are not meaningfully different from one another in the regression. This is probably because school types are more predictive of scores, regardless of what regions the schools are in. BMA schools, for instance, are all in Bangkok, but are one of the low performers. Science schools, by contrast, exist in all regions beyond Bangkok.

By contrast, prior scores and school size are variables that are most predictive of scores (other than school type). There is a significant difference between small and extra-large schools. This is not surprising as literature shows larger schools tend to have better economy of scale and more budget, while smaller schools lack resources.

5.3.2. PISA data: OLS regression results

This section turns to look at estimation results using PISA data. Similar to O-NET, the models are presented stepwise. With (1) showing a model with school types as the only covariates, (2) adding student-level variables, and (3) adding school-level variables. For reasons discussed in section 4.3.3, multilevel models cannot be estimated while also accounting for PISA and TIMSS's weights and plausible values. Hence, the models are simple OLS regressions that account for weights and survey design considerations. Unlike O-NET, school type identifier is only available in 2012. With this, two models are estimated separately for Mathematics and Science and results presented in Table 5-3 and Table 5-4 (Appendix D2 shows estimation results for other subjects and years).

Table 5-3 OLS results for PISA 2012 Mathematics

	(1)	**(2)**	**(3)**
Constant	439.56***	400.23***	378.93***
	(4.97)	(5.84)	(10.97)
School type (Reference: OBEC2)			
OBEC1	17.57	**34.69***	**62.05*****
	(19.31)	(20.03)	(22.84)
Private	-18.98	**-22.47****	-12.64
	(14.09)	(10.28)	(8.66)
BMA	**-39.20*****	**-34.58*****	**-29.96****
	(13.83)	(12.78)	(13.65)
Local	**-33.52*****	**-25.59*****	-11.00
	(10.31)	(8.29)	(7.21)
Satit	**95.87*****	**64.05*****	**76.31*****
	(7.67)	(8.28)	(10.03)
Science	**130.48*****	**106.11*****	**142.89*****
	(5.79)	(6.30)	(8.02)

	(1)	**(2)**	**(3)**
Gender-female		5.86	0.61
		(3.75)	(3.13)
Parental education (Reference: Primary or lower)			
Lower secondary		10.01**	5.86
		(4.51)	(3.95)
Upper secondary or above		**28.61*** **	13.24***
		(4.34)	(3.62)
Books at home (Reference: 0-10 books)			
11-25 books		9.96**	8.61**
		(4.14)	(3.88)
26-100 books		**24.80*** **	**20.44*** **
		(4.42)	(3.95)
101-200 books		**39.46*** **	**31.99*** **
		(7.79)	(7.20)
More than 200 books		**71.78*** **	**55.49*** **
		(7.69)	(7.20)
Urban area			**17.81*** **
			(6.66)
School size (Reference: S)			
M			6.92
			(12.45)
L			**27.23** **
			(12.40)
XL			**58.38*** **
			(11.74)
R-squared	0.039	0.124	0.213
Observations	5,382	5,382	5,382

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using PISA 2012 microdata

Table 5-4 OLS results for PISA 2012 Science

	(1)	**(2)**	**(3)**
Constant	456.02***	416.13***	393.21***
	(4.14)	(6.02)	(8.62)
School type (Reference: OBEC2)			
OBEC1	17.29	**33.60****	**60.97*****
	(12.90)	(13.09)	(13.70)
Private	-16.72	**-19.48*****	-11.03
	(11.36)	(8.38)	(7.70)
BMA	**-41.39*****	**-35.87*****	**-32.14*****
	(10.86)	(9.94)	(12.11)
Local	**-29.28*****	**-21.93*****	-8.55
	(8.88)	(7.39)	(6.33)
Satit	**78.97*****	**52.49*****	**63.10*****
	(5.95)	(6.38)	(8.04)
Science	**109.40*****	**88.43*****	**120.70*****
	(5.10)	(5.37)	(6.53)
Gender-female		12.49***	7.41**
		(3.67)	(3.17)
Parental education (Reference: Primary or lower)			
Lower secondary		7.42*	3.74
		(4.17)	(3.70)
Upper secondary or above		**25.08*****	11.10***
		(3.99)	(3.37)
Books at home (Reference: 0-10 books)			
11-25 books		11.22***	9.96***
		(3.67)	(3.56)
26-100 books		**21.70*****	**17.72*****
		(4.28)	(3.91)
101-200 books		**41.69*****	**34.80*****

	(1)	**(2)**	**(3)**
		(6.83)	(6.43)
More than 200 books		**55.13*** **	**40.30*** **
		(6.50)	(6.07)
Urban area			15.67**
			(6.48)
School size (Reference: S)			
M			11.30
			(8.16)
L			**28.96*** **
			(8.91)
XL			**57.31*** **
			(8.46)
R-squared	0.035	0.114	0.204
Observations	5,382	5,382	5,382

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using PISA 2012 microdata

In model (1), being in Satit or Science schools shows meaningful advantage comparing to enrolling in general public schools. Students in these schools scored around one standard deviation higher in terms of learning outcomes. The advantage is greater in Mathematics than Science, and in Science schools than Satit schools. This result is similar to O-NET, albeit with slightly higher magnitude. Private and OBEC1 schools perform differently from the mean by more than our threshold. However, the difference is not statistically significant. Lastly, students in BMA and Local schools perform meaningfully worse than OBEC2 schools by about 0.4 standard deviation. These clash with the findings from O-NET, which show that these schools perform not meaningfully different from OBEC2 schools.

Part of the reasons why the magnitude and statistical significance is different for BMA and Local schools may be because the multilevel nature of data is not accounted for. We see from O-NET results that multilevel models give lower coefficients for the low-performers. It might be the case that OLS estimations slightly overestimate the effect of being in BMA and Local schools. Another potential reason for the discrepancy may be from differences in skills tested in PISA as compared to O-NET. As PISA tests higher-level skills such as problem-solving and critical thinking, it may be that there are even higher gaps between students in low- and high-performing schools.

In model (2), student-level variables were added. Controlling for parental education and number of books at home, the advantage of being in Satit and Science schools decreased by about 0.2 standard deviations. This implies that part of the advantage of these schools are from enrolling more well-off students. Yet, the advantage remains large and significant, implying that socioeconomic background, like prior scores, cannot fully explain why these schools perform much better than others. It implies that even though these schools do enrol students with better academic and socioeconomic background, they also add value beyond that. Interestingly, in model (3), when both student- and school-level variables were included, both Satit and Science schools' coefficients increased. Science schools show greater advantage comparing to the model without any other covariates. This is probably because most Science schools are located in rural area, and are not as big as other schools, yet perform well academically. Here, school types matter more than school characteristics in predicting variations in achievement.

Other than Satit and Science schools, OBEC1 is another school that show large changes in coefficients in different models. It started out having a positive but insignificant coefficient, then increased significantly with the addition of student and school variables in models (2) and (3). The advantage of being in an OBEC1 school, controlling for student background and school characteristics, is almost on par with being in Satit schools. OBEC1 schools have the least favourable conditions as they are small remote rural schools that enrol poorer students. The positive coefficients show that despite these unfavourable conditions, OBEC1 school students managed to perform quite well. However, it is unlikely that the results truly describe how OBEC1 schools are. This result is based on 2012 data, when OBEC1 schools perform unusually well (see Chapter Six). In 2015, the score of OBEC1 fell by almost one standard deviation, much higher than the average change of 0.3 standard deviation. With this decline, the school type performed much worse than OBEC2 schools. This outlier in the result implies that the 2012 increase in PISA scores were driven mainly from the students of poorer background and less favourable characteristics. It remains unclear why these groups of students improved so much in 2012 but not in other years.

Another low-performers, BMA schools, show consistent disadvantage even after the controls were added. After accounting for student and school characteristics, the disadvantage reduced slightly, but remains large. This suggests that other than enrolling more disadvantaged students, there are still some unexplained reasons why BMA schools lag behind. They might have to do with how the contents were taught or incentives in the school.

Other low-performers, Private and Local schools show increase in coefficients after student background was controlled for. However, they reduced and became insignificant after adding school characteristics. This implies that these schools do not perform differently from the mean once the controls were added. This finding suggests that private schools do not add much value in comparison with public schools in the Thai context.

Next, Table 5-5 presents results of regressions when school type is not included in Mathematics and Science. These are the final models, including both student and school level variables (Models with stepwise inclusion of variables are in Appendix D2).

Table 5-5 OLS results for PISA Mathematics (without school type)

	Mathematics		Science	
	2012	**2015**	**2012**	**2015**
Constant	401.91***	392.98***	416.20***	383.35***
	(10.49)	(8.05)	(9.66)	(5.72)
Gender-female	-0.13	-4.21	6.65**	0.93
	(3.02)	(3.22)	(3.11)	(2.53)
Parental education (Reference: Primary or lower)				
Lower secondary	4.12	-8.86**	2.09	-8.73***
	(4.50)	(4.45)	(4.15)	(3.25)
Upper secondary or above	11.40***	4.90	9.09**	6.21**
	(4.27)	(4.46)	(3.78)	(3.33)
Books at home (Reference: 0-10 books)				
11-25 books	8.08**	6.04	9.376***	9.48***
	(3.94)	(3.99)	(3.60)	(3.31)
26-100 books	**20.72*** **	**22.94*** **	**17.79*** **	**26.69*** **
	(4.12)	(4.69)	(4.06)	(3.53)
101-200 books	**33.53*** **	**25.39*** **	**35.95*** **	**29.16*** **
	(7.10)	(5.75)	(6.37)	(5.12)
More than 200 books	**58.42*** **	**55.90*** **	**42.56*** **	**43.37*** **
	(7.00)	(8.91)	(6.00)	(6.98)
Urban area	**17.30*** **	9.44	15.05**	15.19**
	(6.65)	(8.57)	(6.56)	(7.50)
School size (Reference: S)				
M	-11.51	-.14	-6.94	5.15
	(11.36)	(9.59)	(9.74)	(7.75)
L	4.22	7.18	6.49	**24.14*** **
	(12.76)	(9.25)	(10.53)	(7.36)

	Mathematics		Science	
	2012	**2015**	**2012**	**2015**
XL	**36.36*** **	**41.07*** **	**35.79*** **	**53.66*** **
	(11.08)	(9.42)	(9.96)	(7.49)
R-squared	0.159	0.124	0.148	0.171
Observations	5,382	6,783	5,382	6,783

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using PISA microdata

Results are similar across subjects and over time. The coefficients in the same years are very similar, with only slight differences in magnitudes, aside from a few exceptions. For example, being in a large school gives students more academic advantage in Science in 2015 but not in other years. Across years, the magnitudes show more differences, but nonetheless are in the same general direction.

Overall, number of books at home is the strongest predictor of variation in scores. Students with more books at home perform significantly better than those with a smaller number of books at home. This supports the findings from the previous tables that socioeconomic measures drive differences in scores meaningfully. School size also shows significant coefficients, with students in extra-large schools performing better than small school students. This shows the same results as O-NET. Unlike in the previous tables, the difference is only significant for extra-large schools. Students in medium and large schools do not perform differently from students in small schools (except for Science 2015).

By contrast, gender and level of parental education does not seem to have an impact on achievement level. While it is possible to dismiss that parental education is not a good proxy for socioeconomic status, another alternative explanation is that the way parental education is asked does not capture socioeconomic status very well in the context of Thailand. In PISA, the highest parental education option to choose from is if the parents graduate upper secondary or above. However, policies supporting universal access to education has made upper secondary level of education less selective over time. As of 2016, 91% of grade 9 graduates continue to upper secondary (NESDC, 2020). Nonetheless, university level education is more selective, with only 56% of upper secondary graduates starting year one of bachelor's degree. Hence, someone graduating lower secondary may not be very different from someone who graduates upper secondary, while it is much more prestigious to attend and graduate a university. With this, capping the answer to upper secondary may be ignoring this distinction, making the differences not significant. Data from TIMSS support this, showing that students whose parents graduated bachelor's degree scored 66 points higher than students with parents who graduated upper secondary graduates and 73 points higher than students whose parents did not go to school. This implies that for the tests to be useful in capturing effect of socioeconomic background, these categories should be better designed.

When school types are not included, R-squared is significantly less compared to results in Table 5-3 and Table 5-4. This confirms that school type is a significant predictor of test scores, and not including them may lead to a misspecification. However, this does not lead to widely different inferences. Most coefficients in 2012 are slightly smaller than when school type is in the model, but are similar in magnitudes, directions, and statistical significance. One variable that shows differences in coefficient is school size. In the model without school type, the coefficients are significantly less than when school types are included. With school type, the advantage of being in an extra-large school are much higher. There is also a significant advantage of being in a large school. Without school type, the effect of being in a large school compared to being in a small school is negligible. This is because there are intersections in the effect of school type and size. When controlling for the type of school, being in larger schools gives much more academic advantage. This implies that we may be underestimating the effect of school characteristics when school type is not included.

5.3.3. TIMSS data: OLS regression results

Lastly, OLS results are estimated using TIMSS data. TIMSS is the most restrictive dataset out of the three. There is no school type identifier in the data and school size is only available in 2011. Hence, school type advantage cannot be estimated like in O-NET and PISA. The results should be interpreted with caution as the models are excluding one of the major sources of variation in student performance. Nonetheless, there remains comparative value in estimating the model using TIMSS. As socioeconomic variables in TIMSS are very similar to PISA, regressions can help with understanding if socioeconomic variables have similar impact on TIMSS scores compared to PISA. With this, three models are estimated. Model (1) have only student control variables, model (2) adds in school control variables. In 2011, school size is included in model (3). Table 5-6 and Table 5-7 show TIMSS OLS results for Mathematics and Science respectively[117].

[117] The models are estimated with missing data imputations. Results without this imputation is shown in Appendix D3. Additionally, multilevel models have also been estimated using PISA and TIMSS data (through PV command in Stata). Results are in Appendix D4.

Table 5-6 OLS results for TIMSS Mathematics

	2011			2015	
	(1)	**(2)**	**(3)**	**(1)**	**(2)**
Constant	398.68***	396.87***	385.60***	401.87***	397.13***
	(1.60)	(1.72)	(2.10)	(8.43)	(7.96)
Gender-female	16.01***	14.76***	15.01***	13.83*	13.49***
	(0.35)	(0.36)	(0.34)	(5.06)	(4.37)
Parental education (Reference: Primary or lower)					
Lower secondary	-14.00***	-14.29***	-15.71***	-18.46**	**-19.03*** **
	(1.98)	(1.95)	(1.91)	(7.37)	(6.63)
Upper secondary or above	12.59***	9.39**	2.40	10.72**	3.82
	(3.55)	(3.49)	(2.87)	(7.87)	(6.79)
Books at home (Reference: 0-10 books)					
11-25 books	14.40***	13.40***	9.64**	**19.02*** **	16.83***
	(1.80)	(1.77)	(1.81)	(4.26)	(4.16)
26-100 books	**40.61*** **	**38.58*** **	**31.11*** **	**49.96*** **	**41.65*** **
	(2.38)	(2.34)	(2.44)	(6.15)	(6.02)
101-200 books	**72.92*** **	**68.82*** **	**59.57*** **	**84.25*** **	**70.79*** **
	(3.05)	(3.03)	(3.00)	(10.27)	(9.92)

	2011			2015	
	(1)	(2)	(3)	(1)	(2)
More than 200 books	**84.84*** **	**80.99*** **	**65.78*** **	**94.97*** **	**79.59*** **
	(3.21)	(3.33)	(3.35)	(17.31)	(13.59)
Urban area		**23.35*** **	9.61***		**51.11*** **
		(1.73)	(1.56)		(10.55)
School size (Reference: S)					
M			8.79**		
			(1.21)		
L			**29.42*** **		
			(0.98)		
XL			**51.43*** **		
			(1.05)		
R-squared	0.11	0.12	0.16	0.13	0.18
Observations	5,967	5,967	5,967	6,465	6,465

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using TIMSS microdata

Table 5-7 OLS results for TIMSS Science

	2011			2015	
	(1)	**(2)**	**(3)**	**(1)**	**(2)**
Constant	423.91***	422.29***	411.67***	425.37***	421.28***
	(2.29)	(2.41)	(2.72)	(8.58)	(8.59)
Gender-female	13.03***	11.92***	12.02***	16.17***	15.88***
	(1.47)	(1.48)	(1.46)	(4.61)	(4.11)
Parental education (Reference: Primary or lower)					
Lower secondary	-16.44***	-16.70***	-17.99***	-16.57**	-17.06***
	(1.99)	(1.92)	(1.95)	(6.73)	(6.44)
Upper secondary or above	9.68***	6.81**	0.60	9.62	3.65
	(3.39)	(3.30)	(2.81)	(7.13)	(6.67)
Books at home (Reference: 0-10 books)					
11-25 books	**20.86*****	**19.97*****	16.48***	18.99***	17.09***
	(1.40)	(1.34)	(1.30)	(3.93)	(3.96)
26-100 books	**40.76*****	**38.94*****	**32.20*****	**49.37*****	**42.19*****
	(1.42)	(1.32)	(1.20)	(5.73)	(5.71)
101-200 books	**60.95*****	**57.28*****	**46.36*****	**75.95*****	**64.32*****

	2011			2015	
	(1)	**(2)**	**(3)**	**(1)**	**(2)**
	(2.26)	(2.28)	(2.37)	(8.02)	(8.08)
More than 200 books	**82.18*** **	**78.73*** **	**65.47*** **	**88.14*** **	**74.86*** **
	(1.73)	(1.86)	(1.88)	(12.59)	(10.16)
Urban area		**20.91*** **	9.19***		**44.14*** **
		(1.87)	(1.63)		(8.65)
School size (Reference: S)					
M			8.81***		
			(1.56)		
L			**28.52*** **		
			(1.12)		
XL			**44.92*** **		
			(1.14)		
R-squared	0.10	0.11	0.15	0.14	0.19
Observations	5,967	5,967	5,967	6,465	6,465

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using TIMSS microdata

Similar to PISA, having more books at home is associated with scoring higher in TIMSS. The magnitude is much larger in TIMSS. This may be an overestimation of results as the models in TIMSS include much less variables than those in PISA and O-NET. Parental education, however, shows a different pattern. Students whose parents graduated lower secondary level have lower scores than those whose parents finished primary or having no schooling at all. The gap is also around 0.2 standard deviation. Also similar to PISA, being in a larger school is predictive of scoring higher. Being in urban school does not have a significant improvement compared to being in a rural school in 2011. The gap is much larger in 2015.

In model (3), when school size is included, the effect of number of books at home and being in an urban area is reduced. This shows that school size can explain part of the advantage of being in urban areas. It also stresses the importance of including these school-level variables that are predictive of achievement.

5.3.4. Comparing results between PISA, TIMSS, and O-NET

This section combines the results of different assessments in the same table (Table 5-8) for ease of comparison The year 2012 (or 2011 in TIMSS) is taken as a basis for comparison as it is the year that school type data is available on PISA. Additionally, data is standardised to have a mean of zero and standard deviation of one, so that results across assessments can be compared. Results answer the sub-research question of whether there are differences in the results when different data is used.

Table 5-8 Comparison of the coefficients using 2012 data

	Mathematics			Science		
	O-NET	**PISA**	**TIMSS**	**O-NET**	**PISA**	**TIMSS**
Constant	-0.18***	-0.61***	-0.62***	0.06***	-0.58***	-0.64***
	(0.01)	(0.12)	(0.02)	(0.02)	(0.10)	(.03)
School type (Reference: OBEC2)						
OBEC1	0.11***	**0.68*** **		0.17***	**0.70*** **	
	(0.01)	(0.25)		(0.01)	(0.16)	
Private	0.08***	-0.14		0.10***	-0.13	
	(0.01)	(0.10)		(0.01)	(0.09)	
BMA	-0.05	**-0.33** **		-0.08**	**-0.37*** **	
	(0.03)	(0.15)		(0.04)	(0.14)	
Local	0.10***	-0.12		0.15***	-0.10	
	(0.01)	(0.08)		(0.02)	(0.07)	
Satit	**0.78*** **	**0.84*** **		**0.82*** **	**0.73*** **	
	(0.06)	(0.11)		(0.08)	(0.09)	
Science	**0.61*** **	**1.57*** **		**0.83*** **	**1.39*** **	
	(0.09)	(0.09)		(0.11)	(0.07)	
Gender-female	-0.02***	0.01	0.16***	0.02***	0.09**	0.14***
	(0.00)	(0.03)	(0.00)	(0.00)	(0.04)	(0.02)
Prior scores						
Grade 6 Mathematics score	**0.21*** **			**0.30*** **		

	Mathematics			Science		
	O-NET	**PISA**	**TIMSS**	**O-NET**	**PISA**	**TIMSS**
	(0.00)			(0.00)		
Grade 6 Science score	0.13***			**0.22*** **		
	(0.00)			(0.00)		
Grade 6 Thai score	0.05***			0.12***		
	(0.00)			(0.00)		
Parental education (Reference: Primary or lower)						
Lower secondary		0.06	-0.16***		0.04	-0.21***
		(0.04)	(0.02)		(0.04)	(.02)
Upper secondary or above		0.15***	0.02		0.13***	0.01
		(0.04)	(0.03)		(0.04)	(0.03)
Books at home (Reference: 0-10 books)						
11-25 books		0.09**	0.10**		0.11***	0.19***
		(0.04)	(0.02)		(0.04)	(0.02)
26-100 books		**0.22*** **	**0.32*** **		**0.20*** **	**0.37*** **
		(0.04)	(0.02)		(0.05)	(0.01)
101-200 books		**0.35*** **	**0.59*** **		**0.40*** **	**0.54*** **
		(0.08)	(0.03)		(0.07)	(0.03)
More than 200 books		**0.61*** **	**0.68*** **		**0.46*** **	**0.76*** **
		(0.08)	(0.04)		(0.07)	(0.02)

	Mathematics			Science		
	O-NET	**PISA**	**TIMSS**	**O-NET**	**PISA**	**TIMSS**
School size (Reference: S)						
M	0.02***	0.08	0.09**	0.02**	0.13	0.10***
	(0.01)	(0.14)	(0.01)	(0.01)	(0.09)	(0.02)
L	0.06***	**0.30****	**0.31*****	0.09***	**0.33*****	**0.33*****
	(0.01)	(0.14)	(0.01)	(0.02)	(0.10)	(0.01)
XL	**0.19*****	**0.64*****	**0.53*****	**0.28*****	**0.66*****	**0.52*****
	(0.01)	(0.13)	(0.01)	(0.02)	(0.10)	(0.01)
Urban area	0.00	**0.20*****	0.10***	0.01	0.18**	0.11***
	(0.01)	(0.07)	(.02)	(0.01)	(0.07)	(0.02)
Region (Reference: Lower NE)						
Bangkok and its perimeter	0.07***			0.05***		
	(0.01)			(0.02)		
Central	-0.00			-0.03*		
	(0.01)			(0.02)		
Upper North	0.05***			0.15***		
	(0.01)			(0.01)		
Lower North	0.04***			0.07***		
	(0.01)			(0.01)		
Upper NE	0.04***			0.01		
	(0.01)			(0.01)		

	Mathematics			**Science**		
	O-NET	**PISA**	**TIMSS**	**O-NET**	**PISA**	**TIMSS**
South	0.01			0.02*		
	(0.01)			(0.01)		
East	-0.03*			-0.04**		
	(0.01)			(0.02)		
West	0.00			-0.00		
	(0.02)			(0.02)		
Observations	692,506	5,382	5,967	692,506	5,382	5,967

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using PISA, TIMSS, and O-NET 2012 microdata

General results are similar among the assessments. Satit and Science schools show significant advantage over OBEC2 schools. The magnitude is similar for Satit schools across all three assessments, however, have much higher coefficients in PISA than in O-NET. This is probably driven by the fact that the students taking O-NET in 2012 have not been exposed to the Science curriculum, as previously discussed. Results in 2015 show similar magnitude with PISA. These similarities support that the three data sources are comparable.

Both prior scores and socioeconomic variables help explain why some school types perform better than others. The coefficient of Satit and Science schools greatly reduced with the inclusion of these variables. For PISA and TIMSS, parental education shows similar insignificant coefficients. Books at home results are also similar, but with TIMSS having slightly higher magnitude. It is speculated that this is from not including school type variable in TIMSS. As the advantage remains large even after these variables were added, there are differences beyond these measurable characteristics that enable these schools to score well in all assessments. This aspect is explored in the qualitative part.

5.4. Qualitative results: What could explain score gaps among high- and low-performers?

This section draws on evidence from interviews with teachers and vice principals to shed light on differences not included in the quantitative models. It presents how the context differs for different school types, such as student characteristics, autonomy and accountability, and how that in turn affects practice and how schools prepare for examinations.

5.4.1. Differences in student and school characteristics

Quantitative findings show student and school characteristics that affect learning outcomes such as prior scores and socioeconomic status. This section dives further into mechanisms of how these factors influence teaching practice and learning outcomes.

Prior test scores show the extent in which students are proficient in foundational knowledge and contents from primary level. Almost all low-performing schools in the sample mentioned that some proportion of students in their schools cannot read or write fluently, as well as doing simple calculation. Most students are also not proficient in contents taught in primary levels. This makes it more difficult for teachers to teach contents that build on previous knowledge. It is also difficult for students to be motivated to learn those new topics when they do not have prior knowledge. For teachers, this leads them to having very low expectations on what students can and cannot do, which in turn leads to them teaching easier contents to match what they expect students can do, as illustrated by one teacher.

> *For me, I accept it already, instead of having [students] do 5 exercises, I have them do 1 only. I previously expected that we can do 5 questions today, then I reduced, ok, let's do 1 question instead, but in more detail, teaching all the basics from adding fractions (Academic teacher, BMA school).*

This creates a vicious cycle of low performance. As teachers believe the students cannot learn more difficult topics, they make their classes easier. Students, in turn, do not learn enough to be able to pass examinations, and the results remain low. Additionally, they put in less effort in doing the assessments. When faced with difficult questions, teachers noted students tend to give up and are not motivated to work through the problems, both in classrooms and in assessments.

> *When [students] see difficult questions, questions that require lots of reading, critical thinking questions, they give up, they won't solve it, they would just guess (Academic teacher, OBEC1 school).*

> *[Students] don't see [exams] as important, not at all. When I brought in the exam papers, they wrote their names and slept immediately. And they put in the same answer for the whole exam. Sometimes there are 20 questions, they filled in the answer sheet for 25-30 questions! Some questions require them to pick two answers. They don't even read the questions sometimes! They just sleep in the room (Mathematics teacher, BMA school).*

Unpacking and addressing each element can help improve learning. Research shows that high expectation from teachers is related to student learning outcomes as seen in Baars et al. (2018). In disadvantaged schools that performs well, teachers hold high expectation on pupil learning outcomes. Hence, shifting teachers' mindsets can play an important role in improving learning. Equally, if learning improves, teachers' mindset can change.

Weak foundational knowledge is not a big problem in high-performing schools, as most (if not all) students must pass examinations to gain admission into the schools. Students are motivated to perform well and prepare for assessments without being told to. This results in them performing well in O-NET.

> *The students here are enthusiastic and responsible. It's good. Most of them are like this. This means when there are O-NET exams, they gave it importance, and do it. Even though some didn't prepare fully, but they know, within themselves that when they are going to take an exam, they need to do their best based on their capabilities. After the exam, when I talked to them, even though we didn't have O-NET tutoring at all, actually we don't have tutoring for quite a while in lower secondary, same with upper secondary, the kids will say, they can do this, like they studied the contents, and they can do it (Science teacher, Satit school).*

Other than prior scores, students' socioeconomic status affect their chosen pathways after finishing compulsory education. Many students from low-income background (80% in OBEC1 school) transition into work or vocational schools after graduation. They perceive these pathways as more direct route to employment and income generation compared to attending universities. As vocational schools are relatively easier to get into than universities[118], students focus less on performing well on assessments. O-NET examinations are very low stake for them, and it makes sense that they are not motivated to try hard on the exam.

> *The school has quota for vocational and technical schools, so many of them. We never got quotas for (general track) secondary schools. Hence, [students] would pick already which institutions they would like to go, the one that they can get in without taking an exam. … Some [vocational] places allow students to study for free (Academic teacher, OBEC1 school).*

Aggregation of students of poorer background and lower prior achievement in low-performing schools is partly a product of self-selection. Parents choose to enrol their children in schools with better reputation and academic outcomes. Another effect of this is that low-performing schools have problems retaining students who perform well academically. As top students transition to more prestigious school, low and average performing schools ended up with students of poorer performance (in relation to top schools). This happens both as primary students transition to lower secondary and from lower to upper secondary. When this happens, they lost their best students. This makes it more difficult for the schools then to score higher.

> *For us, when students started grade 10, they are gone, moved to [famous school] and such. The smart, the gifted ones will be gone. It's the problem of the whole country. All the students are stolen by these schools (Academic teacher, Private school (school 6)).*

---

[118] Most students do not need examinations, and schools often have many quotas for free admission to vocational schools.

This movement of students with high achievement leads to an aggregation of top students in a few schools, while leaving the rest with average and low performing students. This contributes to high level of inequality in learning outcomes among schools.

In addition to student characteristics, quantitative results show that low-performing schools have less resources compared to the high-performers. In smaller schools, one of the most important resources lacking are subject-specialised teachers. For OBEC1 school, it is common for teachers to teach multiple subjects. The two Mathematics and Science teachers interviewed reported having had to teach other subjects including Health, Physical Education, History, and Arts over the years due to lack of teachers. Furthermore, they are the only teachers with Mathematics and Science specialisation in the school.

> *It's normal, because we don't have that much capacity, and I am the only Mathematics teacher in the school. In the school, there is only one person with Math specialisation. Previously there were two, but she passed away, so there is only one left, and we don't get additional transfer who specialise in Math. If it's a big secondary school, each teacher teaches his or her own specialisation, but schools like primary schools or opportunity expansion schools like this, you need to teach multiple subjects (Mathematics teacher, OBEC1 school).*

When there is only one specialised teacher per subject, the school is left vulnerable of not having Mathematics or Science teacher when one fell ill, retired, or took leave. Lacking core subject teachers is one of the reasons why smaller schools perform worse than larger schools.

5.4.2. Top-down hierarchical structure of governance hinders achievement

Beyond student and school characteristics, structure of governance can affect learning gaps as practitioners work within the system they are in. When schools have different levels of autonomy and accountability, they vary in practices and eventually learning outcomes.

Thailand has a rigid hierarchical education system[119] with the Ministry designing main guidelines of education policies. While there is scope for the devolved authorities and schools to adapt the policies to suit their contexts, school policies need to align well with the national guidelines. This is true particularly for schools with low to average achievement level (OBEC1, OBEC2, Local, and BMA schools). In these schools, national policies are translated to school-level policies via the local authorities. Schools reported having to implement what was set by the local authorities, rather than having autonomy to set one themselves, as one participant illustrated

> *The principal receives[120] the policy from the ESA. If the ESA sets the goal, like 3% (increase in scores), then the principal sets 3% score increase (Academic teacher, OBEC2 school (school 3))"*.

Hence, these schools seem to have lower level of autonomy in designing school policies. Most schools have to focus on academic aspects and strive towards achieving higher learning outcomes using teaching and tutoring methods decided by the Ministry and devolved authorities.

An advantage of this system is that what is taught and how they are taught is standardised across schools. However, a disadvantage of such a centralised system is that there is little space for schools or local authorities to innovate or try out new practices within the system. This may lead to rigid and outdated practice that contributes to lower achievement. Schools may not be able to contextualise contents to match that of their community needs. It can also create tension between schools wanting to try different things but are limited by policies set by the local authorities.

---

119 Discussed in section 2.2.5.

120 Participants use the word 'receive' in a sense that the policies have been given to them by the higher-ups.

There are exceptions to the current system. Science schools similarly work with a school group committee (local authority). However, as they are special-purpose schools, there is legal space for them to design their own goals and practices that deviate from the national curriculum. For example, they have examinations that are designed at a group level, to be used by all schools in the group. Another group that are less affected by the centralised policies are Satit schools and private schools. The policies are almost exclusively set by principals or the academic department within the school. This allows them more flexibility and autonomy to design teaching, which may partially explain why they perform better than public schools.

One outlier school in the sample (school 7 – Local) managed to break away from the rigid structure by lobbying with the local authorities. The school shifted from heavy emphasis on O-NET to teaching higher-order skills as well as focusing on occupation-related skills.

> *At first, [the local authority] is not ok (with the school not prioritising O-NET), they asked, there are 42 secondary schools in the province, what is our ranking? We got placed 11th-14th, the top school in the province is Science school, where they only have one programme, Science-Math, and they select students from all 17 provinces in the Northern region. I have a responsibility to explain that to our local authority, that at the Science school, there are less students entering teacher training programme, less soldiers, they don't have any national footballers, they don't have any students who earn salary during study, but they ranked first in O-NET because they are all Science-Math programme[121]. Later, they understand that students are different (Principal, Local school).*

This sets the school apart from other Local schools, which may not have the power nor resources to negotiate with the local authorities. This is consistent with many literature showing that school leadership is vital in leading change (Creemers &

[121] Implying that as emphasis has been given to studying Science and Mathematics, students performed well in those subjects.

Kyriakides, 2008; Hallinger & Lee, 2011). It also shows that the system does not reward innovative practices or allow different focus of learning achievement. To do something different, the school has to find its own leeway. With this, the system can be reinforcing outdated practices that are no longer relevant in today's world. Schools that want to change cannot do so easily within the system.

Another limiting factor is budget. Initiatives at school-level remain rare as the schools have to seek funding by themselves. This then depends on schools' own resources and network. Here, a teacher gave an example that principals can have initiatives at a school level, as long as it does not affect the communities negatively and does not use additional budget.

> *Yes, principal can do something not mandated by the BEO, but it has to have no effect on the budget and the community, I mean negatively, if it's positive, it can be done. But he would have to propose the idea to teachers, communities, and education committee. If the committee agrees, then it can be done. It has to have no effect on the main policy too. It has to support [main policies] and have no negative effects (Academic teacher, BMA school).*

5.4.3. Schools experience different pressure to perform on assessments

Another potential source of gap in learning outcomes lie in how schools are pressured to perform well on assessments. While schools are exposed to similar structures for school and teacher evaluation (see section 2.3.4), they still experience different pressure depending on where they are in terms of achievement levels, as well as school type. This affects how schools teach and prepare for assessments, which helps explain why some schools perform much better than others, as well as how they are able to maintain the advantage.

5.4.3.1. Unattainable evaluation criteria perpetuate learning gap

On paper, quality assurance[122] allows the Ministry to ensure schools have adequate teaching quality, and to identify schools that are not performing up to standards. In practice, the system is not conducive to helping schools improve learning outcomes, as the evaluation criteria are the same across the low- and high-performers

All evaluations have similar learning outcomes criteria. That is, schools need to show that student learning outcomes met the set goals, as well as showing improvement every year (ONESQA, 2017). Learning outcomes are usually measured by school examinations and national examinations of O-NET. The benchmark set by most schools is the national average of O-NET. Success is defined by schools scoring above the national average or have certain number of students scoring above the average[123].

It is an unrealistic expectation for all schools to be able to score above the national average, let alone showing improvement every year. As some schools perform above average, others will perform below average. Rewarding schools based on this is problematic for schools at all achievement levels. For schools such as Satit and Science schools, the goal is too low. These high-performers consistently perform above national average and the goal does not challenge them to improve. For low-performers, this goal is challenging as they have students with less favourable characteristics. Many of the low-performing schools are accepting that they cannot score above average in O-NET, nor can improvement be shown consistently as it depends on factors outside their control. Hence, the goal does not lead to improved teaching.

---

[122] Schools and education personnel are evaluated through quality assurances (QA). External QAs are by ONESQA, a public organisation, every four years. Internal QAs are done by local authorities (or schools themselves when no local authorities are present) every year. In addition to the quality assurance, teachers are evaluated for salary and career progression.
[123] The exact goals (such as percentage increase in score) depend on schools and local authorities to define.

> *We cannot [show improvement] (laughs), because [scores] cannot be predicted, and you can't compare previous year and this year's scores because the people who take the exam are different people. The people who took the exam last year already graduated, and for this year, you want to make the scores this year higher than last year, it can't be done (Mathematics teacher, OBEC1 school).*

What these schools did instead was to try to improve scores that they can control, which are learning outcomes at a school level. As these outcomes are solely within teachers' discretion (they design and grade exams themselves), there is a potential for manipulation. By focusing on this, schools can demonstrate improvement as well as showing students passing standards. However, many schools do this in such a way that is detrimental to learning. Some schools pre-emptively provided additional tutorial and coursework for students who might fail. Some allowed students to regrade. This means that students are allowed to retake an exam or resend coursework with results capped at passing grade. Some, like OBEC1 school, just gave students a pass to satisfy the criteria.

> *For this year, the rate at which students fail the subject must decrease, … it's not students, but it's the responsibility of the teachers. Assuming we see the students almost getting a fail grade, we need to have strategies or ways, like do whatever (to pass), maybe extra tutorial or extra coursework, to lower the percentage of students failing the subject (Academic teacher, OBEC2 school (school 3)).*

> *The students cannot fail a subject. And the teachers that fail their students need to write explanations to the principal. In the end, you need to pass the students even when they did not put in effort (Science teacher, OBEC1 school).*

Additionally, the examinations can be made easier so that students pass the criteria.

> *My exams previously, I designed by myself, all of them. But it's all multiple choices. Because at that time, I think, I tried to have some writing questions, but the students cannot answer, so I solve the problem by making it all multiple choices (Academic teacher, OBEC2 school (school 3)).*

Both regrading and making exams easier affect what students learn negatively. This shows that this way of incentivising schools has many drawbacks and may have contributed to the persistent learning gaps. Especially so as the high-performing schools have very different practices on internal examinations. There are checks and balances in school examinations. Science school has exams that were designed and used as a school group. It means that teachers cannot individually make exams easier to artificially inflate scores as done in school 3. Satit and private school teachers co-teach a subject as well as co-design the midterm and final exams. As teachers co-design the exams, it would be more difficult to control students' scores. With these additional standards in place, teachers cannot game the system by making exams easier[124].

> *We can use the same exam because when we prepare for lessons, we always talk to each other, and we use the same worksheets. The teaching method and activities, all the same. The teachers would talk to each other first (Science teacher, Private school (school 5)).*

Another way that the low performers react to the goals is to prepare vigorously for the actual inspection and bring up the average result with non-academic criteria. With this, even with low academic performance, they can score well on the evaluation. For low-performing schools, the participants take the inspection very seriously that teaching time is compromised, both in paperwork preparation and on actual day of the in-person inspection. In a BMA school, for example, students do not study at all during inspection periods as the teachers are busy preparing.

[124] For some schools, co-teaching is simply a by-product of being large schools. School 6 reported having 15 classes of one grade, and it is impossible to have one teacher responsible for all classes.

This has a negative effect on learning outcomes, as well as creating more gaps in results when high-performing schools do not engage in such practice.

> *When [the inspectors] come, it's really busy, there needs to be a teacher to greet them and prepare for food. In the end, the students don't get to study. The big ONESQA evaluation took 3 days. It means that for those 3 days, the kids don't get to study. And it's really busy for the teachers. They request for lots of things like paperwork (Science teacher, BMA school).*

Additionally, some schools went so far as making up evidence solely for the inspection. Teachers show that it is very serious that the school got good outcomes from ONESQA.

> *If ONESQA comes to inspect, we need to get Excellent level, at least Good level. Everyone will see if (this indicator) has a score yet, if not, we have to go back and make up evidence, so that we have it (Mathematics teacher, BMA school).*

> *Nowadays it's all fake. They are coming to inspect the school and we have to arrange it beautifully. What we don't have, we have to find (Mathematics teacher, OBEC2 school (school 3)).*

When evidence is made up to please the inspectors, the schools are not striving towards showing real improvement to school quality. Further, as schools try to sweep the problems under the rug and appear as if they have good quality teaching, the inspectors cannot give them constructive feedback to improve. This can explain why low-performers struggle to improve and keep up with the high-performers.

From the perspective of the high-performers, there is no pressure to improve or add value. For virtually all high-performing schools, external quality assurance poses no pressure to the schools. The schools perform well on all indicators and would have passed anyway. Particularly in academic performance goals, these schools reported passing the criteria regardless of level of preparation.

> *But we don't have any problems, our school, we passed. The teachers at the school has no concern over students' learning outcomes at all, because we believe that most Satit students are smart and prepared, parents also support them. They have a certain level of smartness. And ONESQA doesn't set goals too high that it's impossible. Their criteria are average, for students of the whole country. Hence, we don't really have any worries (Mathematics teacher, Satit school).*

Therefore, there is less incentive for high-performing schools to engage in extensive preparation compared to the sampled low-performing schools. They can instead focus on teaching and maintain their level of learning outcomes. While this is good, by giving them no pressure, there is no incentive for improvement. These schools do not have to strive to improve learning outcomes as much as their standards are already good enough for the criteria.

5.4.3.2. Exam preparation strategies harm actual learning

With pressure to perform well on O-NET, schools engage in various preparation strategies. The strategies differ among schools. Low-performing schools engage in strategies that disrupt learning more than higher-performing schools. While all schools tutor their students before exams to a certain extent (content revision and practicing exams together), low-performing schools do this much more extensively than high-performing schools. The schools spend weeks to months tutoring[125]. To make time for tutoring, schools make modifications to teaching schedule. Subjects that are not in the examinations including boy/girl scouts, clubs, or guidance subjects were cut out in the second semester. While it is important that students learn the contents that will be tested in O-NET, other subjects that teach students life skills or help students decide their future (guidance) are arguably very

[125] OBEC1, OBEC2 (school 3), and BMA school start their tutoring three months before the exam. OBEC2 (school 2) started tutoring intensively around one week before the exam but spent whole day tutoring.

important too. Because these subjects are not tested, they were deprioritised. Additionally, the tested subjects (Thai, English, Science, and Mathematics) also have their contents condensed so that all contents are finished within the first semester of the year. Consequently, class time in the second semester can be wholly dedicated to tutoring.

> *In the first semester, [the school] has the agreement on building foundational knowledge for students, and speed up the contents, to have as much of the second semester's contents be taught in the first semester, so that in the second semester before O-NET exam, we can spend time in …, like in tutoring, teach everything, guessing techniques, bringing in exams for students to practice, tutor by questions, we do all these (Academic teacher, OBEC1 school).*

By condensing the contents of two semesters to finish in the one semester, it is inevitable that teaching quality suffers. Instead of being able to spend time exploring each content, teachers ended up giving lectures on the topics included in O-NET and have students conduct self-study for the rest of the contents. Additionally, for the whole second semester, students only learn how to do well in O-NET. Even if they do become proficient in doing the exam, they are not learning real skills. The knowledge might not carry over to other tests as O-NET emphasises heavily on memorisation. This can explain why they are not performing well in other tests like PISA and TIMSS.

This is in contrast to preparation in high-performing schools, where tutoring is generally optional and occurs outside class time. Actual teaching time is not disrupted to make time for tutoring. Arguably, for the high-performing schools, they are in a much more privileged position than other schools. There is less need for tutoring as students can score well on O-NET already regardless of whether the school tutors them or not.

Even more detrimental to learning, to improve the average scores, teachers chose to focus the contents that are easy. Exam papers were analysed, and easier topics were picked out and emphasised to students. Topics that are deemed difficult were deprioritised, as the teachers perceived that students cannot score points in O-NET exam on those topics anyway. This means that students do not learn the contents according to the curriculum. Teachers only pick out certain topics to teach to students.

> *If I cannot teach all contents with the available time, I will summarise topics that [students] can easily learn. For the difficult topics, it takes time. So, I emphasise the topics that they can learn fast, go fast, and support with exercises and give them answers, and let them review later (Science teacher, OBEC1 school).*

Teachers sometimes also have different strategies for students of different abilities. Students that are "smarter" got intensive preparation so that they would do well on O-NET exams. By contrast, students perceived as "weaker" academically got lower expectations and taught differently. One teacher opted to teach weaker students guessing techniques instead of actual contents. This can be quite detrimental to student learning as they do not learn the same contents as their peers.

> *The strategies are a bit different. For O-NET, I would combine smart and average students together, so that the smart ones can learn very fast, and average students, I try to bring them up a step. But the weak students, if they are really weak, I would teach them guessing techniques. These students are very good at guessing (Academic teacher, OBEC1 school).*

Teaching students with different abilities in using different strategies is common and can be done without disrupting student learning. The strategy done in higher-performing schools is quite different from those previously described. Students are taught differently. However, teachers emphasised that all students still need to master the basic required level of proficiency. Students with higher ability are given more difficult exercises to complete.

> *For the weak students, we might have to emphasise knowledge and memorisation for them first. But for students who are already proficient, we*

> *might give them more difficult questions, so that they feel like it matches their ability and that it's not boring in the class (Mathematics teacher, Private school (School 5)).*

> *When the [midterm exam] results came out, I would give focus on weak and also smart students. I would have activities or questions that are more complex and incorporate other subjects for the good students. It depends on the person, but I do it like this. I have questions in different levels for the students. All students need to be able to do the same sets of questions, but there are additional ones for the smart students (Mathematics teacher, Satit school).*

It seems like teaching to the test has been practiced largely in low-performing schools. The strategy may allow students to perform well in O-NET, but they might not learn much, or be able to apply the knowledge elsewhere. Intensive preparation for examinations is not inherently bad. When the exam is of good quality, teaching to test allows students to learn important contents. However, a test that is heavily reliant on remembering contents, such as O-NET, can be less useful for the students.

5.4.3.3. PISA and TIMSS are less prioritised in relation to O-NET

So far, O-NET results are the main assessment results that matter for quality assurance, career progression, or are tied to other incentives. PISA and TIMSS play less role in school and teacher evaluation since not all schools take the exams, and for those who do, results are anonymised and cannot be linked to them. With this premise, schools are not pressured to perform well on the two tests. This leads to less preparation in low-performing schools.

> *We [took PISA exam], yes. For me, it's not like a life-or-death situation, that we need to score well, score high. It's not like that (Mathematics teacher, OBEC2 school (school 2)).*

The schools view PISA as being quite separate from what is taught in schools. The education system supports memorisation and simple application of knowledge that is tested by O-NET. PISA, on the other hand, focuses on critical thinking and more complex set of knowledge, as one teacher illustrated.

> *Let me say that overall, [PISA] is a good exam, but it's an exam that Thai students never take, never experience, never familiar with. They have not learned this since they were young. … the teaching method, the curriculum, for us, these don't support [PISA exam]. It's like we are not preparing our students for this exam (Academic teacher, Private school (school 6)).*

When students are not exposed to the same type of questions or knowledge required in PISA, it is unlikely that they will perform well, especially without much preparation. For the learning outcomes to improve, the curriculum, pedagogy, and assessment needs to align to focus on these skills that are required in the international assessments. Additionally, when schools view PISA as separate from normal class contents, they perceive themselves to be less accountable for its results. This aligns with research by Hopkins et al. (2008) that shows practitioners perceive policymakers to be accountable to PISA results more than them. With less accountability, schools are less likely to perceive PISA as important. This is as expected since PISA is not designed to have schools accountable for the results. Instead of making schools be more accountable for PISA, the Ministry should integrate desired PISA skills into the national curriculum and assessment.

However, PISA is perceived differently for the high-performers. For Science schools, PISA is part of performance indicators (scoring above certain threshold and the relative rank within the Science schools themselves). This is possible because they know all schools would participate. This certainty, combined with pressure from their committee, makes pressure to perform higher for the Science schools. This results in them preparing for PISA more intensively than other schools, including tutoring and changing internal examination to match PISA. This plays a part in helping them maintain high learning outcomes. Additionally, as they score so well, the Science schools have been asked to help neighbouring schools to prepare for PISA by organising workshops.

For Satit school, there is no formal pressure like Science schools. Nonetheless, teachers are noticeably much more knowledgeable and familiar about PISA compared to the low-performing schools. This might be because individual Satit schools are more likely to be selected to be in PISA sample[126]. The Satit school in the qualitative sample reported being in the PISA sample for almost all rounds. Of 25 Satit schools, 15 were sampled in 2015 (see Appendix C2). This means that taking PISA exams is viewed as a more common occurrence than other schools. As a result, there are teachers assigned to prepare for PISA in multiple rounds. With this, the high-performers prioritise on PISA much more than the low-performers, which can mostly explain the score gap in the international assessments.

5.4.4. Schools focus on different skills

Despite the highly centralised structure, there remains flexibility in teaching in some respects, including in teaching additional subjects[127]. There are clear disparities in what are offered in high- and low-performing schools. For example, the BMA school interviewed added vocational subjects, including hairdressing and sewing as most students go on to vocational schools. By contrast, Satit and Science schools added subjects that allow students to go deeper into the contents required by the curriculum. Satit school has Independent Study subject, which is a project-based subject for students to research what they are interested in while

[126] This is due to there being smaller number of Satit schools overall. Consequently, they are more likely to be included in the sample, compared to other schools. Nonetheless, to calculate the final group-level scores, the weights are applied so that Satit schools are weighted proportional to enrolment, despite being sampled more.

[127] For approximately 27% of teaching hours in secondary, schools and local authorities can set up additional subjects (OBEC, 2010). The additional subjects are contextualised based on students' and communities' needs.

Science school has additional Mathematics and Science subjects that teach deeper contents than the national curriculum.

> *For example, we emphasise Math and Science for the students to study. The indicators we set up additionally so that they are higher than the standards of the core curriculum. And mostly, for Science, there are subjects which train students to think critically. For example, seminar subject, innovation subject, independent work subject, investigation subject (Science teacher, Science school).*

The subjects the high performers focus on clearly teach higher level skills that would benefit them in taking tests such as PISA. Students learn skills such as critical thinking and problem-solving, which are not normally listed in the national curriculum. In Satit school, these skills are integrated into school curriculum and vision. Students are expected to have soft skills such as "*able to think, have the courage to think, discuss, act, and be leaders academically (Science teacher, Satit school)*". This partially explains why Satit and Science schools did well in PISA.

Even within the core subjects, the skills emphasised are different and contents are taught differently in different schools. All schools are mandated to teach the same contents based on the indicators listed in the core curriculum. However, low-performing schools have trouble teaching all contents within the allotted time. This is partly from students' poorer foundational knowledge. Teachers have to spend time revising prior knowledge before teaching the actual grade-appropriate contents. Hence, even with the same amount of time, low-performing schools spend more time teaching the same contents than high-performing schools. Many resorted to lecturing the contents and assign many contents for students to self-study at home.

> *When I'm teaching it's like, [I tell students to] write these down. I will give the answers in class and have them review. But whether they review or not, I don't know. … But I try to get them to do exercises. Right or wrong answers, it's ok, I just want them to try. Then when I give the answers, they can write down and read along (Science teacher, OBEC1 school).*

The way the contents are taught is quite different in high-performing schools. A science teacher in Satit school recounted the way he taught students about tides. Rather than telling students what tidal forces are, he had students analyse real-world data of sea water levels, before discussing with students what forces they think are behind the changing water levels. Then, students experiment with a scientific model that simulates tidal waves. The lesson finishes with students coming to their own conclusion of how tidal forces work and how that can be used to explain natural phenomena. With this, students learn about the contents by conducting experiments themselves, rather than being lectured like in OBEC1 school. Furthermore, the teacher's assessment of knowledge is very focused on application and critical thinking.

With the skills such as problem-solving and critical thinking embedded in teaching and assessment, it is logical that students from these schools are more proficient in these skills and can do well in PISA. It is troubling that these skills have not already been included in the curriculum, but instead have to be an add-on from the schools. This shows a misalignment among the curriculum, pedagogy, and assessment. As curriculum and assessment focuses on memorisation and simple application of knowledge, these are the skills that are taught to students. If Thailand wants its PISA results to improve, the curriculum, pedagogy, and assessment need to be adjusted to include PISA-related skills.

5.5. Qualitative results: What had changed over time?

In addition to factors affecting scores, qualitative interviews also include what participants perceived as changes over time that can influence learning outcomes. Even though their answers cannot be used to directly answer research question 3, they provide important insights of changes in practices. During the interviews, a specific question of "What had changed over time, especially between your first and most recent experience of PISA or TIMSS?[128]" was asked to the participants. It was also used as prompts when participants talked about aspects such as their teaching/management practices, assessment practices, quality assurance, accountability system, etc. With this, we can dive into what teachers perceive on students' skills, and what may have caused students to perform worse over time. For example, teaching practices or schools' priorities on skills may have changed over time, driving the score changes. Asking Science school participants about the changes can also yield recommendations related to school improvement. They can discuss about what other adjustments the school made when they became Science schools and had to teach a different curriculum. Additionally, participants can introduce other explanations that were not captured in the available quantitative data.

As the interviews were conducted in 2020, many participants discuss changes that happened more recently than our time frame of interest. These changes are included in 0 as they do not explain the trends during 2011-2015. Notably, there were recent changes made to include competencies and application of knowledge in the national curriculum, and subsequently assessment and pedagogy. Understanding these recent changes allow us to know the direction Thailand is moving towards to solve the problem of low learning outcomes.

Below, changes over time as perceived by the participants (within study time frame) are discussed.

5.5.1. Changes in pedagogy

Most schools reported changes in pedagogy as a result of curriculum and assessment changes after 2017. They switched from focusing on memorisation to

[128] Or from when they started working at the school/from 2011 – 2012 comparing to now, so that the experience mirrors the time frame of the study.

application of knowledge (see 0), with changes in O-NET as the main driving force for their changes in practice.

For some schools, the shift came earlier (happened within the study time frame). For the outliers in the sample, Local school and Private school (school 5), the schools made the shift away from memorising contents to teaching PISA-like skills and critical thinking starting from 2007 and 2014 respectively. This is driven primarily by visions of the principals and vice principals at the school. For both schools, the management team at the school saw the importance of including skills beyond memorisation and pushed for change to happen. This shows that when schools have visionary leaders, change can happen within the system. Other than the vision, both principals seem to have resources in making the shift. They recognise that equipping teachers with skills are of importance, and both had connections with instructors who can train teachers to teach critical thinking. Local school participants mentioned that this biannual training plays a big part in their teacher professional development.

> *We are like this because we are lucky that the principal's teacher was an instructor specialising in PISA, in analysing and reading. So the principal coordinated with his teacher so that he came to instruct the teachers on designing exams following PISA guidelines, twice per semester, every year, we have teacher professional development to make them able to design exams according to PISA. … In 2007, we were under the principal who took interest in active learning style of teaching, so he sent the managing committee and teachers to learn how to do it (Mathematics teacher, Local school).*

Satit and Science schools also shifted to teaching beyond memorisation earlier than the rest of the schools. The previous section attributed the schools' score advantage partly to their ability to teach further and deeper than the curriculum requires. Even though both schools seem to have always been teaching like this, it is only recently that the schools started to incorporate PISA-style of examinations in (questions asking students to provide short answers and those with elaborated context provided). Both schools started to increase the importance given to PISA after the results of PISA 2012 came out. For Science schools, PISA became part of their indicators for internal quality assurance. That made the schools incorporate PISA in teaching and internal examinations. This higher focus on critical thinking and PISA style of questions can explain why these schools perform much better than others in both rounds of PISA.

5.5.2. Greater focus on O-NET

One of the biggest changes that all participants mentioned is having greater focus on O-NET. Compared to in 2008, when O-NET was first tested in grades 6, 9, and 12, O-NET has increased its importance over time, and especially during 2012 and 2015. It plays bigger part in quality assurance and evaluation of teachers and schools, as well as in students' entrance to universities. Participants mentioned increase in pressure to perform well on O-NET, both to strive for score increase and scoring above the national average (see section 5.4.3.).

This probably came partly as a response to poor O-NET results themselves but also poor PISA results. From 2008, O-NET results have not been up to the standards desired by the Ministry of Education. Average score is consistently below 50% correct for each subject, and scores in core subjects of Mathematics, Science, and English have always been lower than other subjects (OECD & UNESCO, 2016). Another part why the pressure increased came from PISA results, especially those of the 2012 and 2015 rounds. Even though Thailand has been participating in PISA since 2000, the assessment itself became known to the wider public as well as among teachers after the results of PISA 2012 came out. Media highlighted the poor performance of Thailand in terms of ranking and also proportion of students not reaching minimum proficiency levels defined by the OECD (Lathapipat & Sondergaard, 2015).

With this, the Ministry increased pressure to schools to improve learning outcomes. As previously discussed, since PISA and TIMSS are sample based, schools cannot be held accountable for their PISA scores. Hence, the pressure falls on the schools to improve their O-NET scores, another measure of learning outcomes. The policy is communicated down to the local authorities and to schools.

> *Lately, the ESA gave greater importance and attention [to O-NET]. Probably in the past 5 years. … they come in and play a bigger part. I think it's because of the national examinations. I think it started with PISA results, which are considered poor, so they gave more importance on the national assessment (Academic teacher, OBEC2 school (school 3)).*

To respond to the increased pressure, school engage in many of the preparation strategies discussed earlier (tutoring, exam analysis, etc.). Many of the preparation strategies have always been there. However, teachers reported that tutoring has increased in terms of intensity in the recent years. Teachers also started incorporating O-NET questions into their lessons. As time passed, teachers themselves also gained more experience in O-NET and reported that they can teach them better than before.

> *The result analysis has been there for quite a while, like analysing which departments need improvement, but we didn't do complex analysis, … we just see the data. But lately there are more complex analysis for each department, what contents (need focus). Because in each department, there are many indicators. We look at it in detail, by indicators now. Because O-NET exams mostly are based on indicators. But not everything is in the exam. Some topics they only cover some indicators (Academic teacher, OBEC2 school (school 3)).*

These preparations help familiarise students with O-NET exam questions and learn skills tested in O-NET. As O-NET exams during 2012 and 2015 were heavily emphasised on memorising and simple applications of knowledge, it is arguable that students are now better in lower level skills with the extensive preparation by the schools. With this, it may explain why students seem to be performing better in memorising and less so on higher level skills as seen in TIMSS and PISA.

Extensive preparation and teaching to the test are not necessarily negative practices. When the tests include the skills that are beneficial for the students to learn, being focused on the test can help ensure students do learn the contents. However, in trying to improve O-NET scores during 2012 and 2015, when O-NET exams are very content heavy and rely on memorisation, it is arguable that students learn less useful skills.

5.5.3. Change in Science schools

Another change we saw from O-NET data is the increased in advantage of being in Science schools in 2015 as compared to 2012. This can be attributed to changes in the admission policies and school curriculum. Arguably, the schools have made a successful transition, with the new cohorts of students scoring exceptionally well on both national and international assessments, as well as competing in global stages. This section dives deeper to see what had changed from the participants' perspectives. Interviews were made with teachers who had remained at the school before and after the transition period. Knowing this can contribute to policy recommendations in school improvement. We can learn from what adjustments are needed to make the schools able to teach the curriculum that is evidently more demanding than the national curriculum. This learning can be adjusted to help other schools transition to changes made by the Ministry.

The school went through intensive preparation before admission policies changed. Most of the preparation is borne out of a partnership with Mahidol Wittayanusorn school (MWIT). MWIT is the first special-purpose schools that cater specifically to students gifted in Mathematics and Science (MWIT, 2022). It is funded by the government and received 325 million Bath (£7.5 million) in the 2019 financial year (Budget Bureau, 2019). The funding is much greater than normal public schools.

With this, MWIT is able to pay teachers more and attract those who graduated at Master or PhD level to become teachers at the school.

While MWIT was able to enhance learning for those gifted in Science and Mathematics, enrolment per year is very small, there are only upper secondary levels, and the school is located in Bangkok. The creation of Science schools is, in some way, an aspiration to expand the success of MWIT to schools in different regions and grade levels. While the budget allocated to Science schools is greater than OBEC2 schools, the difference is less than when OBEC2 is compared to MWIT. In some ways, Science schools are more similar to OBEC2 schools. Teacher allocation is based on the same system as OBEC2, and they do not have budget to pay teachers more. Resources the school get are similar to public schools. Additionally, while there is periodical hiring of new teachers, most teachers were the same teachers who had been there prior to the change to become Science schools.

At first, in 2008, each of the 12 Science schools enrol one pilot classroom that uses the Science curriculum and special examinations developed by MWIT. The rest of the classes were taught with the national curriculum as usual. During this period, teachers attended boot camps with MWIT teachers to familiarise themselves with the new curriculum and train themselves to teach different skills. This cooperation helped teachers adjust to teach in a new way without putting too much pressure on them. MWIT teachers act as mentors throughout the period.

> *[MWIT] started training teachers to design exams, training teachers to be able to train students, meaning we are in a boot camp with them during summer break. We trained teachers rigorously, so that they have the experience and the skills. We had a think, for students focusing on Mathematics and Science, what skills should they have? If the students are to have skills like MWIT students, the teachers need to have these characteristics and knowledge. They were trained like that. Then after they were trained, they came back to teach students like at MWIT (Vice Principal, Science school).*

Then, in 2011, the enrolment of grades seven and ten began and by 2013, the schools have fully become Science schools. MWIT staffs still act as mentors, but the schools were gradually supported to be able to operate independently of MWIT. At first, the internal examinations were written by MWIT. Then, when the teachers were trained, Science school teachers were involved in the examination design process jointly with MWIT. In this adjustment period, the exam questions came both from the MWIT experts and Science school teachers. The teachers got feedback on whether the exam items were of good quality or not, and what needs to be adjusted. Now, the school forms a committee among the 12 schools to design exam together. Teacher training with MWIT is still conducted every two years to help teachers keep up with new teaching methods or trends in Science teaching.

This adjustment allows the Science schools to have a successful transition from normal schools to special-purpose science schools. Note that the change takes years to implement, and many supports were given along the way. This is in contrast to the way the curriculum has been implemented in all schools. From the participants' perspectives, there is very little time to adjust to the curriculum and teachers have to adjust by themselves. If the schools are supported in this way, they can better implement the policy changes made by the Ministry.

5.6. Discussion

This chapter shows that the quantitative results are similar among the assessments. Regressions show similar effect of being in certain school types, as well as the effect of having specific student and school characteristics. This supports the validity of comparing results among the assessments, and extending beyond simply comparing mean and dispersion of outcomes as in Brown et al. (2007). It also confirms their usefulness in guiding policies. O-NET, PISA, and TIMSS should be used jointly to form understanding of Thailand's nature of learning outcomes. Specifically, due to different data availability and contents tested, all datasets could be used to complement one another to derive insights about the nature of achievement profile of Thai students and inequality of the outcomes. For instance, PISA and TIMSS's wealth of socioeconomic variables are useful in analysing gaps between students of different socioeconomic background. O-NET, by contrast, can show effect of schools by looking at value-added when prior scores are taken into account. Currently, reports have been produced for PISA and TIMSS after the country's participation. However, it is unclear how the data has been used for improvement. At a school level, qualitative results show that schools analyse O-NET data but not others. Particularly, most schools use data on which topics the schools need improvement to plan for teaching. This shows how much influence O-NET has on influencing practice. It implies that O-NET could be used strategically to help mould teaching practices.

Results show that *there is a large and persistent gap among the top performing schools (Satit and Science) and the average and low performing schools*. When observable characteristics were controlled for in regression analyses, the advantage of the high-performers reduces. Particularly, inclusion of prior scores and socioeconomic variables reduce the coefficient significantly. This shows that the advantage these schools have over other schools partially stem from enrolling students who are more well-off and perform better academically. This is similar to selective grammar schools in the UK that enrol students from better socioeconomic and academic background. Even though academic merit plays a part in getting the students into these schools, research shows that other factors, such as well-off families paying for private tutoring can result in clustering of pupils with high socioeconomic status (Jerrim & Sims, 2019).

Nonetheless, this does not tell the whole story as Satit and Science schools still show large advantages even after these characteristics were included in the model. This shows that there are other differences not captured in the model that affect the differences. For the low-performers, the results diverge for O-NET and PISA. For O-NET, the modelling results show that once the background characteristics were controlled for, the low-performing school types do not perform differently from general public schools. In PISA, OBEC1 shows large and positive coefficient while BMA shows negative coefficient, implying that OBEC1 performs well despite the unfavourable characteristics and BMA performs much worse. The qualitative results dived into the rationale behind why this is the case.

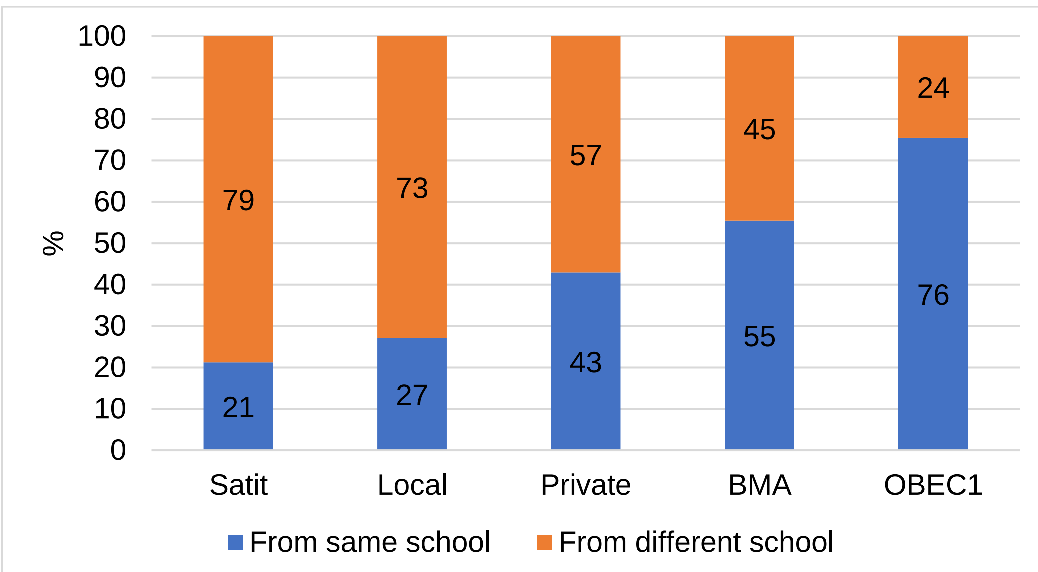
The advantage stems largely from differences in student and school characteristics. Results from both quantitative and qualitative data point in the same direction in this aspect. Nonetheless, most of what affect low performance are beyond simple observable characteristics. Similar to results found in World Bank (2012a), qualitative results show that students in low-performing schools have weak foundational knowledge (as reflected in lower O-NET prior scores). Rationale given by World Bank (2012a) for weak knowledge is that schools such as OBEC1 and BMA are not popular with parents. Those enrolling there are because they cannot get into what parents perceive to be "better schools", which usually require academic examinations. Hence, the schools end up enrolling primarily students with lower academic ability. Consequently, it takes more time for them to be taught the same number of contents compared to high-performing schools, where students are more academically ready. This shows that low learning outcomes at grade nine are the cumulated result of many years of students not learning the contents. To improve learning outcomes at secondary, the education system needs start improving at primary levels and ensure that students learn what are expected of them. This chapter shows that teacher practice affects what students learn. Additionally, factors such as lack of resources such as subject-specialised teachers and self-selection of students also affect learning outcomes.

Teachers in average and low-performing schools mention the problem of brain drain, with their best students choosing to study elsewhere when they have the opportunity. When schools are unable to retain their best students, this created a vicious cycle of low performance. Quantitative data seems to support this hypothesis. Using O-NET data of 2014, Figure 5-9[129] shows proportion of students

[129] The figure excludes Science schools and OBEC2 schools as they have 100% of students from different schools due to their nature. Science schools do not have primary levels. Hence, all

who stayed in the same school from primary to lower secondary (those who had taken O-NET grade 6 and 9 at the same school), and those who had moved from a different school. The figure shows that aside from OBEC1 schools, where most students tend to continue in the same school, there are quite large movements among schools as students transition from primary to lower secondary. Students are more likely to move schools during these grade levels.

Figure 5-9 Proportion of grade nine students who studied primary in the same school, by school type



Source: Own calculation using O-NET 2014 microdata

Data on prior scores supports teachers' narratives partly that students who moved schools seem to be those who perform better academically. Table 5-9 shows average Mathematics score difference of each group of students. That is, their average score minus average score by school type. In both grade levels, students enrolled in OBEC2 and Satit schools who had transitioned from other schools have

students came from different schools. For OBEC2 schools, secondary schools are separate from primary schools. They have different principals and adopt different EMIS codes. Therefore, all students are tagged as moving schools. Note that 58% of students remained in OBEC2 school type from primary to secondary.

higher academic achievement. The difference is larger in grade 12. This shows that students who moved to these schools are indeed those who performed well academically in primary. Then, after finishing lower secondary in these schools, the top-performers moved to other schools. When looking at proportion of new enrolment in lower secondary, top students in OBEC2 schools are more likely to move to Science schools and top students in Private schools are more likely to move to Satit schools. Schools lose their best students as they choose to transition to Satit and Science schools, and their average outcomes remain low.

Table 5-9 Mathematics score differences of students in different groups

		Based on grade 6		**Based on grade 12**		
	School type score average	**Studied primary in the same school**	**Studied primary in different school**	**Continue in same school**	**Continue in different school**	**No O-NET score in grade 12**
OBEC1	-0.12	0.00	-0.01	N/A	0.13	-0.05
OBEC2	0.16	N/A	N/A	0.16	0.37	-0.31
Private	0.17	0.15	-0.12	0.20	0.16	-0.29
BMA	-0.21	-0.01	0.01	0.21	0.16	-0.05
Local	-0.05	-0.04	0.02	0.11	0.15	-0.11
Satit	1.74	-0.32	0.09	-0.13	0.64	-0.39
Science	2.87	N/A	N/A	0.23	-0.39	-0.02

Source: Own calculation using O-NET 2014 microdata

It is likely that students moved to what Nakornthap (2018) refers to as famous prestigious “Name” schools. These schools have reputations of high academic prestige and virtually all Satit schools are listed there. This highlights a structural problem in the Thai education system. Students who perform better academically self-select themselves into schools with prestige such as Satit and Science schools. This left other schools with students who perform less well and presumably cannot get into Satit and Science schools. Some of these students may have been switching out to Private schools, as the data shows Private schools enrolling new students with poorer performance. Usually, Private schools have later enrolment deadlines than OBEC and Satit schools. Hence, it is possible that those students failing to enrol in the prestigious schools may have settled with Private schools instead. This makes the outcomes remain unequal as schools cannot stop students from enrolling into new schools.

By contrast, students enrolling in BMA, OBEC1, and Local schools do not differ significantly in their academic levels, regardless of whether they came from a different school (see Table 5-9). This shows that those who score higher did not choose these schools to transition to. New students coming also have very similar achievement levels to those who had not transitioned. While the high-performers move to more prestigious schools, it seems the low-performers are more likely to remain where they are, or at most move to a different school in the same school type with roughly the same academic performance. This disparity shows that students with differing academic performance have different transition trajectories.

Note also that students without O-NET scores in grade 12 all perform much worse than those who continued. Most of students who did not take O-NET in grade 12 are either those who dropped out, continued in vocational schools, or chose alternative education. This supports teachers' narrative that these students who do not wish to pursue university education focus less on scoring well on O-NET (at grade nine) and lack motivation.

To reduce inequality in education, Thailand has recently introduced several measures. The Ministry of Education had provided a budget for the Equitable Education Fund (EEF) to reduce education inequality. This has been primarily through giving cash transfers directly to disadvantaged students (EEF, 2021a). They also recently provided capacity support to schools in the rural areas through the Teacher School Quality Program: TSQP, where schools partner with NGOs to receive support in teaching through activities such as mentoring. The organisation is relatively young (established in 2018) and has a relatively limited reach. It is supporting around 700,000 students yearly. While this certainly helped reduce inequality, there certainly remains areas of improvement as the problems of weak foundation and motivation have not been addressed.

To tackle the problem of lack of resources in small schools, a solution is proposed by the World Bank. As they are difficult to manage efficiently, they proposed for small schools to be closed (Lathapipat & Sondergaard, 2015). It is hugely unpopular and has not been implemented because of the pushback. This shows that the change needs to be gradual and provide support through for example transport. Another way to increase number of teachers in the rural area is through “Kru Kuen Tin” (MOE, 2012), where teachers can choose to be relocated back to their hometowns. This helped putting qualified teachers in provinces outside Bangkok. However, little research has been done regarding how much of this actually help staff small schools.

Another potential reason that children did not learn the contents may stem from the curriculum itself. The curriculum includes many indicators and contents teachers are required to teach (see section 2.3.3). However, when asked what teachers believe students should know once they finish compulsory education, very few mentioned curricula contents. Instead, teachers want students to have skills that they can use to survive and thrive in everyday life. When prompted what those skills are, many mention basic literacy and numeracy. Beyond that, students should have occupational skills and knowledge they need for further studies. 21st century skills such as collaboration and empathy are also mentioned as critical for students to be able to live harmoniously with others and face challenges of a globalised world. These skills are highlighted as being demanded of the global workforce in recent years (Joynes et al., 2019). Identifying as needed, these are the skills that teachers themselves try to incorporate into lessons in addition to the content requirements. Other than literacy and numeracy, these skills are not in the curriculum. 21st century skills were mentioned briefly in the curriculum. Nonetheless, there are no guidelines how those skills can be incorporated into lessons. Perhaps including these in would motivate teachers to teach more and make school more relevant for students as well.

Updating to make the curriculum remain relevant is especially important when many students lack motivation to learn the current contents. The Ministry should reflect on what knowledge and skills are important for students to have. Now, the governance structure is very rigid and standardised, which can be good at ensuring the contents are taught uniformly. However, when the national curriculum does not reflect the current demand, schools have to find their own leeway to incorporate the contents and skills they find important. To make learning more relevant as well as improving learning outcomes in PISA, these skills should not be optional. They need to be integrated as part of the curriculum and assessment to ensure they are taught in all schools. Part of the solution is to review the curriculum frequently. Countries such as Finland have set time frame, when the Education Decree (and subsequently curriculum) are reviewed and updated (Halinen, 2018). When this is done regularly, the curriculum is made more relevant to the current demand. The Ministry needs to reflect on what is needed in the curriculum as well as the national assessment to ensure there is a right balance between what the stakeholders want and what skills should students have that would benefit the country. Arguably, higher order skills such as critical thinking and problem-solving should be included in addition to basic literacy and numeracy, as they are included in both PISA and what the teachers see as important.

Another way to make curriculum more relevant is to allow schools and local authorities to adapt the curriculum to local context to a greater extent. There seems to be a movement towards less centralisation in terms of curriculum. The new curriculum in Wales, for example, allows flexibility for schools to write their own school curriculum that is appropriate with the school context (Hwb, 2021). Similarly in Thailand, schools have students with varying background and participants advocate for devolving this decision. Currently, there are some adaptations, such as BMA schools opening courses for hairdressing. There are some pros and cons associated with this. While this allows schools to contextualise, it can also be a cause of achievement gap as high-performing schools chose to go deeper into academic contents while low-performing schools chose to teach practical subjects. With this, high-performing schools ended up teaching skills that enable them to perform well in international tests. Hence, it is important for the government to rethink what should be learned in schools, so that schools such as OBEC1 also include the skills.

In addition to curriculum being the constraint of achieving higher learning outcomes, the quality assurance plays a part as well. Results show that assessment results are the main source of pressure for low-performing schools, while not pressuring high-performing schools enough. This stems from the criteria being the same for all schools. To ensure quality assurance leads to improvement, the criteria should be different from school to school. In some respect, the system is already in place. On paper, the principals or head of each department are allowed to set their own goals. However, one teacher said the goals cannot be set lower than what the local authority (and the Ministry) already sets. Hence, it seems there is not much flexibility. Additionally, there is pressure for scores to increase every year. Schools cannot set goals that the scores would be lower or even remain the same.

> *[Principal] won't allow us to set lower goals. And when we have a meeting and we propose [the goals], he will consider again whether that should be increased further, it shouldn't be this low, like this (Science teacher, OBEC2 school (school 2)).*

I argue that schools (or local authorities) should be allowed to set their own goals and justify them. This would allow goals to be tailored to the school context more. Schools and local authorities should be supported to make this decision. Lacking budget and personnel capacity are parts of the reasons why local authorities were unable to create a fully local curriculum (OECD & UNESCO, 2016). Hence, they should be given this support. This problem is also acknowledged in Science schools. At the time of the interview, the vice principal mentioned that they are currently drafting new indicators for the schools, so that they incorporate other performance indicators such as academic competitions and higher O-NET goals. With more challenging goals, the high-performers may be more motivated to improve over time.

To ensure the system runs smoothly, pedagogy needs to also align with the curriculum. A teacher emphasises that the way the contents are taught affect how well students would perform on different assessments, especially PISA that requires students to know higher-order skills. Hence, teachers should be supported to teach the new skills.

> *If we teach students to memorise the formula to use, I don't think they would do well in PISA. But if in our teaching, we allow students to think by themselves or create their own knowledge, students can build upon their knowledge more. And I think they will score better on PISA (Mathematics teacher, Satit school).*

Another part of the puzzle is assessment. Results show how much assessment dictate what happens in classrooms, with teachers changing taught contents to match O-NET. Hence, we need to ensure that O-NET measures what we want to include in the curriculum.

One of the unintended consequences brought about by the school evaluation is that schools are pitted against other schools to be at the top of the ranking table. This means that schools are not collaborating as much as they should. In education system of high-performing countries, such as British Columbia, teaching professionals share their best practices in professional learning communities (S. Brown et al., 2017). As a result, teaching quality improves. In Thailand, these learning communities are not as common among schools. Participants mentioned that there are some learning communities within school. Teachers would share what they did to help increase O-NET scores, for example. However, between different schools, this kind of sharing is not common nor systematic. Professional learning communities can allow high-performing schools to help low-performing schools improve. In addition to the expertise, resources could be shared among them including teachers. For instance, specialised teachers could be shared among schools where possible.

Another aspect of the results worth diving into is the peculiarity of OBEC1's performance in PISA 2012. Quantitative results show that OBEC1 performed exceptionally well in 2012 and shows sharp decline in 2015. While this is purely speculative, it is possible that this change is driven by change in PISA exam format. For PISA, it is important to note that in 2015, the exam changed from being paper-based to being computer-based. In qualitative interviews, some participants offer some insights on how this change can affect the scores. Challenges that come with computer-based exams may affect groups of students differently. For example, students who are not familiar with typing may take significantly longer time to compete the test. In some schools, students prepare for PISA using practice exams on computers. Science school, for example, explicitly said training how to use computer is one of the prioritised strategies. In other schools, practice tests are done on paper. While the effect is difficult to isolate quantitatively, section 6.2.1 shows that OBEC1 is most deprived of all school types in terms of computer resources. With this, it is possible that students who might not be so familiar with computer-based tests, such as those in OBEC1 schools, perform less well in PISA 2015.

Further, what teachers noted as changes over time elucidate many features of the Thai education. One limitation of this research is that many changes mentioned by the participants happened in more recent years than the time frame of this study. This is probably because the fieldwork was conducted in 2020 – 2021, and participants are more likely to remember changes that are more recent as opposed to the changes of 8 – 9 years ago. These changes are summarised in Appendix F.

While interviews have provided insights on what Thailand has been doing to improve learning outcomes, it provides less explanation on why there are discrepancies in trends over time. Part of the answers lie in the greater focus on O-NET over the time period of interest. With schools focusing more on the test that are mainly memorisation, students learn more of the memorised contents, and less on application of knowledge. This shows how important O-NET is in influencing what is taught in classrooms. When this is aligned with what the Ministry wants students to learn, it can be a powerful force for change.

Changes made to Science schools also offer strategies that can be adapted to help low-performing schools. The schools are supported to change the curriculum. Not only that the schools started with a pilot classroom, MWIT worked closely with them to help teachers adjust to teach the new curriculum and design internal assessment. This kind of collaboration is currently absent with other schools. Currently, there are small informal collaboration between high-performing schools and schools in their vicinity. Both Science school and Satit school in the sample said they helped low-performing schools prepare for O-NET and PISA, and well as advising them in terms of best teaching practices. Arguably, the support can be more formal and mandated by the Ministry or the local authorities. If the low-performing schools had support from the top performers, they may be able to teach better.

Even though this is not part of the study time frame, changes made to curriculum and assessment in 2017 offers important insights to how change should be implemented. After implementation, some participants, especially those from low-performing schools, struggled to keep up with the changes. Some switched to easier textbooks or chose easier topics to focus on. This arguably happened because the curriculum revision was probably done without many stakeholder consultations. There have not been many public documents on curriculum review process. Additionally, within the curriculum itself, the aims and rationales behind the shift do not mention any views from stakeholders beyond IPST[130] and the Ministry (IPST & MOE, 2017). When the curriculum is imposed in a top-down manner such as this one, it can lead to less buy-in from the stakeholders. Additionally, while the implementation is staggered (starting from grade seven in the first year, then eight and nine in the subsequent two years), there were no adjustment periods for the schools. Participants also did not mentioned support

[130] IPST partnered with the Ministry to help write curricula contents.

given to schools beyond teachers' guides provided. For schools that already struggled with the previous curriculum, adding application of knowledge can be challenging. These schools should have been better supported to ensure smooth transition and that the curricula changes can improve learning outcomes.

Lastly, as different data is available for different assessments, data availability can be a limiting factor in analysing learning outcomes. TIMSS has shown to be the most restrictive dataset out of the three. The report itself does not provide a sampling framework, nor the weights it used to calculate the scores. From section 5.2.2, sample comparison with enrolment data cannot be done with TIMSS, and it is not possible to analyse if TIMSS's sample is representative or not. Additionally, school type subgroup is not available in the microdata of TIMSS and PISA 2015 (the latest PISA 2018 also omits this information). This hinders the ability of researchers to conduct secondary data analyses, especially when school type is shown to be one of the most important factors in predicting achievement. IPST's PISA reports show detailed descriptive statistics by school types. However, most of them are related to achievement. Examples of statistics provided by school type are overall scores, proficiency levels, and scores by sub-topics. Other results are very limited. We do not know, for example, levels of enrolment by school type, or levels of different socioeconomic indicators by school type[131]. These are important insights that can aid policymakers. When school type identifier is not available in microdata, these types of analyses are not possible.

[131] The report provides limited number of socioeconomic indicators by school type, including just indices of ESCS and resources shortage. We do not know, for example, distribution of parental education or by home educational resources.

Chapter Six: In What Ways Are the Trends from PISA/TIMSS/O-NET Similar and Different?

6.1. Introduction

This chapter answers the second research question of “In what ways are the trends from PISA/TIMSS/O-NET similar and different?”. It starts by exploring the key differences in assessments that can potentially drive the differences in trends as well as gaps in learning outcomes. Next, the overall trends of PISA, TIMSS, and O-NET are revisited. Then, score differentials among different groups are analysed to gain insight on where the inequalities are and to what extent. After which, the trends of different groups are analysed. The chapter concludes with discussion of implications on what the findings contribute to what we know about changes in learning outcomes and what this means to the Thai education.

6.2. Differences in the assessments

Before diving into the findings, it is important to discuss differences in the assessment designs themselves and how they can affect results. This section analyses differences in contents tested and sampling characteristics of the three assessments.

6.2.1. Content differences

It is possible that the differences observed in trends over time stem from the inherent differences in contents and skills being tested in the three assessments. Section 2.2.4shows that due to differences in the test objectives, the contents tested are slightly different. This section expands on this by discussing the differences in characteristics and proportions of different items included.

Item types

Table 6-1 shows distribution of types of questions asked in each of the assessment. PISA relies less on simple multiple-choice compared to other types of items. TIMSS has a similar proportion of multiple-choice questions and constructed response questions. O-NET, by contrast, relies heavily on simple multiple-choice questions. Arguably, this makes O-NET the simplest test among the three (and possibly the easiest). With many questions being multiple-choice, it is easier for students to guess when they do not know the answers.

Table 6-1 Item type distribution

Types of items	Mathematics			Science		
	PISA	TIMSS	O-NET	PISA	TIMSS	O-NET
Simple multiple-choice	25%	55%	80%	29%	58%	80%
Complex multiple-choice	17%	-	-	36%	-	20%
Constructed response	58%	45%	20%	35%	42%	-

Sources: Own calculation based on released item characteristics of PISA 2015 (OECD, 2017b), TIMSS 2015 (TIMSS & PIRLS International Study Center, 2017), and O-NET 2015 (NIETS, 2021)

Cognitive domains

Other than item type, the assessments vary in terms of cognitive domains. As shown in Table 6-2, there are clear discrepancies in the emphasis of each cognitive domain among the assessments. In Mathematics, PISA and TIMSS focus more on the second domain (Employ/Evaluate/Applying), while O-NET focuses more on the lowest domain of Knowing. This implies that PISA and TIMSS may be testing higher-level skills and are more difficult overall. In Science, TIMSS and O-NET has a more similar distribution while PISA focuses on the Explain and Evaluate domain. This means O-NET Science scores may be more comparable with TIMSS than with PISA as they are more similar.

Table 6-2 Cognitive domain distribution

PISA's cognitive domains	TIMSS's cognitive domains	Mathematics			Science		
		PISA	TIMSS	O-NET	PISA	TIMSS	O-NET
Formulate / Explain	Knowing	28%	31%	43%	48%	36%	33%
Employ/ Evaluate	Applying	43%	45%	27%	21%	41%	42%
Interpret	Reasoning	28%	24%	30%	30%	23%	24%

Sources: Own calculation based on released item characteristics of PISA 2015 (OECD, 2017b), TIMSS 2015 (TIMSS & PIRLS International Study Center, 2017), and O-NET 2015 (NIETS, 2021)

When O-NET, the national assessment, focuses less on constructed response and higher cognitive domains, there can be implications on what students learn. Students taught with O-NET as a goal may perform worse on these types of questions if they are not emphasised in the test. This can affect their performance on PISA and TIMSS, which include more of these items.

Content domains

Table 6-3 and Table 6-4 shows distribution of each content. In Mathematics, while PISA shows roughly equal distribution in each content, TIMSS tests students more in Numbers O-NET in Geometry. In Science, similar emphasis is on Biology. Majority of O-NET questions are in Physics, which is the area less prioritised by TIMSS. If students perform better in one topic compared to others, this can affect exam results as they would perform better in tests that have more coverage of that topic.

Table 6-3 Mathematics content domain distribution

PISA's content domains	TIMSS's content domains	PISA	TIMSS	O-NET
Numbers	Quantity	26%	31%	16%
Algebra	Change and Relationships	25%	28%	19%
Geometry	Shape and Space	23%	21%	29%
Data and chance	Uncertainty and Data	26%	21%	16%[132]

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b), TIMSS (TIMSS & PIRLS International Study Center, 2017), and O-NET (NIETS, 2021)

[132] The number does not add up to 100% here because the assessment also includes 20% of points allocated to "Mathematical Skills and Processes". See full breakdown of exams by learning units in 0.

Table 6-4 Science content domain distribution

PISA's content domains	TIMSS's content domains	PISA	TIMSS	O-NET
Biology	Living Systems	40%	36%	30%
Chemistry	Physical Systems	33%	19%	12%
Physics			24%	44%
Earth Science	Earth and Space Systems	27%	21%	14%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b), TIMSS (TIMSS & PIRLS International Study Center, 2017), and O-NET (NIETS, 2021)

To conclude, there seems to be some degree of differences in the items included in PISA, TIMSS, and O-NET. O-NET includes more simple items that requires lower-level skills, while PISA includes more higher-level items. Contents coverage varies and each assessment has slightly different focus. These differences can drive variation in scores as well as change over time. As O-NET is a national test linked with school evaluation, there is less incentive for schools or students to learn the contents or skills present in PISA and TIMSS but not O-NET. Thailand has no control over what is tested in PISA or TIMSS. What they can control is what is tested in O-NET. If the goal is to improve Thailand's global rankings and learning outcomes, the contents and skills should be emphasised in both the national curriculum and assessment.

6.2.2. Sample differences

Other than differences in the tested contents, there may be meaningful differences in sample of students being tested, both as compared to the national population and among the assessments themselves. The section starts with comparing targeted population of the assessments to the actual population. Then, characteristics of sample are compared.

It is important to note the similarities and differences in target population of the three assessments. As discussed in section 2.2.2, all assessments' target population are students who are enrolled in school. By focusing on only those enrolled in schools, the assessments are excluding out-of-school children. Spaull (2017) noted that specifically for PISA, enrolled students are not necessarily representative of the population of actual 15-year-olds in the country. In the context where many students are out-of-school, PISA sample can include disproportionately more well-off students.

Inclusion of out-of-school children

In Thailand, out-of-school children may not be as major a concern as in Turkey. Based on Multiple Indicator Cluster Survey[133] (MICS) that Thailand participated in 2015 and 2019, very few children and youths are out-of-school. At primary level, enrolment is virtually universal and only 1% of age-appropriate children are not enrolled at all. At secondary level, as shown in Table 6-5, there are more out-of-school youths, with the majority from upper secondary levels. Even though the proportion is small, it must be acknowledged that when looking at performance of students in PISA, TIMSS, and O-NET, we are ignoring at least 15% of age-appropriate children who are out-of-school.

[133] A nationally representative household survey, which collects data on both enrolled and out-of-school children. Note that even these household surveys still likely fail to account for the poorest out-of-school children, especially those who are homeless or nomadic (Carr-Hill, 2012). Hence, there may be more out-of-school children than reported

Table 6-5 Percentage of enrolled VS out-of-school children

Percentage of children	2015-2016	2019		
	Secondary	**Overall secondary**	**Lower secondary**	**Upper secondary**
% attending school at the right level	81%	79%	87%	70%
% attending school at lower level	5%	11%	10%	13%
% not attending school at all	14%	10%	3%	18%

Sources: NSO & UNICEF (2016, 2020)

Comparison with enrolment data

Enrolment data also suggests that less students continue in upper secondary. While almost all primary students continue to lower secondary, transition rate from lower secondary to upper secondary remains around 58% from 2011 to 2020 (NSO, 2021). Additionally, net enrolment rates are 104%, 97%, and 80% in primary, lower secondary, and upper secondary levels respectively. This difference in transition rates means that upper secondary is more selective than lower secondary, which reflects the structure that education is compulsory up until lower secondary (OECD & UNESCO, 2016). This self-selection can potentially drive differences in the assessment results as the assessments target different grade levels. While TIMSS and O-NET focus on students lower secondary, the majority of PISA sample (73%) are from upper secondary. Consequently, PISA sample includes more students who already self-select themselves to be in upper secondary. These students may be from a more well-off family or are more academically motivated.

Next, the proportion of students in PISA and O-NET sample[134] is compared with enrolment by school types (Table 6-6). Regardless of grades, most students are enrolled in OBEC2 schools. The proportion remains constant over time for each grade level. Transitioning from grade 9 to grade 10, there is no OBEC1 school type anymore. This is because by definition, OBEC1 schools do not have upper secondary level (grade 10-12). The number of students also fell significantly as students enter grade 10. This reflects the transition at end of compulsory education at grade 9. Nonetheless, public school enrolment (OBEC1 and OBEC2) of the three grades remains constant.

In terms of alignment, O-NET's composition very closely match that of the enrolment data of grade 9 students. This is to be expected as O-NET exams are compulsory for grade 9 students. For PISA, the sampling frame matches grade 10 enrolment more closely. More OBEC2 students are sampled when comparing to enrolment. However, grade 9 composition differ slightly from the enrolment. The share of OBEC schools match that of the population enrolment data. Yet, more OBEC1 schools are represented in PISA compared to OBEC2. Since PISA sampled based on number of 15-year-olds, we may not expect the sample to match the enrolment exactly as there might be more 15-year-olds in some school types more than others. More sampled grade 9 students in OBEC1 schools might indicate that OBEC1 schools have more overaged students compared to other school types. With this, we can imply that while O-NET sample mirrors enrolment, PISA sample is more similar to enrolment in grade 10 and not grade 9. The difference can drive the score differences when OBEC1 schools perform significantly different from other schools.

[134] TIMSS was excluded as even though TIMSS provided a sampling frame, the weighting of each school type was not provided.

Table 6-6 Enrolment and proportion of test takers by school type and grade

	Grade 8	Grade 9			Grade 10	
	Enrolment	**Enrolment**	**O-NET**	**PISA**	**Enrolment**	**PISA**
OBEC1	22%	22%	21%	31%	-	-
OBEC2	54%	55%	55%	47%	79%	83%
Private	14%	13%	15%	12%	13%	11%
BMA	1%	1%	1%	1%	0.2%	0.1%
Local	7%	7%	7%	9%	6%	5%
Satit	0.4%	1%	1%	0.1%	1%	1%
Science	N/A	N/A	0.2%	0.1%	N/A	0.4%
Number of students	770,226	770,874			453,588	

Sources: Enrolment data based on NESDC (2019), O-NET and PISA percentages are from own calculation of microdata.

Coverage and exclusion rates

Representativeness can also be viewed from coverage and exclusion rates. Both PISA and TIMSS report very high coverage of the targeted population. PISA has 1.6% exclusion rate while TIMSS reported 0.2% exclusion rate. This means that less than 2% of schools were not part of the sampling frame. For PISA, excluded schools are those which are inaccessible or teach only ineligible students (such as those with functional disabilities) (OECD, 2016a). For TIMSS, excluded schools are very small schools with less than five students and special-needs schools (Mullis et al., 2016). It is likely that the schools excluded are not at random, and that they enrolled more disadvantaged students than the sampled schools. Nonetheless, the percentage of schools not covered is very small, and is unlikely to significantly affect results. Similarly, student-level exclusion is small. There are not many students selected to be in the sample but are absent on the test date.

Composition by socioeconomic indicators

Next, group composition is compared by socioeconomic indicators of number of books in the home (Table 6-7), and parental education (Table 6-8). PISA shows higher proportion of students with more books at home. While this can be interpreted as PISA having students from more well-off background, it may be reflecting students of different grade levels. Students in PISA sample are older than students in TIMSS sample. Hence, they might naturally have more books at home. By contrast, students participating in TIMSS seems to have higher parental education than PISA. More than 1/3 of parents of PISA sampled students got education less than lower secondary. It is important to note that parents of TIMSS sample may be relatively younger than parents of PISA sampled students as most TIMSS sample are 14-year-olds and PISA are 15-year-olds. Being younger, TIMSS students' parents may have access to more education opportunities as the government engaged in education expansion over the years. Even if we account for this, the figure still shows a considerable difference of a two-fold jump in the lower secondary level. It can be implied that PISA sample may be less well-off than TIMSS.

Table 6-7 Proportion of students in terms of books at home

Books at home	PISA	TIMSS
0-10 books	21%	25%
11-25 books	34%	48%
26-100 books	30%	19%
101-200 books	8%	4%
More than 200 books	6%	3%

Sources: Official PISA and TIMSS 2015 reports (Mullis et al., 2016; OECD, 2016a)

Table 6-8 Proportion of students in terms of mother and father's education

Mother's education	PISA	TIMSS	Father's education	PISA	TIMSS
Some primary or lower secondary or did not attend school	36%	14%	Some primary or lower secondary or did not attend school	36%	12%
Lower secondary	19%	45%	Lower secondary	18%	43%
Upper secondary or above	45%	41%	Upper secondary or above	46%	45%

Sources: Official PISA and TIMSS 2015 reports (Mullis et al., 2016; OECD, 2016a)

To conclude, the differences in sample among PISA, TIMSS, and O-NET is greatest in terms of grade levels included. PISA includes large proportion of students in upper secondary, while TIMSS and O-NET include exclusively students from lower secondary. As education is compulsory up until lower secondary, PISA sample is included students who want to study beyond compulsory education, perhaps with intention to enrol in tertiary education. These are likely to be students who may be more motivated academically. Other than this, the sampling of PISA and TIMSS is robust, mirroring national enrolment in terms of school type composition. When comparing proportion of socioeconomic indicators, TIMSS may sample more advantaged students than PISA.

Other than the differences in contents tested and sample, it is important to note other differences that can drive variation in scores as well as change over time including the change to computer-based test in PISA 2015. Nonetheless, it is not possible to investigate the extent to which these drive change in scores.

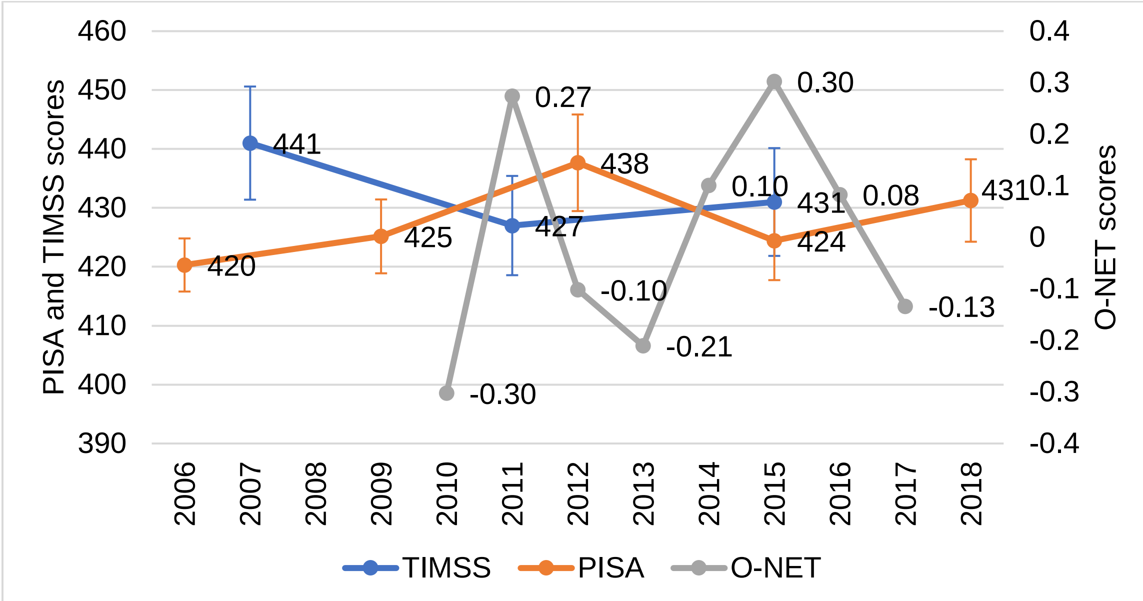
Knowing these differences, the next section compares the overall trends of the three assessments.

6.3. Overall trends

As seen initially from Chapter One, Figure 6-1 and Figure 6-2 are brought up again here for in-depth analysis. From 2006-2018, PISA results show stagnation (slight, but insignificant improvement) in Mathematics. In Science, however, the increase is 0.2 standard deviation and is meaningful. TIMSS shows insignificant decline. O-NET scores swing up and down with no clear trends in Mathematics. Science scores appear to go up, then come back down after 2014. A continual decline has been shown in recent years. Additionally, PISA shows an unusual spike in 2012, before coming down again in 2015.

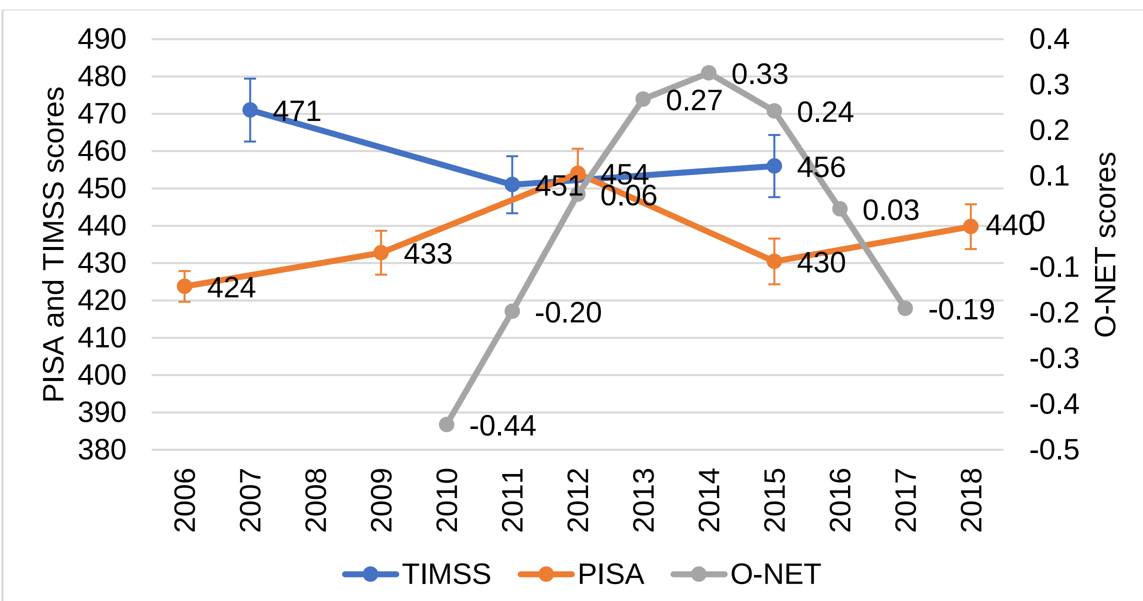
For policymakers, the graphs give a confusing picture of how the Thai education system performs as the trends are not coherent. For example, from 2006-2012, PISA scores went up while TIMSS went down. Therefore, it is difficult to say whether the learning outcomes improved or not. The most problematic trends that prompt need for further investigation is that of O-NET. Unlike PISA and TIMSS, where results are very similar in Mathematics and Science, O-NET shows very different trends. Additionally, the year-to-year score changes are large and meaningful, as opposed to simply being noise. It is possible that there are differences in test design or stakeholders' actions that cause this. Another interesting point is 2012 PISA, which is unlikely to reflect true trends, as the scores went back to 2009 levels in 2015 before showing slight improvement in 2018. Knowing why there is increase can help policymakers understand if something temporary was being done right to improve learning outcomes, or whether it is simply a glitch in the assessment implementation. These aspects will be explored throughout this chapter and the subsequent chapter.

Figure 6-1 Trends in Mathematics performance between 2006-2018



Source: Own calculation using PISA, TIMSS, and O-NET microdata

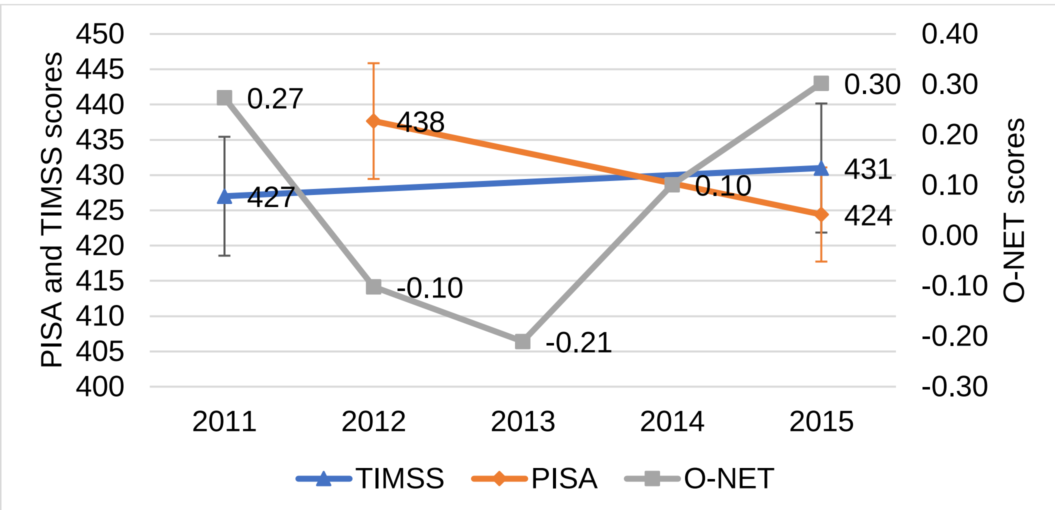
Figure 6-2 Trends in Science performance between 2006-2018



Source: Own calculation using PISA, TIMSS, and O-NET microdata

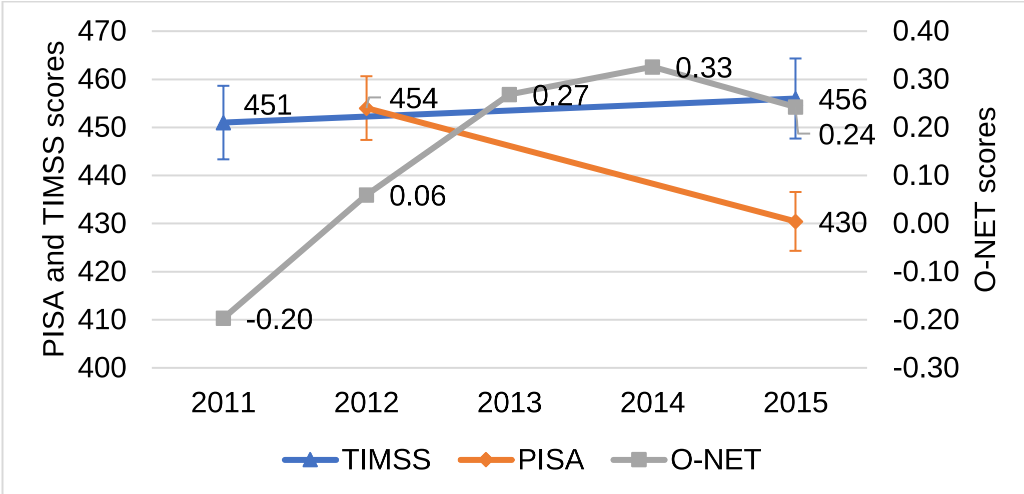
For the analysis to be more focused and in-depth, a time period of 2011 – 2015 was selected for investigation in this thesis. Comparing PISA scores of 2012 and 2015 can also shed light on what makes 2012 different than other years. Figure 6-3, Figure 6-4, and Table 6-9 show score trend over this period for Mathematics and Science. PISA shows insignificant decline in score in Mathematics and meaningful decline in Science scores. TIMSS shows insignificant increase. O-NET Mathematics trend shows a u-shape, which can be seen as an increase if looking from 2012-2015, or a stagnation if looking from 2011-2015. For the purpose of the analysis in the thesis, O-NET scores of 2012 is used when calculating change over time. It is important that this is interpreted with caution.

Figure 6-3 Trends in Mathematics performance between 2011-2015



Source: Own calculation using PISA, TIMSS, and O-NET microdata

Figure 6-4 Trends in Science performance between 2011-2015



Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table 6-9 Change in overall scores

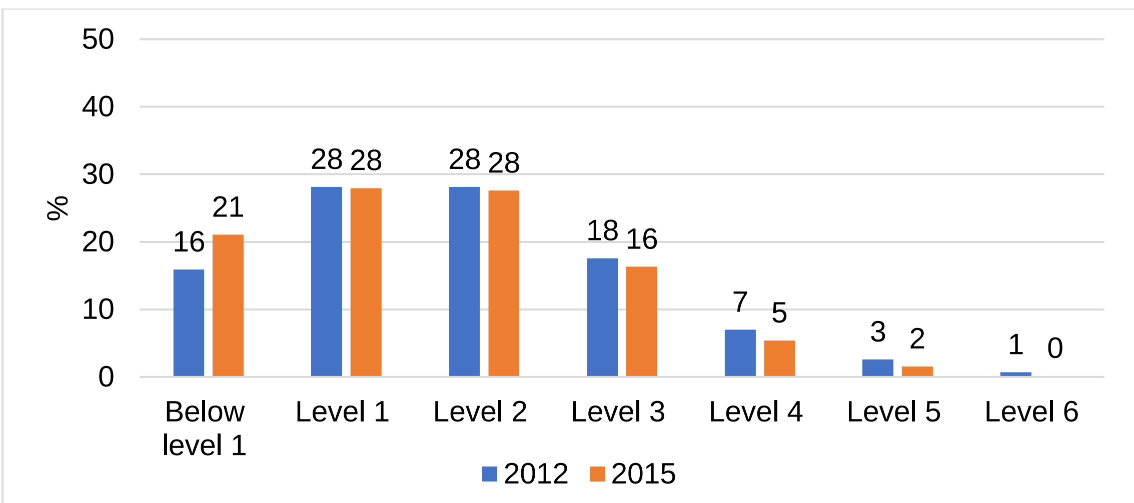
Subject	PISA 2012-2015	TIMSS 2011-2015	O-NET 2011-2015	O-NET 2012-2015
Mathematics	-13	4	0.03	**0.40**
Science	**-24**	5	**0.44**	0.18

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

The scores can be described in terms of proficiency levels, which describes what students can and cannot do. In both subjects, average scores of PISA in Thailand are in level 2 and level 1 in TIMSS. As proficiency level two is considered minimum proficiency levels for PISA and TIMSS (see section 2.2.9), Thailand's average scores are below minimum international requirements in TIMSS.

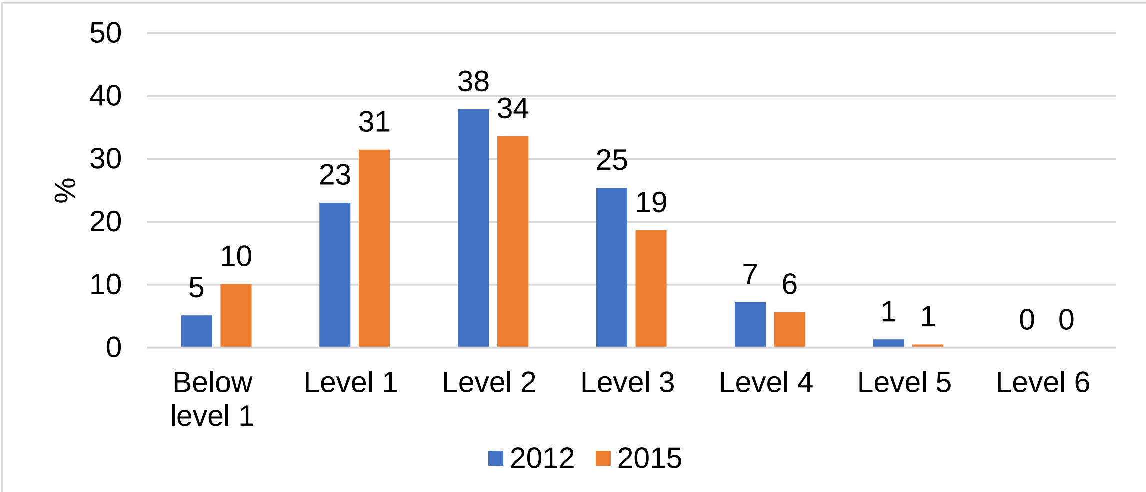
Analysing beyond the mean, Figure 6-5 and Figure 6-6 show PISA score distribution by proficiency levels in Mathematics and Science, respectively. Both subjects show a right-tailed distribution, with the majority of students performing at level 1 and 2. This means that most students are performing at the lower end of the distribution. In Mathematics, the distribution is similar in both years, with an increase of five percentage points in proportion of students scoring at the lowest level. This reflects a small decline in average PISA score shown in Table 6-9. In Science, by contrast, the largest change observed from 2012 to 2015 are in the groups of students below baseline proficiency levels. Proportion of students below level 1 doubled from 5% to 10%. Meanwhile, proportion of level 1 scorers increased by eight percentage points. This change reflects a decline in learning levels and can also be a cause of concern as most of the decline is from proportion of students scoring level 3.

Figure 6-5 PISA Mathematics: Percentage of students at each proficiency levels from 2012-2015



Sources: Official PISA 2012 and 2015 reports (OECD, 2014, 2016a)

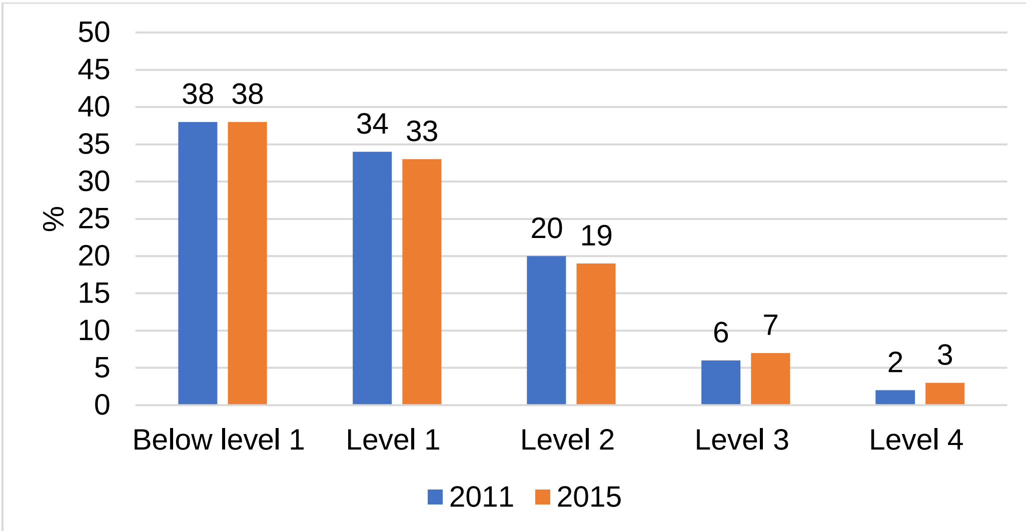
Figure 6-6 PISA Science: Percentage of students at each proficiency levels from 2012-2015



Sources: Official PISA 2012 and 2015 reports (OECD, 2014, 2016a)

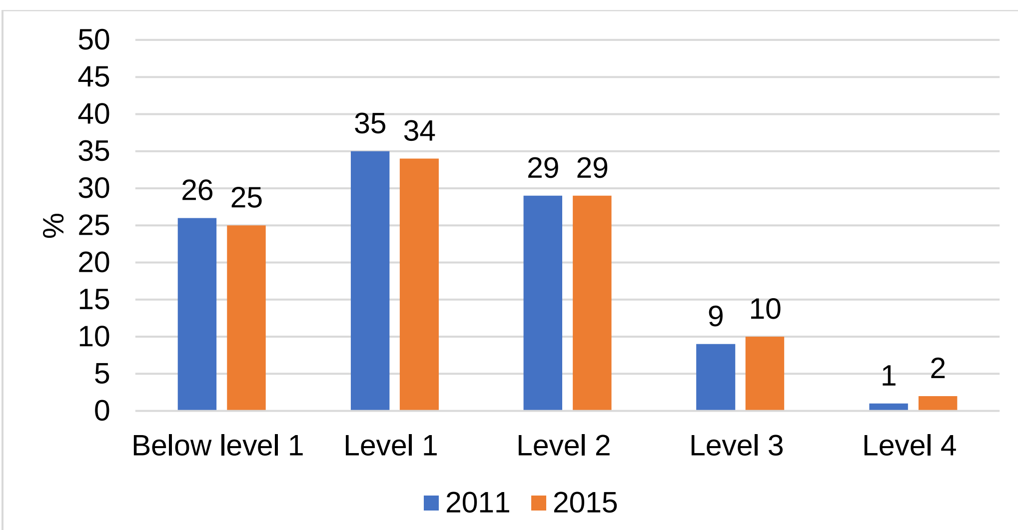
As for TIMSS, Figure 6-7 and Figure 6-8 show that, similar to PISA, the score distribution is right-skewed, with the majority in the proficiency levels of one and below one. However, there is much greater proportion of students who are below level 1 than in PISA. There is no significant change over time. The proportion of students remains similar across 2011 and 2015.

Figure 6-7 TIMSS Mathematics: Percentage of students at each proficiency levels from 2011-2015



Source: Official TIMSS 2015 Thailand report (IPST, 2017)

Figure 6-8 TIMSS Science: Percentage of students at each proficiency levels from 2011-2015



Source: Official TIMSS 2015 Thailand report (IPST, 2017)

The scores show some similarities between the two assessments. In addition to both assessments showing right-skewed distribution, it can be implied that Thai students perform better in Science than in Mathematics. In both assessments, larger proportion of students perform at low levels in Mathematics than Science. In PISA, while 24% of students perform below level 1 in Mathematics, only 1% are at that level in Science. In TIMSS, 38% are in level 1 in Mathematics while 25% are in level 1 in Science.

To conclude, both TIMSS and PISA score distribution shows that the majority of Thai students perform lower than baseline levels of competencies, confirming the low levels of learning outcomes. PISA data shows that this got worse over time, with more students in lower than baseline proficiency levels. While TIMSS's distributions remain roughly the same, it is worrying that many students perform below level 2. Thailand's learning outcomes would improve significantly if these students' learning levels can be improved. Another point to note is that the learning outcomes are not uniformly poor. There are students performing at baseline or above. In PISA, this accounts for around half the students. In TIMSS, it is around 40% in Mathematics and 30% in Science. A few performs at the highest level of proficiency, matching that of the high-performing countries. With this, it is important to identify who these high performers are, and what can be learned from them to help improve the overall outcomes. Hence, the next sections analyse scores by subgroups to identify where the learning gaps are as well as their magnitudes, and how they change over time.

6.4. Gaps in learning outcomes

The section starts with overview of the magnitude of gaps for all the interested subgroups, before diving into each subgroup to see who the high-performers are.

6.4.1. Overview of gaps in learning outcomes

Table 6-10 presents gaps (in standard deviation) between the learning outcomes of the highest performing groups compared to lowest performing groups for each category[135]. This gives a snapshot of the extent of problems of inequalities in Thailand and where the highest gaps are.

Except for gender, the gaps are large and meaningful. Notably, the gaps among the school types are larger than other subgroups in all three assessments. The top school types, Science and Satit schools are special-purpose schools with competitive admission policies and are reputed in academic rigour (see section 2.3.2). Given their very selective nature, it is likely that they perform better than other schools. The gaps narrowed when these schools were excluded, but remain high, indicating that the learning outcomes are very unequal even among schools that are more similar. The magnitudes of school type gaps are now around the same level with gaps by regions, urban-rural, and school sizes. This observed inequality is alarming as it signals the deeper structural inequality in the Thai education. For performance percentiles, the gaps are higher in the upper part of the score distribution (P90-P50) compared to P50-P10. This shows that the gap is mainly from the high performers scoring much higher than average performers. While it is concerning that the inequality is large, this also gives a reason to be optimistic. If we can adapt some practices from the high performers to the low performers, we may be able to improve learning outcomes.

The last three rows of the table show gaps by item characteristics. Data here is presented differently as percentage instead of standardised scores, as it is calculated from percent correct of each category in the subgroup. The numbers here represent the difference in percentage points of the highest and the lowest category. For example, students got 42% of simple multiple-choice questions correct and 23% of constructed response in PISA, hence the difference is 19%.

[135] Unstandardised results are in Appendix E3.

Data shows that there is some degree of differences in terms of item type and cognitive domain. With this, it can be implied that Thai students are much less skilled in answering questions that are more complex, regardless of the topic. This is one of the issues that need addressing.

Table 6-10 Gaps in learning outcomes by each subgroup

Subgroups	Subjects	PISA	TIMSS	O-NET
School types	Math	**2.01**	**1.72**	**3.33**
	Science	**2.28**	**1.63**	**2.79**
School types – excluding Satit and Science schools	Math	**0.50**	**0.45**	**0.40**
	Science	**0.63**	**0.48**	**0.40**
Regions	Math	**0.67**	**0.67**	**0.52**
	Science	**0.67**	**0.69**	**0.40**
Urban-rural	Math	**0.39**	**0.69**	**0.46**
	Science	**0.54**	**0.68**	**0.40**
School size	Math	**0.66**	**0.44**	**0.74**
	Science	**0.88**	**0.46**	**0.66**
Performance percentile (P90-P50)	Math	**1.31**	**1.29**	**1.69**
	Science	**1.43**	**1.20**	**1.61**
Performance percentile (P50-P10)	Math	**1.21**	**1.07**	**0.93**
	Science	**1.17**	**1.21**	**0.97**
Number of books at home	Math	**0.89**	**1.11**	N/A
	Science	**0.84**	**1.15**	N/A
Parental education	Math	**0.35**	**0.56**	N/A
	Science	**0.41**	**0.57**	N/A
Gender	Math	0.01	0.19	0.21
	Science	0.08	**0.23**	0.23
Item types	Math	19%	19%	N/A
	Science	23%	22%	N/A

Subgroups	Subjects	PISA	TIMSS	O-NET
Cognitive domain	Math	20%	6%	N/A
	Science	8%	11%	N/A
Content domain	Math	15%	15%	N/A
	Science	11%	9%	N/A

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

From the overview, the gaps among subgroups are large and significant, especially for school types. The next section dives further into identifying who the top and bottom performers are, and how each group performs relative to average score.

6.4.2. Gap by school types

From Table 6-11 and Table 6-12[136], all test results agree on which school types are the best performers (Science and Satit schools), the lower performers (OBEC1, Local, and BMA), and the average performers (OBEC2). Both Science and Satit schools perform above international average, with Science schools' scores equalling the top-scoring countries. By contrast, OBEC1 schools perform meaningfully worse than the mean. The pattern is also similar across subject. Nonetheless, there remain slight differences in performance that are more difficult to pinpoint why. For example, BMA and Local schools perform worse than the mean in only PISA and O-NET, and Private schools perform at the mean level in TIMSS and O-NET, but not PISA.

From section 5.2.2, OBEC1 schools are included in the PISA sample disproportionately more than actual enrolment in grade 9. As their learning outcomes are much worse than average, including more of OBEC1 schools can contribute to the low scores. A former Thai minister (2018) asserted that the Thai

[136] PISA and TIMSS scores are standardised with mean and standard deviation of pooled 2011, 2012, and 2015 scores. The tables with unstandardised scores are shown in Appendix E3.

PISA sample includes more disadvantaged students compared to actual enrolment, and that this explains why the scores are low. There seems to be some truth in it. However, weighting the scores with proportion from enrolment only increases the scores by five points. Hence, greater proportion of OBEC1 students only contribute marginally to explaining the low scores.

There is significant level of inequality in terms of scores by school types, with the top and bottom differing by around two standard deviations. In both subjects, the top performers significantly scored above the mean by more than one standard deviation. The difference is largest in O-NET, followed by TIMSS and PISA. While the bottom performers scored from 0.3 – 0.5 standard deviation below the mean, the top performers scored 1 – 3 standard deviations above the mean. This means that the large gaps are driven by the top performers. This highlights the importance of knowing what those top performers did well to achieve such high level of learning outcomes, and whether those practices can be adapted to lower performing schools. Especially so as the low performers scored significantly worse than the average outcomes. If OBEC1, BMA, and Local schools can be brought up to average, mean learning outcomes will improve.

Table 6-11 Mathematics ranking by school type

School type	PISA Rank	PISA Score	PISA Dif from mean	TIMSS Rank	TIMSS Score	TIMSS Dif from mean	O-NET Rank	O-NET Score	O-NET Dif from mean
Science	1	1.51	**1.60**	1	2.13	**2.10**	1	3.33	**3.03**
Satit	2	0.87	**0.95**	2	1.44	**1.42**	2	2.09	**1.79**
OBEC2	3	0.00	0.08	4	0.09	0.07	4	0.39	0.09
Private	4	-0.32	**-0.24**	3	0.17	0.15	3	0.40	0.10
OBEC1	5	-0.34	**-0.26**	7	-0.28	**-0.30**	6	0.00	**-0.31**
BMA	6	-0.49	**-0.41**	6	0.02	0.00	7	-0.01	**-0.31**
Local	7	-0.50	**-0.42**	5	0.04	0.02	5	0.03	**-0.28**
Average		-0.08			0.02			0.30	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

Table 6-12 Science ranking by school type

School type	PISA Rank	PISA Score	PISA Dif from mean	TIMSS Rank	TIMSS Score	TIMSS Dif from mean	O-NET Rank	O-NET Score	O-NET Dif from mean
Science	1	1.59	**1.75**	1	2.01	**1.98**	1	2.73	**2.48**
Satit	2	0.86	**1.02**	2	1.35	**1.31**	2	1.77	**1.52**
OBEC2	3	-0.06	0.10	4	0.09	0.06	3	0.34	0.10
Private	4	-0.42	**-0.26**	3	0.20	0.16	4	0.28	0.04
OBEC1	5	-0.58	**-0.43**	7	-0.28	**-0.31**	5	0.00	**-0.25**
Local	6	-0.66	**-0.50**	5	0.06	0.02	7	-0.06	**-0.30**
BMA	7	-0.69	**-0.53**	6	0.06	0.02	6	-0.03	**-0.28**
Average		-0.16			0.03			0.24	

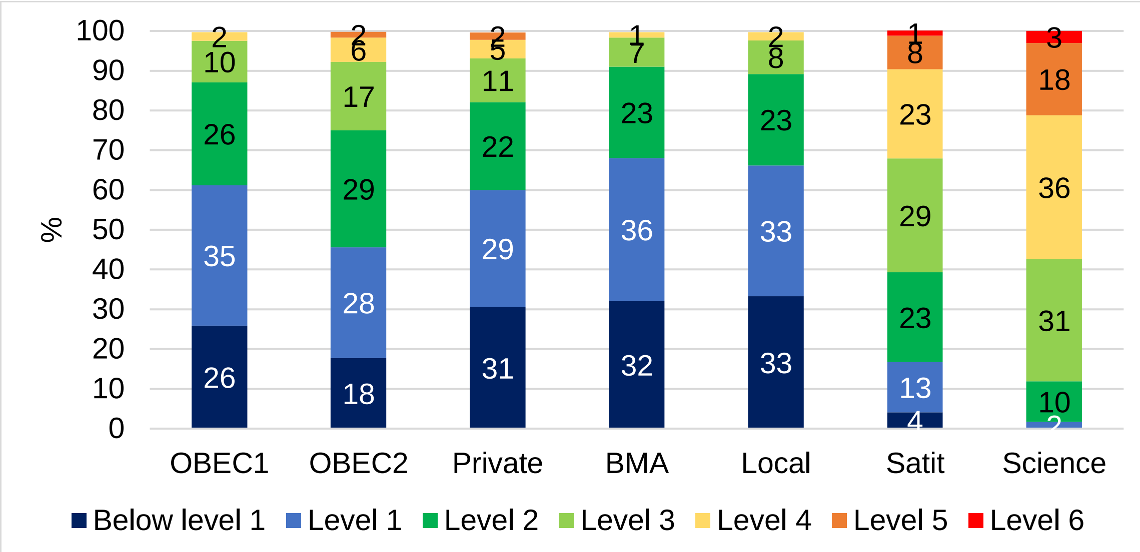
Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

To illustrate the differences further, Figure 6-9 and Figure 6-10 show distribution of scores by proficiency levels in PISA[137]. Satit and Science schools have markedly different distributions compared to other school types. Most students perform above baseline level 2. The rest of the school types have score distributions that are more right skewed. The schools that perform worse, such as BMA, Local, and OBEC1 schools show greater proportion of students below level 1 and at level 1, as well as very small number of students above level 3. The pattern is similar across subjects, but the difference is greater in Science compared to Mathematics. This shows that the majority of students in low performing school types are not performing up to the standards. Chapter Five shows that this gap is from both differences in student background and teaching practices.

Additionally, students have higher learning levels in Mathematics compared to Science. There are virtually no students below level 1 in Science while up to 1/3 of students are below level 1 in Mathematics. This reaffirms the results from section 5.3. Despite the decline in Science scores in 2015, score distribution is better in Science than Mathematics.

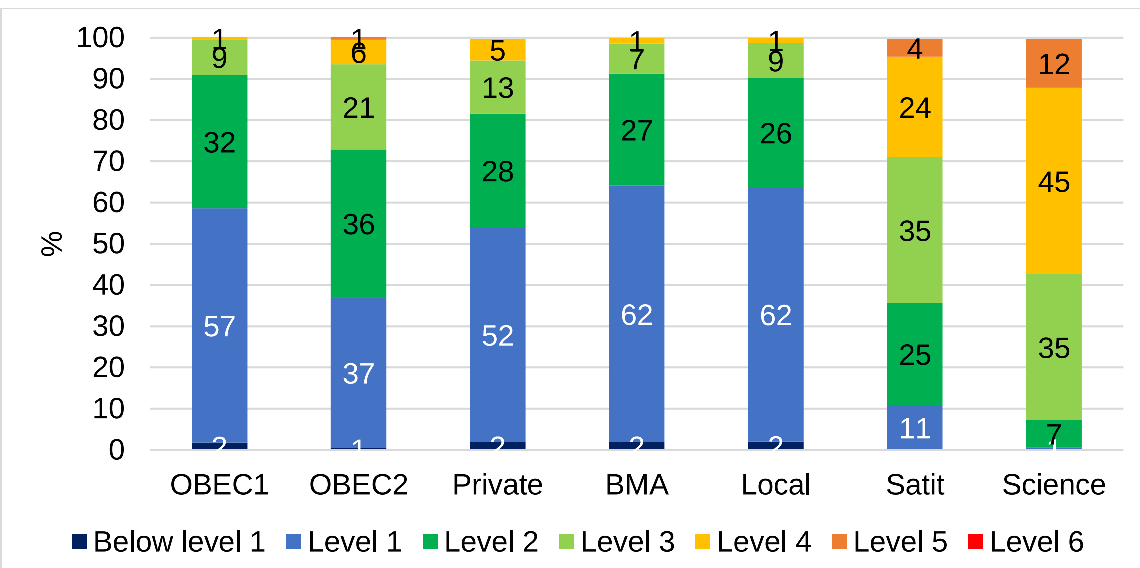
[137] Only PISA is explored here as this information is not available in TIMSS.

Figure 6-9 Proportion of students at different proficiency levels in PISA Mathematics 2015, by school type



Source: Own calculation using PISA 2015 microdata

Figure 6-10 Proportion of students at different proficiency levels in PISA Science 2015, by school type



Source: Own calculation using PISA 2015 microdata

6.4.3. Gaps by other subgroups

Gaps by other subgroups show smaller magnitude compared to school type. Results show generally high level of similarities among subjects and assessments. The highest and lowest rank groups are virtually the same across assessments and subjects[138]. The rankings of these groups are generally as expected based on literature (for instance, schools in Bangkok perform better than the rest of the regions). The magnitudes are also similar across assessments. Other than those in the highest and lowest rank, most of the other subgroup categories do not perform statistically differently from the mean. Results are summarised below.

Gap by school characteristics

Table 6-13 shows the subgroup with highest and lowest performance in terms of region, urban/rural location, and school size.

Region shows the most disagreement in gaps. As most regions show no significant different from the mean, the differences here may be due to sampling of PISA and TIMSS. With large score gaps observed among school types, if some school types are sampled more in some regions than others, this can affect the average score of that region. However, it is not possible to test this hypothesis since PISA and TIMSS do not have region identifiers. What the assessment results agree on is that the region of lower northeast performs significantly worse than others. Bangkok and its perimeter is the region that outperforms the rest in PISA and O-NET but curiously ranked fourth in TIMSS. By contrast, East is the best performing region in TIMSS by quite a large margin.

As expected, students from urban and extra-large schools perform significantly above the mean. In terms of school size, it seems contradictory with literature in

[138] Except for region and gender

higher income countries, that found smaller school sizes and class sizes to be associated with better quality and better learning outcomes (DfE, 2011; Egalite & Kisida, 2016). Especially because these schools also have larger class sizes compared to other schools (average of 43 students per class compared to 27 students per class in small schools[139]). One of the rationales from the literature is that small schools have much less resources compared to larger schools (Lathapipat & Sondergaard, 2015). They are usually located in the rural areas. This is supported by PISA data, where 95% of small schools are in the rural area while only 59% of extra-large schools are[140]. Additionally, there remains teacher shortages in the core subjects of Mathematics, Science, and English (Siribanpitak, 2018). The differences in resources should be taken into consideration when making recommendations to improve learning outcomes.

[139] Own calculation of PISA 2015 data
[140] Own calculation of PISA 2015 data

Table 6-13 Summary of ranking by school characteristics

	PISA		TIMSS		O-NET	
	Math	**Science**	**Math**	**Science**	**Math**	**Science**
Region						
Highest rank	BKK and perimeter		East		BKK and perimeter	
Dif from mean	**0.39**	**0.27**	**0.34**	**0.36**	**0.32**	**0.26**
Lowest rank	Lower NE		Lower North		Lower NE	
Dif from mean	**-0.27**	**-0.40**	**-0.32**	**-0.33**	-0.20	-0.14
Urban/Rural						
Highest rank	Urban					
Dif from mean	**0.31**	**0.45**	**0.53**	**0.52**	**0.29**	**0.25**
Lowest rank	Rural					
Dif from mean	-0.08	-0.09	-0.16	-0.16	-0.17	-0.15
School size[141]						
Highest rank	Extra-large					
Dif from mean	**0.38**	**0.38**	0.09	0.08	**0.41**	**0.38**
Lowest rank	Small					
Dif from mean	**-0.28**	**-0.42**	**-0.35**	**-0.38**	**-0.33**	**-0.28**

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: A mix of information from official Thailand reports (IPST, 2017; OECD & IPST, 2018) and own calculation using 2015 microdata

[141] TIMSS data is from 2011 due to lack of school size information in 2015

Gap by socioeconomic indicators

From Table 6-14, as expected, students with more books at home scored better than those with less books. For parental education, curiously, students whose parents attain lower levels of education[142] perform around the mean while students whose parents graduate lower secondary level perform meaningfully worse than the mean. For PISA, the difference between the two groups is not significant. Therefore, it is possible that this is noise in the data. In TIMSS, however, the difference is about 0.2 standard deviation. There is no obvious explanation why this is the case. Note also that this result is not replicated in other countries. In the OECD official report, being in higher socioeconomic status (using PISA index of economic, social and cultural status) is related to having higher scores (OECD, 2016a).

[142] Only getting some levels of primary, did not attend school, or did not graduate lower secondary.

Table 6-14 Summary of ranking by socioeconomic indicators

	PISA		TIMSS	
	Math	**Science**	**Math**	**Science**
Number of books at home				
Highest rank	More than 200 books			
Dif from mean	**52**	**41**	**78**	**71**
Lowest rank	0-10 books			
Dif from mean	**-21**	**-25**	**-29**	**-29**
Parental education				
Highest rank	Upper secondary or above			
Dif from mean	11	13	**31**	**28**
Lowest rank	Lower secondary			
Dif from mean	**-18**	**-19**	**-21**	**-20**

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA and TIMSS 2015 microdata

Gap by gender

In contrast with other subgroups, male and female students perform around the mean. Hence, gender gap may not be as concerning as gaps by other school characteristics.

Table 6-15 Summary of ranking by gender

	PISA		TIMSS		O-NET	
	Math	**Science**	**Math**	**Science**	**Math**	**Science**
Highest rank	Male		Female			
Dif from mean	0	0.06	0.09	0.11	0.10	0.11
Lowest rank	Female					
Dif from mean	-0.01	-0.02	-0.09	-0.06	-0.11	-0.12

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

6.4.4. Gap by item characteristics

Table 6-16 shows percentage of students answering the questions correctly by item types[143]. The average percent correct is less than half in both tests in all categories. Overall, students perform better in simple multiple-choice questions compared to constructed response questions. The differences are large, with about twice more students being able to answer simple multiple-choice than they can in constructed response question. This can be attributed to the Thai education placing less emphasis on constructed response type of question. As seen from O-NET exam format (0), the national test has put heavy emphasis on simple multiple-choice type of question (80% of total score). In Science exams, all questions are multiple choice. With O-NET being a high-stake examination, it is likely that teachers and students would prepare towards it. If we want to see improvement in more complex questions, these types of questions should be more emphasised in O-NET.

[143] Only trend items are used in this section.

Table 6-16 Percent correct by item types

Item types	Mathematics		Science	
	PISA	TIMSS	PISA	TIMSS
Simple multiple-choice	42%	40%	47%	47%
Complex multiple-choice	34%		39%	
Constructed response	23%	21%	24%	25%
Average	29%	31%	36%	36%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b) and TIMSS (TIMSS & PIRLS International Study Center, 2017)

Table 6-17 shows percent correct by cognitive domains[144]. TIMSS provide additional information on scores by cognitive domain, which account for scores based on all exam items, not just the trend items, as shown in parentheses in the table. Additionally, as the scores were calculated based on item-response theory, different weights are given to questions with differing level of difficulty. The table shows that the highest percent correct for both PISA and TIMSS is in the lowest cognitive domain of Formulate/Explain/Knowing. This shows that Thai students did well in questions requiring lower-level skills. Additionally, in Science, TIMSS scores of the Knowing domain is higher than the average (though not making the cut-off of 0.2 standard deviation difference), as well as significantly higher than the lowest category score of Reasoning.

The picture is less clear which of the two higher-level cognitive domains students perform better in. In PISA, students perform worst in the highest domain of ‘Interpret’. The gap is higher in Mathematics than Science. This shows students perform better in lower-level skills than they do in higher-level that requires more critical thinking. In TIMSS, the picture is less clear cut. Curiously, percent correct is higher in Reasoning domain compared to Applying, which should be of higher skill requirement. The scores also show that in Mathematics, students perform best in Reasoning and worst in Knowing. Nonetheless, the score differences are not

[144] O-NET is not included here because item-level scores are not available.

meaningful, and the differences may be simply noise. In Science, performance in Applying and Reasoning are also not statistically different from each other. Performance in Knowing, however, is significantly higher than both Applying and Reasoning. With this, a general conclusion can be drawn that Thai students perform better in lower-level skills. There remain improvements that can be made on the higher-level skills.

Table 6-17 Percent correct by cognitive domains

PISA's cognitive domains	TIMSS's cognitive domain	Mathematics		Science	
		PISA	TIMSS	PISA	TIMSS
Formulate/Explain	Knowing	38%	37% (425)	40%	46% (469)
Employ/Evaluate	Applying	31%	22% (431)	37%	25% (450)
Interpret	Reasoning	18%	31% (435)	32%	35% (447)
Average		29%	31% (431)	36%	36% (456)

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b) and TIMSS (TIMSS & PIRLS International Study Center, 2017). TIMSS scores are taken from IPST (2017)

Lastly, percent correct show slightly different picture on the performance of Thai students by each content domain (Table 6-18 and Table 6-19). In Mathematics, students' performance on Quantity/Numbers are above average in PISA. In TIMSS, the domain that the students perform well in is instead Data and chance. In Science, students perform better in Physical systems in PISA and in Biology in TIMSS. This discrepancy may be due to the differences in contents tested. For instance, students may be able to answer basic questions about statistics in TIMSS but fell short on more difficult questions asked in PISA.

Table 6-18 Percent correct by content domains in Mathematics

PISA's content domains	TIMSS's content domains	PISA	TIMSS
Quantity	Numbers	37%	32% (430)
Change and relationships	Algebra	26%	26% (429)
Shape and space	Geometry	22%	28% (429)
Uncertainty and data	Data and chance	30%	41% (425)
Average		29%	31% (431)

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b) and TIMSS (TIMSS & PIRLS International Study Center, 2017). TIMSS scores are taken from IPST (2017)

Table 6-19 Percent correct by content domains in Science

PISA's content domains	TIMSS's content domains	PISA	TIMSS
Living systems	Biology	32%	38% (466)
Physical systems	Chemistry	43%	36% (445)
	Physics		31% (437)
Earth and space systems	Earth Science	37%	40% (459)
Average		36%	36% (456)

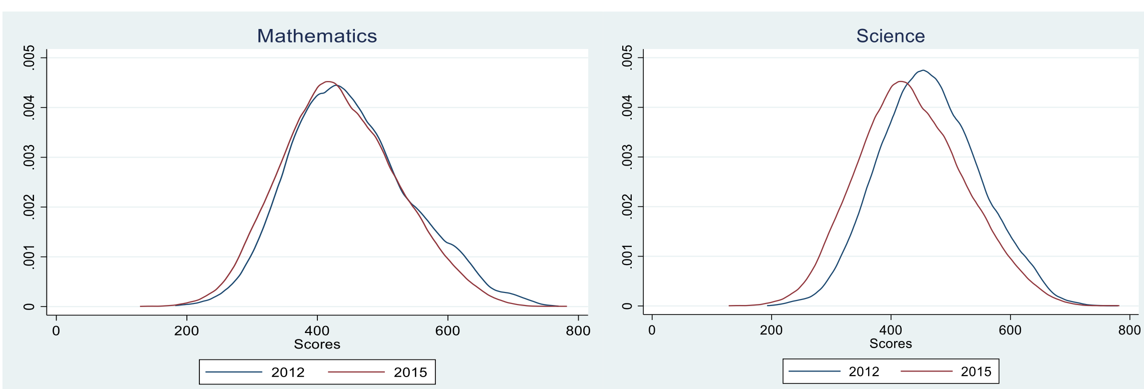
Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b) and TIMSS (TIMSS & PIRLS International Study Center, 2017). TIMSS scores are taken from IPST (2017)

In short, section 5.4 shows gaps in learning outcomes by different subgroups. Gaps are large and concerning. There are groups of high-performers scoring sometimes well above the OECD average. At the same time, many low-performing groups of students scored below a minimum acceptable standard. The largest gaps observed are by school types, with Science and Satit schools' performance matching those of the top countries such as Singapore.

6.5. Differences in trends over time

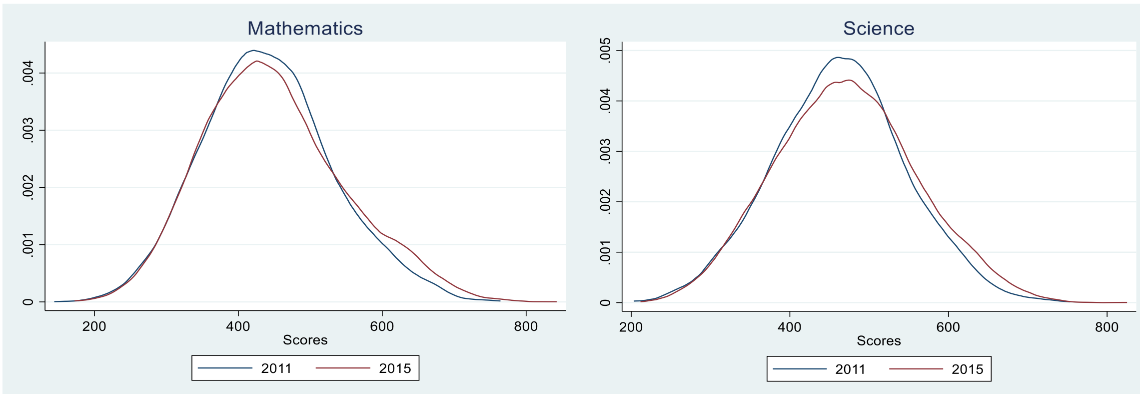
To show how performance has changed over time for students in different parts of the score distribution, k-density graphs are produced below, comparing proportion of students at different points of scores. For PISA, as shown in Figure 6-11, the distribution does not change much in Mathematics, reflecting no significant change in score in this period. For Science, the distribution changed for students at the lower and the middle end of the performance distribution. In 2015, there are visibly more students performing below 400 (approximately cut off point for proficiency level 2). For TIMSS, as shown Figure 6-12, both subjects have slight changes in the distribution. There is virtually no change at the bottom end of the distribution. Meanwhile, there are more students at the top end of the distribution and less at the middle. However, by our threshold, this change is too small to be meaningful. For O-NET (Figure 6-13), both subjects' distribution shifts more to the right, but with greater magnitude in Mathematics. In short, the figures show that score distributions changed meaningfully for subjects and assessments with significant change over time (PISA Science and O-NET Mathematics), while the rest have similar distribution across the years. There are clear shifts in the distributions around the middle in PISA. In O-NET, there are more students scoring more than 50%.

Figure 6-11 PISA score distribution in 2012 and 2015



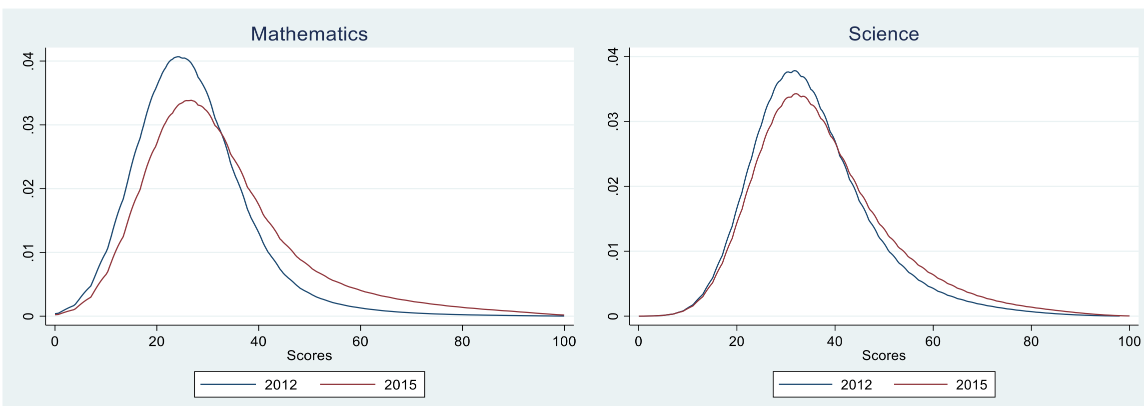
Source: Own calculation using PISA microdata

Figure 6-12 TIMSS score distribution in 2011 and 2015



Source: Own calculation using TIMSS microdata

Figure 6-13 O-NET score distribution in 2012 and 2015



Source: Own calculation using O-NET microdata

This section explores changes in gaps of learning outcomes. Next, differences in trends over time of each group are presented to identify which groups are making more (or less) progress than others.

6.5.1. Change in gaps

Table 6-20 shows change in the gap between the highest performing group and the lowest performing group. Over time, most of the gaps remain roughly the same size, with the changes not exceeding our 0.2 standard deviation threshold. Those that changed significantly, however, are almost all positive changes, and most are in O-NET. This means that the gap increased over time, and inequality levels are higher in 2015 than they were in 2011/2012. This trend shows that Thailand has not made many improvements towards equalising learning outcomes over time, which can be key in improving average outcomes.

By contrast, region subgroup in TIMSS is the only subgroup with decreased level of inequality. However, at a closer inspection, this decrease in gap is from the top-performing region, East, having large decline in score, while the bottom performer, Lower Northeast, has similar learning levels compared to previous round. Hence, even though the inequality had decreased, this is arguably not the desired outcome for Thailand, as the decrease is from the top-performers scoring lower and have more similar learning levels with the low-performers.

One of the most striking findings from the table is the increase in inequality among school types. Except for PISA Mathematics, other assessments show significant increase in gaps by school types. In O-NET, the increase is mainly from a dramatic increase in score of Science schools (1.3 and 2.5 standard deviation in Mathematics and Science respectively). In TIMSS, the increase is also from including Science schools in 2015 (previously, Science schools were not included in the sample and Satit schools were the top-performers). In PISA, by contrast, the increase in gap is from the bottom performer, BMA schools' large decline in scores, while Science schools' performance remain the same. The sources of the increase in gap are worrying, as it shows the low performers are struggling to keep up with the top performers.

Despite the changes in gap, virtually the same groups of students remain the top and bottom performers. Other than Science schools that were added in TIMSS

2015, other top/bottom performers remain quite stable. Groups such as urban schools and students with the greatest number of books at home maintain their ranking. Some subgroups where the performance is not meaningfully different among the groups see some shift. For example, PISA's medium-sized schools are the bottom performer in 2012 while small schools are the bottom performer in 2015. Nonetheless, the difference in scores between the two school sizes are only marginal (2-3 points). Therefore, the change in ranking is not meaningful. Similar with rankings by regions, where both top and bottom performers changed for PISA and TIMSS. This is probably because most regions do not perform differently from the mean. The change is not of a concern. Overall, we observe no large change between rankings of groups. This suggests inequality remains a pressing problem in the Thai education, but it has not changed much over time.

Table 6-20 Change in gaps in learning outcomes by each subgroup

Subgroups	Subjects	PISA	TIMSS	O-NET
School types	Math	-0.05	**0.82**	**1.90**
	Science	**0.36**	**0.83**	**0.97**
School types – excluding Satit and Science schools	Math	-0.19	0.05	**0.22**
	Science	-0.10	0.03	0.07
Regions	Math	0.16	**-0.30**	**0.24**
	Science	0.15	**-0.22**	0.05
Urban-rural	Math	-0.16	**0.31**	**0.22**
	Science	0.02	**0.30**	0.07
School size	Math	-0.15	N/A	**0.40**
	Science	0.13	N/A	0.14
Performance percentile (P90-P50)	Math	-0.08	0.13	**0.70**
	Science	0.11	0.07	**0.32**
Performance percentile (P50-P10)	Math	0.06	-0.04	**0.23**
	Science	0.01	-0.04	0.00
Number of books at home	Math	-0.13	0.14	N/A

Subgroups	**Subjects**	**PISA**	**TIMSS**	**O-NET**
	Science	0.00	0.10	N/A
Parental education	Math	-0.09	0.04	N/A
	Science	0.02	0.02	N/A
Gender	Math	-0.08	0.01	0.19
	Science	-0.08	0.06	0.12
Item types	Math	0%	0%	N/A
	Science	9%	4%	N/A
Cognitive domain	Math	0%	-4%	N/A
	Science	2%	2%	N/A
Content domain	Math	1%	2%	N/A
	Science	0%	-1%	N/A

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

The following sub-sections look at how the gaps changes over time for each group of students. This in-depth analysis allows us to know who were driving the changes observed in the gaps in the previous table.

6.5.2. School types

Table 6-21 shows score change over time for each school type in all assessments. In PISA, OBEC1 schools show largest decline in learning outcomes compared to other school types. The table also shows the discrepancies in trends by assessment and subject. Within PISA and O-NET, all school types change in the same direction (negative for PISA but positive for O-NET) but in different magnitudes. The trends also differ by subject. In O-NET, Mathematics scores increased meaningfully for all school types, but in Science, only OBEC2 and Science schools showed significant increase. By contrast, the decrease is uniformly felt in Science, but only significantly decreased in some school types in

Mathematics. In TIMSS, only Private schools increased their scores significantly. Similar to the overall trends, this table gives a confusing picture of what changed over time as there is no agreement on the trends. Change over time of Science school is markedly higher than other school types in O-NET. This can be explained by changes made in school curriculum and admission process as discussed in section 5.3.1.

Table 6-21 Score change over time by school type

School type	Mathematics			Science		
	PISA	**TIMSS**	**O-NET**	**PISA**	**TIMSS**	**O-NET**
OBEC1	**-0.65**	0.00	**0.22**	**-0.96**	0.03	0.12
OBEC2	-0.11	-0.02	**0.46**	**-0.23**	-0.03	**0.22**
Private	-0.16	**0.27**	**0.48**	**-0.35**	**0.34**	0.18
BMA	-0.11	-0.02	**0.24**	**-0.33**	0.01	0.18
Local	-0.19	0.09	**0.22**	**-0.45**	0.09	0.00
Satit	**-0.38**	0.14	**0.91**	**-0.29**	**0.20**	0.16
Science	-0.17		**2.50**	0.03		**1.31**

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table 6-22 and Table 6-23 show change in score distribution by school type. There is larger change in the distribution of Science than Mathematics. This mirrors overall trends that Science shows significant decrease in performance while changes in Mathematics scores are not above the 0.2 standard deviation threshold. The decline in scores is marked by increase in number of students performing below level 2 in 2015. OBEC1 schools see largest change in scores, mirroring Table 6-21. In 2012, OBEC1 schools perform above the national average, but the fall in scores pushed them down to scoring below the national average in 2015. It is important to note that in other years, OBEC1 schools have always perform below the national average. Hence, 2012 is an outlier for OBEC1 schools.

Table 6-22 Change in proportion of students at different proficiency levels in PISA Mathematics, by school type

School types	Below level 1	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
OBEC1	12.9	13.7	-1	-12.4	-8	-4	-1.2
OBEC2	3.2	0	-0.3	-0.3	-1.3	-0.9	-0.5
Private	9.4	-3.6	-4.6	-0.7	-0.5	-0.1	0.2
BMA	4.3	-1.7	1	-1.2	-2.3	-0.2	N/A
Local	10.3	-1.4	-7.5	-2.2	0.6	0.2	0.1
Satit	2.4	5.3	4.4	0.6	-2.2	-7.2	-3.2
Science	-0.2	-0.1	1.2	7.7	1.4	-6	-3.9

Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table 6-23 Change in proportion of students at different proficiency levels in PISA Science, by school type

School types	Below level 1	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
OBEC1	-0.6	39.0	-4.0	-23.7	-8.5	-2.1	0.0
OBEC2	-4.0	13.9	-2.7	-5.0	-1.6	-0.6	0.0
Private	-7.2	24.7	-7.2	-10.4	1.2	-1.2	0.0
BMA	-11.3	24.3	-6.1	-6.0	-1.0	0.1	0.0
Local	-7.5	32.4	-15.4	-9.1	-0.4	0.2	0.0
Satit	-0.9	6.4	3.6	1.3	-7.6	-2.2	-0.8
Science	-0.2	0.1	-1.6	-1.5	4.2	-0.8	-0.2

Source: Own calculation using PISA, TIMSS, and O-NET microdata

6.5.3. School characteristics

For trends over time by school characteristics, there appears to be no general agreement among the assessments (Table 6-24 to Table 6-26). Bangkok and its perimeter, for example, did worse in TIMSS, no change in PISA, but increase in scores in O-NET. The disagreement in trends shows the complex picture of the nature of learning outcomes in Thailand over time. For policymakers, it is clear on who performs well and who do not, but it is difficult to know how the results fare over time.

In the case of rural and urban location, TIMSS and O-NET show similar trends, with urban students made greater improvement than their rural counterparts, increasing gaps in learning outcomes. This is undesirable, and policies should support rural students to catch up. In PISA, the trend differs between subjects. While the two groups show similar decline in Science, in Mathematics, urban students show greater negative changes than rural schools counterparts. Even though this brings down the learning gap, the change is also undesirable from the policy viewpoint.

Table 6-24 Score change over time by school region

	Mathematics			Science		
Region	**PISA**	**TIMSS**	**O-NET**	**PISA**	**TIMSS**	**O-NET**
BKK and perimeter	0.05	**-0.46**	**0.54**	-0.10	**-0.48**	**0.23**
Central	0.18	**0.33**	**0.38**	0.04	**0.28**	0.15
Upper north	**-0.29**	**0.26**	**0.48**	**-0.38**	**0.27**	**0.20**
Lower north	-0.12	-0.15	**0.38**	**-0.33**	-0.16	0.18
Upper NE	-0.10	0.05	**0.30**	**-0.24**	0.09	0.13
Lower NE	-0.17	-0.19	**0.30**	**-0.33**	-0.14	0.18
South	-0.16	**0.42**	**0.42**	**-0.32**	**0.42**	0.18
East	-0.19	**-0.32**	**0.44**	**-0.38**	**-0.24**	**0.21**
West	**-0.52**	-0.07	**0.42**	**-0.58**	-0.09	0.19

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: PISA and TIMSS taken from official Thailand reports (IPST, 2017; OECD & IPST, 2018) and own calculation using O-NET microdata

Table 6-25 Score change over time by school location

School location	Mathematics			Science		
	PISA	**TIMSS**	**O-NET**	**PISA**	**TIMSS**	**O-NET**
Rural	-0.07	-0.03	**0.24**	**-0.25**	-0.02	0.06
Urban	**-0.23**	**0.28**	**0.66**	**-0.23**	**0.28**	**0.37**

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table 6-26 Score change over time by school size

School size	Mathematics		Science	
	PISA	**O-NET**	**PISA**	**O-NET**
S	**-0.20**	**0.22**	**-0.46**	0.11
M	-0.05	**0.26**	**-0.29**	0.13
L	**-0.24**	**0.34**	**-0.30**	0.15
XL	**-0.25**	**0.62**	**-0.29**	**0.25**

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA and O-NET microdata

6.5.4. Performance percentile

Table 6-27 show whether the change in performance is driven by the top, median, or bottom performers. For Mathematics, the percentile score changed meaningfully only in O-NET and PISA. In PISA, the significant change happens only for the 90th percentile. In O-NET, it is for both 50th and 90th percentiles. In Science, however, PISA shows significant decrease in all percentiles. O-NET shows large increase just for the 90th percentile. Top-performing students' scores increased significantly in O-NET compared to students at other percentiles. This improvement has made performance gap even larger over time.

Despite the large gap, inequality changed little over time in PISA and TIMSS outcomes. O-NET, however, due to large increase of the top performers' scores, the gap increased significantly. Even though average scores show students have performed better, the increase is not equitable or uniform. Rather, it is concentrated in the top-performing groups of students. This is arguably not the direction Thailand wants to move towards.

Table 6-27 Score change over time by performance percentile

Performance percentile	Mathematics			Science		
	PISA	**TIMSS**	**O-NET**	**PISA**	**TIMSS**	**O-NET**
P10	-0.18	0.04	0.00	**-0.31**	0.07	0.16
P50	-0.12	0.00	**0.23**	**-0.32**	0.03	0.16
P90	**-0.21**	0.13	**0.93**	**-0.21**	0.10	**0.48**
Gaps						
P90-P10	-0.02	0.09	**0.93**	0.11	0.03	**0.32**
P90-P50	-0.08	0.13	**0.70**	0.11	0.07	**0.32**
P50-P10	0.06	-0.04	**0.23**	-0.01	-0.04	0.00

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

6.5.5. Socioeconomic indicators

Table 6-28 and Table 6-29 present score changes by numbers of books in the home category and parental education. For PISA, in Mathematics, most change over time is not statistically significant except for two higher groups. The decline in scores is greater for students with higher socioeconomic status (proxied using number of books at home, with higher number of books/parental education being associated with higher socioeconomic status). Even though this slightly narrows the gap between high and low socioeconomic groups, this is also an undesirable situation as learning outcomes fell more for the high socioeconomic groups. In

Science, however, the decline happens in all groups in PISA, whereas in TIMSS, there is generally no meaningful change except for the increase in score of the 101-200 books group.

Table 6-28 Score change over time by books at home category

Number of books at home	Mathematics		Science	
	PISA	TIMSS	PISA	TIMSS
0-10 books	-12	-2	**-28**	1
11-25 books	-14	3	**-26**	0
26-100 books	-12	9	**-17**	12
101-200 books	**-25**	13	**-35**	**20**
More than 200 books	**-23**	11	**-28**	10

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA and TIMSS microdata

Table 6-29 Score change over time by parental education

Parental education	Mathematics		Science	
	PISA	TIMSS	PISA	TIMSS
Some primary or lower secondary or did not attend school	-3	8	**-19**	6
Lower secondary	**-23**	-1	**-35**	1
Upper secondary or above	**-19**	3	**-25**	3

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA and TIMSS microdata

6.5.6. Gender

Even though gaps by gender do not change significantly over time, some of the individual scores changed significantly as shown in Table 6-30. Here, PISA shows decline in scores greater for female students, TIMSS shows no significant change, and O-NET shows greater increase for female students, again showing a confusing picture of what trend over time look like.

Table 6-30 Score change over time by gender

School location	Mathematics			Science		
	PISA	**TIMSS**	**O-NET**	**PISA**	**TIMSS**	**O-NET**
Female	-0.19	0.05	**0.49**	**-0.33**	0.08	**0.24**
Male	-0.11	0.04	**0.30**	**-0.26**	0.02	0.12

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

6.5.7. Item characteristics

In terms of performance by item type[145] (Table 6-31), the trends over time differ for PISA and TIMSS. In PISA, all change over time is negative. In Science, the decrease is most significant in constructed response questions. This may imply that students perform less well on higher skills items over time. By contrast, there is no significant changes in TIMSS's percent correct of the trend items, nor do the trends differ largely between item types.

[145] The statistics were calculated for trend items only as these are the items that can be compared between 2011 and 2015.

Table 6-31 Score change over time by item types

Item type	Mathematics		Science	
	PISA	TIMSS	PISA	TIMSS
Simple multiple-choice	-4%	1%	-2%	1%
Complex multiple-choice	-6%		-2%	
Constructed response	-4%	1%	-11%	-1%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b) and TIMSS (TIMSS & PIRLS International Study Center, 2017)

As for performance by cognitive domains/capabilities, Table 6-32 shows that in PISA in Mathematics, percent correct of each item shows similar decrease in each domain. In Science, students appear to perform worse in Interpret domain over time, with a reduction in percent correct of 7%. Meanwhile, TIMSS shows no significant change over time. Percent correct is highest for the Knowing domain and lowest in the Applying domain. It seems that, particularly in Science, students perform worst in higher level skills over time.

In TIMSS, scores were also reported separately for each cognitive domain (as shown in parenthesis). This shows slightly different picture of achievement. Most sections show positive change in scores. However, the magnitude is not meaningful except for changes in score of Science in the Knowing domain where the change is greater than 2 standard deviation. A closer inspection of the trend items shows some questions in the Knowing domain that changed significantly over time. For example, a question named ‘Good source of calcium’ has large change in percent correct, from 18.7% in 2011 to 86.2% in 2015. Hence, it is possible the change in scores of the Knowing domain is primarily driven by questions like these. Additionally, the increase only in the Knowing domain may suggest tutoring. This cannot be confirmed from the data but can be explored in the qualitative part of the research.

Table 6-32 Score change over time by cognitive domain

PISA's cognitive domains	TIMSS's cognitive domain	Mathematics		Science	
		PISA	TIMSS	PISA	TIMSS
Formulate/Explain	Knowing	-4%	0% (+2)	-5%	1% **(+26)**
Employ/Evaluate	Applying	-4%	2% (+3)	-4%	0% (-1)
Interpret	Reasoning	-4%	1% (+6)	-7%	0% (-6)

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b) and TIMSS (TIMSS & PIRLS International Study Center, 2017)

As shown in Table 6-33 and Table 6-34, change over time seems to be more uniform across content domains than it is in cognitive domains. There is similar reduction in scores over time in PISA, with the largest change being a reduction in Biology. In TIMSS, most domains show no change in percent correct as well as scores. Increase in Geometry scores is the largest, yet it still falls below our 0.2 standard deviation threshold. Therefore, change in performance by content domains are not as significant as change by cognitive domains and item types.

Table 6-33 Score change over time by content domain for Mathematics

TIMSS's content domain	PISA's content domain	PISA	TIMSS
Numbers	Quantity	-5%	1% (+5)
Algebra	Change and relationships	-2%	1% (+4)
Geometry	Shape and space	-5%	0% (+14)
Data and chance	Uncertainty and data	-3%	1% (-6)

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b) and TIMSS (TIMSS & PIRLS International Study Center, 2017)

Table 6-34 Score change over time by content domain for Science

TIMSS's content domain	PISA's content domain	PISA	TIMSS
Biology	Living systems	-7%	1% (+6)
Chemistry	Physical systems	-4%	0% (+9)
Physics			0% (+7)
Earth Science	Earth and space systems	-5%	0% (-7)

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b) and TIMSS (TIMSS & PIRLS International Study Center, 2017)

6.6. Discussion

In this chapter, results show how the learning outcomes of PISA, TIMSS, and O-NET are similar and different. Key issues are discussed below.

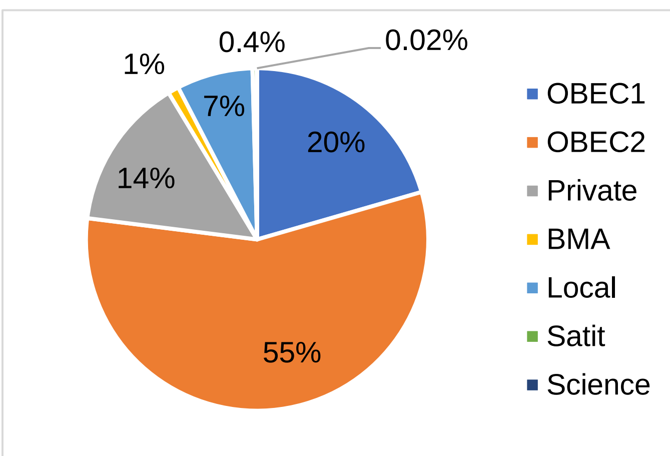
Firstly, there are differences in the contents tested among the three assessments that can potentially drive differences in trend results. TIMSS and O-NET show some general alignment in terms of contents tested and coverage of topics. Almost all topics tested in TIMSS are included in the Thai national curriculum. At the same time, arguably, TIMSS tests the contents that require higher level skills than O-NET. There are more percentage of constructed response questions in relation to simple multiple-choice questions. Additionally, in Mathematics, TIMSS tests less of Knowing domain than O-NET. By contrast, PISA is quite different from TIMSS and O-NET. The assessment shows more focus on the higher-level skills. Both cognitive and content domains extend beyond what is tested in TIMSS and O-NET. This implies PISA tests higher-level skills than the other two assessments. This aligns with Wu's (2010) analyses that TIMSS places more importance on solving routine problems rather than reasoning, which requires higher skill level. It also aligns with TIMSS's assessment framework and objectives to focus on curricular contents and simple applications of knowledge (Mullis & Martin, 2013).

It is likely that this difference in content tested affect the score differential and trend differences observed. As seen from analysis of item characteristics, students perform worse in types of items requiring higher level of skills (constructed response VS multiple-choice questions or Knowing VS Reasoning). Additionally, over time, particularly in Science, students seem to perform worse in PISA's higher level items (Interpret and constructed response questions) and better in TIMSS's lower level items (Knowing). This may suggest that overall, Thai students perform worse higher-level skills and better in lower-level skills. This may result in a decline in PISA performance, which tests higher skills, and increase in O-NET performance, which tests lower level of skills.

Going beyond using this insight to describe trends, knowing what skills Thai students are good at has many policy implications. It is not surprising that students are good at lower-level skills such as remembering or understanding concepts as they are heavily emphasised in both the national curriculum and assessment. Even though these lower-level skills are useful in themselves, as well as serving as a foundation for attaining higher-level skills, it is becoming clear that having only these skills is not enough to ensure students perform well in international assessments like PISA and TIMSS. Policymakers should reflect on what skills Thai students should have when they finish compulsory education, and whether that includes skills tested in PISA or not. Currently, it is clear that Thailand wants to improve PISA results, partly to signal good education quality and increase international competitiveness (section 1.2). To improve the scores, the Thai education system needs to shift its emphasis to teach higher-level skills. Arguably, this should be done via O-NET and the national curriculum, that Thailand has full control of. As O-NET is a national assessment that stakeholders are accountable to, it can be used to ease stakeholders into incorporating desired skills in their lessons.

Secondly, results show Thailand has a high level of inequality. All assessments show this to a certain level, with some reporting more unequal results than others. Learning outcomes are far from being uniformly low. Gaps in learning outcomes are high across many subgroups. Most notably, gaps among school types are large and significant. There are both groups of schools/students that perform meaningfully above and below the mean. This corresponds with the Thai media's portrayal of learning levels in Thailand (Mala, 2016; Thairath, 2018) and with academic research (Phuaphansawat, 2021). It is true that top schools (Satit and Science schools) perform very well in international assessments, in line with Singapore and Japan. Despite the high performance, average learning levels in Thailand remain low as Satit and Science schools only account for less than 1% of total enrolment (see Figure 6-14). It is unclear how much of the advantage is from selection: accepting students from better socioeconomic or academic backgrounds, or from better teaching quality. Knowing this would be beneficial for policy as we can dissect what can be learned from them or what practices can be applied to other schools

Figure 6-14 Proportion of grade 9 students as of 2015



Source: NESDC (2019)

The proportion also shows that the majority of pupil enrol in OBEC2 schools. Looking back to section 5.4.1, even though the learning levels are not at the same level with the high performers, they are arguably satisfactory. Most students are at baseline level 2. When Satit and Science schools are excluded, OBEC2 schools become the top performer. Arguably, if the low performers can be brought up to have learning levels similar to OBEC2 schools, the average scores will improve. This can be a more practical and achievable goal that Thailand can strive towards. The next logical question would be what kind of policies can be targeted to the low performers to help them improve. The key lies in the extent that these schools differ from OBEC2 schools and the high performers of Satit and Science schools, which was explored in the previous chapter.

Thirdly, trends over time show a confusing picture of learning outcomes. There is less agreement on the extent to which scores and inequalities have improved or worsened over time. For most subgroups, gaps in learning outcomes show little change over time. For some subgroups, however, inequality seems to be increasing over time, especially among school types. The increase in gap is driven primarily by the top-performers making more progress than the bottom performers. Arguably, this is not a desirable outcome for the education system. Thailand should strive towards more equitable outcomes where the low performers are supported to improve to be up to an acceptable standard. In cases where inequality declines, such as among regions, it is because the top-performing regions scored much worse over time, narrowing the gap. This is also not a desirable outcome.

Lastly, despite the differences in the contents tested, the three assessment results show striking similarities in terms of who the top and bottom performers are. This alignment supports comparability of the three assessments.

Chapter Seven: What Contribute to Change Over Time in Scores?

7.1 Introduction

This chapter presents a systematic investigation of what potentially affect observed changes over time in scores and how the discrepancies can be reconciled. Firstly, the comparison is made between school and student samples of each assessment over time in various characteristics. The comparison is also made for item-level characteristics, to see if the assessments are testing similar things in the two cycles. Next, modelling results are presented to show whether changes over time are affected by changing sample composition or not.

7.2 Quantitative results: Descriptive statistics

Average score trends can be influenced by changing sample characteristics over time. It is especially important when the characteristics are very predictive of scores as demonstrated by Aloisi and Tymms (2017). Hence, this section explores changes in elements of sampling and test items The descriptive statistics in this section closely mirror those presented in Chapter Six. They are presented here with emphasis on change over time to answer research question 3.

7.2.1 School type proportion

As shown in Table 7-1, proportion of students enrolled in each school type changed little over time from 2012 to 2015. There are a smaller number of students enrolled in all grade levels, which reflect a smaller population overall. The decline is not uniform, with 5% decline in grade 8, 10% decline in grade 9, and 12% decline in grade 12. Otherwise, the proportion of enrolment by school type remains roughly the same. Therefore, it is unlikely that the change in scores is driven by changes in enrolment of students.

Table 7-1 Enrolment rates by school types

School type	Grade 8		Grade 9		Grade 10	
	2012	2015	2012	2015	2012	2015
OBEC1	22%	22%	22%	20%	-	-
OBEC2	54%	54%	55%	55%	79%	78%
Private	13%	14%	13%	14%	13%	13%
BMA	1%	1%	1%	1%	0.2%	0.2%
Local	7%	7%	7%	7%	6%	6%
Satit	1%	0.4%	1%	0.4%	1%	1%
Science	N/A[146]	N/A	N/A	N/A	N/A	N/A
Student number	813,900	770,226	854,937	770,874	517,284	453,588

Source: NESDC (2019)

Proportion of students who took O-NET and PISA[147] also remains is roughly the same over time. Overall changes are less than 3%, as shown in Table 7-2. Hence, it is not likely that the trends in assessment scores are caused by changing composition of school type included. While this does not give us clarity on why there are discrepancies, it shows that PISA sample is robust and does not change over time. This gives confidence that the trend result is meaningful.

[146] Science schools are listed as OBEC2 schools and not as a separate category.
[147] Similar to section 6.2.2, TIMSS is excluded as it does not provide sampling frame.

Table 7-2 Proportion of each school type in O-NET and PISA samples

School type	O-NET		PISA	
	2012	2015	2012	2015
OBEC1	22%	21%	8%	9%
OBEC2	55%	55%	73%	72%
Private	14%	15%	11%	12%
BMA	1%	1%	0%	1%
Local	7%	7%	6%	6%
Satit	1%	1%	1%	1%
Science	0.3%	0.2%	0.4%	0.3%

Sources: PISA percentages based on official report (OECD & IPST, 2018), O-NET percentages based on own calculation of microdata

As discussed in the previous chapter, PISA proportion of students look different from O-NET because it includes a mix of students from different grade levels. Table 7-3 shows the proportion of students at different grade levels in 2012 and 2015. It seems that there are more grade 9 students tested in 2015, which lowers the scores slightly as grade 10 students perform significantly better than grade 9. Nonetheless, the change is small and does not affect the scores much.

Table 7-3 Proportion of each grade level in PISA

Grade	2012	2015
Grade 7	0%	0%
Grade 8	0%	1%
Grade 9	26%	29%
Grade 10	71%	67%
Grade 11	3%	3%

Source: Own calculation based on PISA microdata

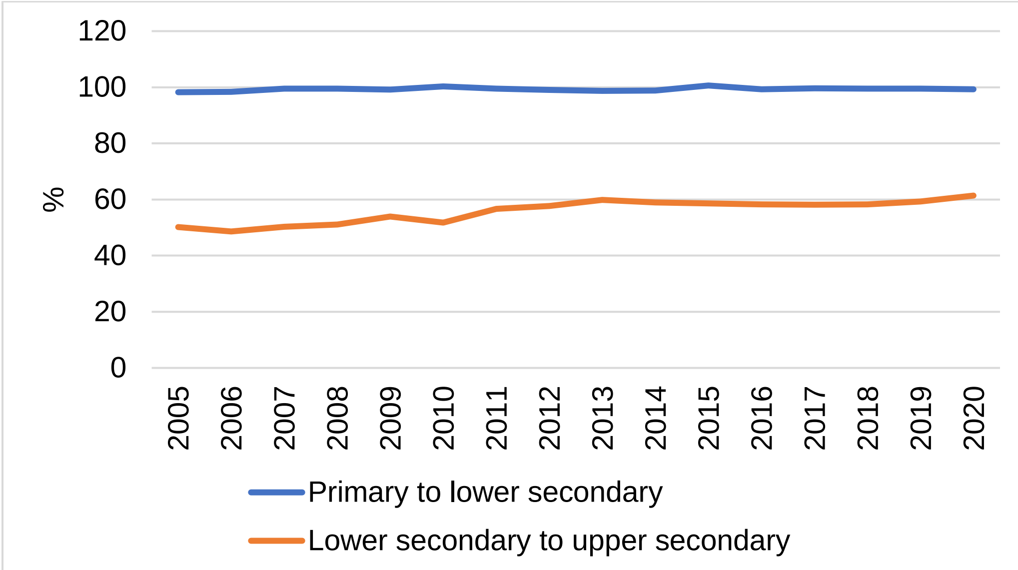
Data here shows that proportion of school types included does not seem to be driving large changes in scores. This is in contrast to a claim made by a Thai academic, who asserted that Science schools drove the improvement in 2012, and subsequently caused the decline as they do not participate in 2015 (Mala, 2016). Data shows that this is false. Science schools participated in both rounds of PISA. Additionally, as the proportion of enrolment in Science schools is so low compared to other schools, excluding them does not change the scores significantly. This shows that there remain misunderstandings surrounding what drive PISA results, and that the media and expert opinions can sometimes contradict the data as seen in Morsy et al. (2018). It is important to then fact-check what the media reports.

7.2.2 Transition rates

Monitoring transition rates from lower to upper secondary may also be relevant to our analysis. O-NET and TIMSS test lower secondary students while PISA samples both lower and upper secondary. Changes in transition rate can affect characteristics of students in PISA sampling frame. As access to education improves, transition rates increase over time. Consequently, greater proportion of school-age children and youth have access to education. These students are usually from poorer socioeconomic background and would not have access before. Hence, greater enrolment and transition rates can affect scores negatively.

Figure 7-1 shows transition rates over time both from primary to lower secondary and from lower to upper secondary (NESDC, 2019). Transition rate from primary to lower secondary has been virtually universal from 2005 onwards. There is little variation over time. The transition rate of our interest, from lower to upper secondary, shows a slight increase over time, but remain stable at around 60%. Specifically, at the period of focus, 2011 to 2015, there is only around 1% of change. Hence, this could not explain the decline in scores either.

Figure 7-1 Transition rates from different stages of education over time



Source: NESDC (2019)

7.2.3 Socioeconomic status

Data is available for comparing group composition of rural and urban students, number of books in the home, and parental education for each assessment. The proportion of each group varies somewhat between surveys as presented below.

For the proportion of rural and urban students (Table 7-4), O-NET data shows quite consistency over time, with somewhat increased percentage of urban students over time. Both TIMSS and PISA data sampled more students from rural schools than O-NET. Particularly, PISA sampled more rural students over time, by 12%. The changing proportion, however, affects the score trend only minimally, with changes of 5 points when the proportion of 2015 was reweighted.

Table 7-4 Proportion of rural and urban students

	2011	2012	2013	2014	2015
TIMSS					
Rural	74%				76%
Urban	22%				24%
PISA					
Rural		69%			79%
Urban		31%			21%
O-NET					
Rural	63%	63%	63%	62%	62%
Urban	37%	37%	37%	38%	38%

Source: Own calculation using PISA, TIMSS, and O-NET microdata

Number of books in the home proportion (Table 7-5) changed little over time as well and is unlikely to affect score trends.

Table 7-5 Proportion of students in terms of books at home

	PISA			**TIMSS**		
Books at home	2012	2015	Change	2011	2015	Change
0-10 books	19%	21%	+2%	28%	25%	-3%
11-25 books	33%	34%	+1%	45%	48%	+3%
26-100 books	31%	30%	-1%	20%	19%	-1%
101-200 books	9%	8%	-1%	5%	4%	-1%
More than 200 books	7%	6%	-1%	3%	3%	0%

Source: Own calculation using PISA, TIMSS, and O-NET microdata

Parental education composition as shown in Table 7-6 and Table 7-7 shows larger and quite significant change over time. Parents of students taking both tests are more educated in 2015 than in 2012. Less parents completed only primary level of schooling and the majority now have secondary education. Disparity between mother's and father's education is less in 2015 as well, showing less inequity in enrolment over time.

Table 7-6 Proportion of students in terms of mother's education

	PISA			TIMSS		
Mother's education	**2012**	**2015**	**Change**	**2011**	**2015**	**Change**
Some primary or lower secondary or did not attend school	48%	36%	-12%	35%	14%	**-22%**
Lower secondary	14%	19%	5%	33%	45%	12%
Upper secondary or above	38%	45%	7%	32%	41%	9%

Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table 7-7 Proportion of students in terms of father's education

	PISA			TIMSS		
Father's education	**2012**	**2015**	**Change**	**2011**	**2015**	**Change**
Some primary or lower secondary or did not attend school	41%	36%	-5%	29%	12%	-18%
Lower secondary	14%	18%	4%	32%	43%	11%
Upper secondary or above	44%	46%	2%	39%	45%	7%

Source: Own calculation using PISA, TIMSS, and O-NET microdata

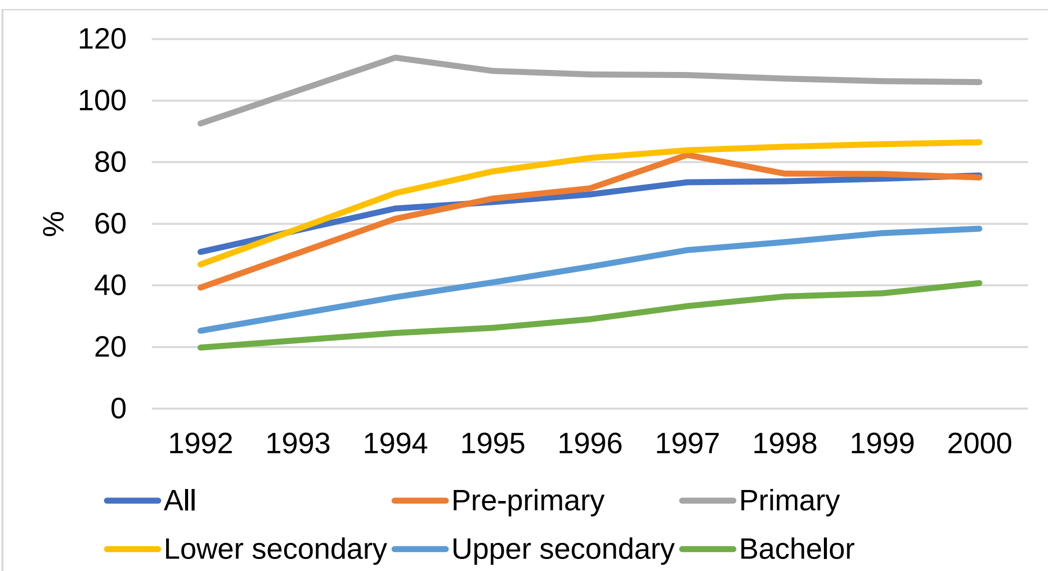
Change in patterns of enrolment rates can partially explain why the change in parental education is large. Assuming the parents are 40 years old by the time their children are 15, then they would have been in compulsory education during the 1990s. As Table 7-8 and Figure 7-2 show, during that time, Thailand was expanding access to education rapidly. Large changes of 20% were recorded during 1992 and 1994. As a result, parents born in different years may have very different level of schooling. This might account for why parents of the 2015 PISA and TIMSS cohort have significantly higher level of schooling.

Table 7-8 Gross enrolment rate over time

Level	1992	1994	1995	1996	1997	1998	1999	2000
All	51%	65%	67%	70%	74%	74%	75%	76%
Pre-primary	39%	62%	68%	72%	82%	76%	76%	75%
Primary	93%	114%	110%	109%	108%	107%	106%	106%
Lower secondary	47%	70%	77%	81%	84%	85%	86%	86%
Upper secondary	25%	36%	41%	46%	51%	54%	57%	58%
Bachelor	20%	25%	26%	29%	33%	36%	37%	41%

Source: NESDC (2019)

Figure 7-2 Gross enrolment rate over time



Source: NESDC (2019)

Another source of data, the household socioeconomic survey, also shows gradual improvement in parental education (Table 7-9). The survey has been collected yearly from 2006 onwards and is sampled to be representative of the Thai population in all 77 provinces. One information collected is the education level of the head of the household. The improvement is in the same direction as shown in the enrolment data and the questionnaires from PISA and TIMSS, but at a smaller magnitude[148].

Table 7-9 Household head education level from the National Household Socioeconomic Survey

Education level	2011	2012	2013	2014	2015	Change 2015-2011
No schooling	6%	6%	6%	5%	5%	-0.6%
Primary	64%	61%	62%	58%	58%	-6.5%
Secondary	16%	18%	18%	19%	20%	3.6%
Vocational certificate	5%	6%	6%	6%	7%	1.2%
Bachelor's	7%	8%	8%	9%	9%	1.8%
Above Bachelor's	1%	1%	1%	2%	2%	0.4%
Others	0%	0%	0%	0%	0%	0.1%

Source: NSO (2019)

Data shows that socioeconomic status improves over time in terms of parental education. More parents graduated higher levels of education. However, this seems to be a natural progression from the expansion of access to education. Whether this change affects the score much depends on whether it changes behaviours of parents towards their children’s education. From the data, it seems that the score differential for both TIMSS and PISA does not change much when

[148] This might be because not everyone in the household is asked the question, just the household head. Therefore, it might not be representative of overall parental education. Additionally, around 65% of household heads who answered the survey are male.

they were reweighted. Hence, it seems changes in socioeconomic background does not explain the change in scores.

7.2.4 Change in school type characteristics

Since school type is the variable most predictive of test scores, changes in characteristics of school types could influence score change. This section investigates if characteristics of each school type had changed over time[149].

Table 7-10 shows proportion of school sizes by each school type. The largest change is Science schools, which became smaller. This resonates with the change in policy in 2011, when the school admits smaller number of students and limit class size. OBEC1 schools seem to become smaller over time while Satit and Local schools became larger. As previously discussed, in Thailand, being larger schools come with many advantages such as ability to attract more qualified teachers and being allocated more budget. This implies that OBEC1 schools may be more disadvantaged in 2015 than they were in 2012. Satit schools, by contrast, became larger and accepting more enrolment.

[149] Due to data limitation, this comparison can only be done using O-NET data. PISA school type data is only available in 2015. Hence, subgroup analysis can only be done in 2012 and cannot be compared over time.

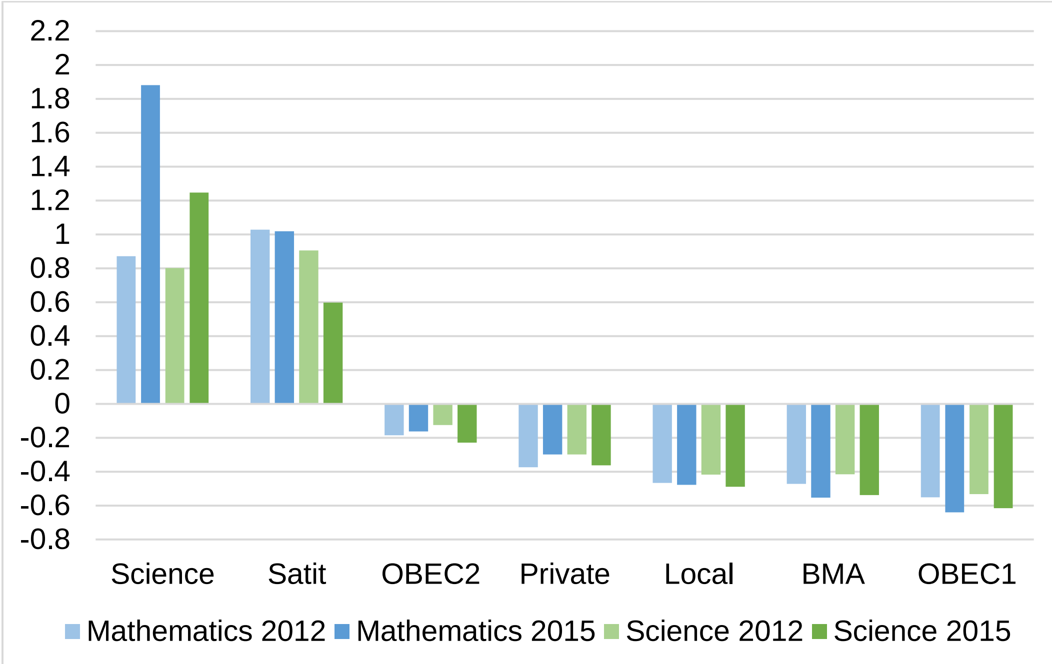
Table 7-10 Change in proportion of school size by school type

School type	Small	Medium	Large	Extra-large
OBEC1	7%	-8%	1%	0%
OBEC2	1%	-2%	-1%	2%
Private	-1%	0%	0%	1%
BMA	1%	3%	-3%	0%
Local	1%	0%	-4%	4%
Satit	0%	-2%	-4%	6%
Science	0%	31%	-31%	0%

Source: Own calculation using O-NET microdata

Prior scores of students show little change over time in most school types (Figure 7-3), except for Science and Satit schools. Reflecting the change in Science schools, the figure shows they admit students with significantly higher prior scores. Satit schools see students with lower prior Science scores. It is possible that the students with better scores chose to enrol in Science schools instead of Satit schools. This shift in enrolment may have contributed to Science schools outperforming Satit schools in O-NET 2015.

Figure 7-3 Prior score by school type over time



Source: Own calculation using O-NET microdata

In contrast with the previous characteristics, percentage of students continuing to general track does not change much over time (less than 2% changes in all school types). It implies that students pursuing different tracks after finishing compulsory education remains the same for all school types, with students in high-performing schools more likely to continue to universities.

7.2.5 Change in item characteristics

There are small changes in the formats of the items over time, which are unlikely to meaningfully influence scores (Table 7-11 and Table 7-12). In Mathematics, PISA and TIMSS's item distributions remain roughly the same. O-NET changed some questions from complex multiple-choice to constructed response. In Science, the proportion of item types remain roughly the same. PISA reduced the proportion of simple multiple-choice and increased complex multiple-choice.

Table 7-11 Mathematics item type distribution

Item type	PISA		TIMSS		O-NET	
	2012	2015	2011	2015	2012	2015
Simple multiple-choice	25%	22%	54%	55%	80%	80%
Complex multiple-choice	14%	16%	-	-	20%	-
Constructed response	61%	62%	46%	45%	-	20%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b), TIMSS (TIMSS & PIRLS International Study Center, 2017), and O-NET (NIETS, 2021)

Table 7-12 Science item type distribution

Item type	PISA		TIMSS		O-NET	
	2012	2015	2011	2015	2012	2015
Simple multiple-choice	34%	29%	51%	58%	80%	80%
Complex multiple-choice	30%	36%	-	-	20%	20%
Constructed response	36%	35%	49%	42%	-	-

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b), TIMSS (TIMSS & PIRLS International Study Center, 2017), and O-NET (NIETS, 2021)

Proportion of items by cognitive domain shows larger change than item type (Table 7-13 and Table 7-14). In TIMSS Mathematics, there is a slightly greater focus in Applying compared to Knowing over time. In O-NET, the change is in another direction, with emphasis changing from the second domain to the lowest domain. The reduction is 10% of total items. In Science, PISA increased questions on the easiest domain and reduced questions on Interpret domain. O-NET increased number of items in the highest domain and decreasing the items in the lowest domain. This shows that O-NET may have become more difficult over time in Science and easier in Mathematics. As a result, it may be problematic in terms of trends comparison, particularly when the positive results in Mathematics could have been driven by increase in number of easier questions rather than increase in quality (from 2012 to 2015, Mathematics results show significant increase while

Science shows no meaningful changes). This shows that for trend comparison to be useful for policymakers, test difficulty should remain similar over time and should be factored in the test design process.

Table 7-13 Percentage of items in each competencies/cognitive domain in Mathematics

PISA's capabilities	**TIMSS's cognitive domains**	**PISA**		**TIMSS**		**O-NET**	
		2012	2015	2011	2015	2012	2015
Formulate	Knowing	32%	30%	36%	31%	33%	43%
Employ	Applying	43%	42%	39%	45%	37%	27%
Interpret	Reasoning	25%	28%	25%	24%	30%	30%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b), TIMSS (TIMSS & PIRLS International Study Center, 2017), and O-NET (NIETS, 2021)

Table 7-14 Percentage of items in each competencies/cognitive domain in Science

PISA's capabilities	**TIMSS's cognitive domains**	**PISA**		**TIMSS**		**O-NET**	
		2012	2015	2011	2015	2012	2015
Explain	Knowing	38%	48%	32%	36%	44%	33%
Evaluate	Applying	26%	21%	44%	41%	44%	42%
Interpret	Reasoning	36%	30%	24%	23%	11%	24%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b), TIMSS (TIMSS & PIRLS International Study Center, 2017), and O-NET (NIETS, 2021)

For all assessments, the proportion of questions by content domain seems to change little over time (Table 7-15 and Table 7-16). O-NET again shows the greatest change in proportion of items in each content, with more items of Biology and Physics. This is problematic as the change over time may be from the change in content emphasis. As found in a study by Carnoy and Rothstein (2013), when students perform much better in some contents than others, including more of those contents can artificially drive up the average scores. As item-level data in O-NET is not available, it is not possible to know if students perform much better in one topic compared to others. Hence, it is unclear what the effect of changes in content focus would have been.

Table 7-15 Percentage of items in each content domain in Mathematics

TIMSS's content domain	**PISA's content domain**	**PISA**		**TIMSS**		**O-NET**	
		2012	2015	2011	2015	2012	2015
Numbers	Quantity	25%	26%	29%	31%	16%	16%
Algebra	Change and relationships	25%	23%	33%	28%	16%	19%
Geometry	Shape and space	25%	25%	19%	21%	29%	29%
Data and chance	Uncertainty and data	25%	26%	19%	21%	19%	16%[150]

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b), TIMSS (TIMSS & PIRLS International Study Center, 2017), and O-NET (NIETS, 2021)

[150] The number does not add up to 100% here because the assessment also includes 20% of points allocated to "Mathematical Skills and Processes". See full breakdown of exams by learning units in Appendix E1.

Table 7-16 Percentage of items in each content domain in Science

TIMSS's content domain	PISA's content domain	PISA		TIMSS		O-NET	
		2012	2015	2011	2015	2012	2015
Biology	Living systems	45%	40%	37%	36%	20%	30%
Chemistry	Physical systems	32%	33%	20%	19%	18%	12%
Physics				25%	24%	36%	44%
Earth Science	Earth and space systems	23%	27%	18%	21%	22%[151]	14%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b), TIMSS (TIMSS & PIRLS International Study Center, 2017), and O-NET (NIETS, 2021)

7.2.5.1 Further investigations in item-level PISA Science

Changes in item characteristics can influence scores, especially when Thai students perform worse in higher-order skills in Science (Chapter Six). In the 2015 round of PISA, 99 new items in Science were included (see Table 7-17). This accounts for more than half of items included. Hence, this section investigates whether these new items are similar to the previous items or not.

[151] Similar to Mathematics, the number does not add up to 100% as 6% of the contents are on "Nature of Science and Technology".

Table 7-17 Number of items per categories

Item type	2012	2015
Trend 2012/2015	53 (100%)	53 (29%)
Trend 2006/2015		32 (17%)
New items		99 (54%)
Total	53	184

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b)

In terms of proportion of items, it can be concluded that the new items of PISA 2015 may be more cognitively demanding and difficult than items in 2012. Proportion of item types and cognitive demand support this. There are less simple multiple-choice questions in the new items compared to the trend items as shown in Table 7-18. Constructed response questions are slightly less as well, but the difference may well be negligible. The majority of the new items are complex multiple-choice. Similarly, in terms of cognitive demand (Table 7-19), there is an increase in medium level of cognitive demand in 2015 new items with a slight reduction in high cognitive demand items. This makes most items fall in the medium level, with greater reduction in low skill items rather than the high skill ones. Overall, this may imply that the 2015 new items may be slightly more difficult than 2012. For proportion of items by cognitive domains, however, there is significantly larger proportion of domains requiring lower skills such as ‘Explaining phenomena scientifically’ as shown in Table 7-20. This might counterbalance the changes in item types and cognitive demand discussed above.

Table 7-18 Proportion of each item types

Item type	Trend 2012/2015	New items 2015	Difference
Simple multiple-choice	34%	25%	-9%
Complex multiple-choice	29%	41%	**+12%**
Constructed response	36%	33%	-3%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b)

Table 7-19 Proportion of items by cognitive demand

Cognitive demand	Trend 2012/2015	New items 2015	Difference
Low	36%	30%	-6%
Medium	53%	63%	**+10%**
High	11%	7%	-4%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b)

Table 7-20 Proportion of items by cognitive domains

Cognitive domain	Trend 2012/2015	New items 2015	Difference
Explain phenomena scientifically	48%	48%	-
Evaluate and design scientific enquiry	19%	23%	+4%
Interpret data and evidence scientifically	33%	28%	-5%

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b)

As for percentage correct, results show that there exists a significant difference between the old and new items (Table 7-21, Table 7-22, and Table 7-23.). For item types, the students perform less well in the constructed response questions, with only 17% of students answering the questions correctly on average. The students perform similarly between the new and trend items for the other two item types. In cognitive domains, the results are different. Thai students scored better in the interpret domain of the new items while scoring worse for the other two lower domains. In terms of cognitive demand, it shows that students perform significantly worse in the new high demand items, with only half of the students getting the answers correctly compared to the trend item.

Table 7-21 Percent correct by item types

Item type	Trend 2012/2015	New items 2015	Difference
Simple multiple-choice	47%	47%	-
Complex multiple-choice	39%	36%	-3%
Constructed response	24%	17%	**-7%**

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b)

Table 7-22 Percent correct by cognitive domains

Cognitive domain	Trend 2012/2015	New items 2015	Difference
Explain phenomena scientifically	40%	31%	**-9%**
Evaluate and design scientific enquiry	37%	26%	-**11%**
Interpret data and evidence scientifically	32%	40%	**+8%**

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b)

Table 7-23 Percent correct by cognitive demand

Cognitive demand	Trend 2012/2015	New items 2015	Difference
Low	43%	46%	+3%
Medium	33%	28%	-5%
High	33%	16%	**-17%**

Sources: Own calculation based on released item characteristics of PISA (OECD, 2017b)

The results are inconclusive, but it is suggestive that the new items added are of slightly higher level than the previous trend items. Additionally, Thai students are performing less well on them, especially in the higher-level items. This may be the reason why the score dropped significantly during 2012 and 2015. If this is the case, Thailand’s focus needs to shift from improving basic skills to higher-order skills to be able to perform well in PISA. It is no longer enough to teach students to memorise facts and do simple applications.

7.3 Quantitative results: Modelling results

This section presents additional modelling results from the microdata of the three assessments. Data of two years were combined in the regression and year variable was used as a main predictor. In model (1), the coefficient shows the effect of the year dummy on learning outcomes. Model (2) adds school-level predictors and model (3) adds student-level predictors (the variables mirror those used in the models presented in Chapter Five). Partial results from PISA, TIMSS, and O-NET are shown in Tables 7-24 to 7-26. The tables show only the coefficient of the year dummy as it is the main focus on the analysis. To see if change over time is influenced by change in sample composition of school or student characteristics or not, the effect of year can be observed as school and student determinants are added to the model.

Table 7-24 The effect of year on learning outcomes in PISA

	(1)	(2)	(3)
Mathematics			
Year	-13.53***	-15.69***	-14.76***
	(4.80)	(4.28)	(4.02)
Science			
Year	**-23.93*****	**-25.87*****	**-24.55*****
	(4.23)	(3.74)	(3.37)
Observations	12,165	12,165	12,165

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using PISA microdata

Table 7-25 The effect of year on learning outcomes in TIMSS

	(1)	(2)	(3)
Mathematics			
Year	5.12	4.65	4.21
	(4.21)	(3.57)	(3.55)
Science			
Year	4.53	3.97	3.48
	(4.66)	(3.92)	(3.87)
Observations	12,432	12,432	12,432

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using TIMSS microdata

Table 7-26 The effect of year on learning outcomes in O-NET

	(1)	(2)	(3)
Mathematics			
Year	**0.39*** **	**0.33*** **	**0.33*** **
	(0.00)	(0.00)	(0.00)
Science			
Year	0.17***	0.10***	0.10***
	(0.00)	(0.00)	(0.00)
Observations	1,320,808	1,320,808	1,320,808

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using O-NET microdata

Model (1) shows the effect of year on Mathematics and Science scores. The results mirror the trends shown previously in Chapter Six. TIMSS shows insignificant effect of year on achievement as change over time in TIMSS scores are not meaningful nor statistically significant. Similarly, PISA Mathematics and O-NET Science subjects show statistically significant year effect, but the magnitude is lessor than 0.2 standard deviation. By contrast, in the subjects that show significant and meaningful changes, PISA Science shows a negative coefficient, indicating that the average score declined in 2015 compared to 2012 while O-NET Mathematics score shows an increase.

In all assessments, results from models (2) and (3) show little change in year coefficient after student and school characteristics have been controlled for. This implies that change in sample composition in key student and school characteristics cannot explain the score change over time. This confirms the findings from the descriptive analysis, which shows the characteristics (such as composition of school type) do not change significantly over time. The findings strengthened the claim that the discrepancies are not caused by sampling bias.

7.4 Discussion

Quantitative results investigated many factors with potential to affect change over time, from changes in sample characteristics to test item characteristics, and modelling results. Sources of discrepancies are summarised in Table 7-27. The table shows the hypothesised sources of trends discrepancies, the evidence the claims are based on, and a judgement of the strength of the evidence (whether the claim is more speculative or well-supported by evidence).

Table 7-27 Summary of evidence related to sources of trends discrepancies

Sources of discrepancies	**Evidence**	**Strength of evidence**
Changes in sampling composition and sample characteristics do not significantly affect change in scores	Quantitative – Analysis of sampling frame and coverage rates (Chapter Six) and change in sample characteristics (Chapter Seven)	High
O-NET is an unreliable test	Both quantitative and qualitative data supports this claim	High
Thai students perform worse in high-skilled items and better in low-skilled items	Both quantitative and qualitative data supports this claim	High
PISA test in 2015 includes more high-skilled items	Quantitative – Descriptive analysis of item-level data	Medium
OBEC1 schools perform worse in PISA 2015 as they are less familiar with computers	Both quantitative and qualitative data supports this claim	Medium
Greater focus on O-NET increased O-NET scores	Qualitative – Teachers' accounts of change over time	Low
Science schools increased performance because of policy change	Qualitative and quantitative	Low

Results show many of the aspects that do not change significantly over time, such as proportion of students from different school types, grade levels, and transition rates. Many of the school type characteristics also do not change significantly that it can influence scores. Regression analyses help strengthened the claim that sample composition does not change significantly over time. Adding student and school determinants of learning outcomes do not change the effect of time on the learning outcomes. This analysis helps rule out bias in sampling and change in sample composition as a potential source of discrepancies. As this claim is based on analysing nationally representative data using both descriptive and regression analysis, the evidence is of high credibility. Additionally, this helps increase confidence in analysing assessment results over time, particularly of PISA. It suggests that PISA sample in Thailand is comparable over time and is representative of the population. This contributes to literature investigating reliability of international assessments over time such as Jerrim (2013). It shows that, for Thailand, PISA results are reliable.

While sampling is unlikely to affect change over time in scores, change in item characteristics can influence the scores. The quantitative study investigated item characteristics in 2012 and 2015. Results show that in both subjects of O-NET, there are changes in exam format and proportion of items requiring different cognitive demands. In Mathematics, complex multiple choice was switched to short answer or fill in the number questions, and there are more items requiring less skills (more focusing on the Knowing domain). In Science, there seems to be more items requiring higher level skills, and the emphasis on different content area changed. This suggests that O-NET's exam difficulty may have changed over time. In the case of time period between 2012 and 2015, Science exams may have become more difficult while Mathematics exam easier.

This aligns with the qualitative part of the research in Chapter Five, which shows teachers perceiving that O-NET exam difficulty changed over time. Teachers speculated that O-NET exam difficulty is different each year. Some years, the exams were easier, and students performed well. In other years, the questions were more difficult. Changes in difficulty happen every year, but most prominently after 2015, when the examinations have been increasingly difficult owing to IPST's involvement in designing the exams (see 0). This suggests that changes in O-NET's results may not be reliable over time. The change in scores can be from change in exam difficulty as much as improvement in actual learning outcomes. This is similar to what the report by OECD and UNESCO (2016) concludes, as NIETS fail to provide explanations on measures taken to ensure trends are reliable. When the change is not reliable, making comparison in scores over time become difficult. This may suggest that O-NET trends may not reflect true change in learning outcomes. Levels of achievement over time are likely to be stagnated, or show a slight decline, as PISA and TIMSS suggested, rather than increasing like suggested by O-NET.

Another implication of O-NET's trends being unreliable is that the data is less useful for the Ministry and the schools. In Chapter Five, practitioners claim that O-NET is used extensively for school accountability. When a flawed test is placed such importance, the system may not be rewarding the right schools. It also makes it more difficult for practitioners to prepare for the assessment when the focus can change from year to year. For O-NET to be more useful, some changes needed to be made in the exam design. For example, O-NET can include some ungraded anchor items, or ensure that each year, roughly the same proportion of items are present.

Another solid explanation for the discrepancies is that Thai students perform better in lower-skilled items (such as memorisation), and worse in higher-skilled items (such as problem-solving). Item-level data shown in this chapter and Chapter Six imply that over time, the students perform less well in higher cognitive domains tested in PISA, and better in memorisation questions tested in TIMSS. Descriptive analyses in Chapter Six show that PISA tests higher skills than TIMSS and O-NET. Additionally, this chapter shows suggestive evidence that PISA Science may have become more difficult over time by adding more higher-skilled items[152]. Thai students are also performing less well on these new items. Consequently, this provides an explanation why PISA results show a decline in Science, but not in TIMSS or O-NET, which test lower-skilled items. Further, qualitative results support that teachers focus less on higher-skilled items over time, as they emphasised more of O-NET, which focuses heavily on memorisation. This shift may contribute to students performing less well in higher-skilled items.

The next claim that partially explains the discrepancies lies with the performance of OBEC1 schools. Chapter Six shows that the school group shows the greatest decline in PISA scores in 2015 compared to 2012. It is possible that the decline may be from the schools' unfamiliarity in computers (as seen in Chapter Five showing the schools' lack of IT resources). As PISA switched from paper-based to computer-based test in 2015, OBEC1 schools may perform less well simply because the students are not proficient in computer skills. Additionally, one of the interviewees believe this to be the case but noted that this is purely his own assertion. The evidence is rated as medium as this is more speculative. As discussed in Chapter Three, it is not possible to test the effect of switching to computer-based test as done in a research by Jerrim (2016).

Two explanations are rated as being more speculative as they are based on qualitative research and not supported by quantitative data. Chapter Five provide some explanations for change over time in learning outcomes based on the interviews. This includes a claim that greater focus of O-NET tests led to improvement in O-NET scores. The increase in focus may have led to improvement in scores. However, there is solid evidence that O-NET is not a reliable test. As a result, it cannot be ruled out that the increase may be from change in test difficulty rather than practice. Another explanation for the change is the improvement in Science school performance. However, this change only

[152] The evidence is suggestive, but less credible than the claim of Thai students' performance as some characteristics of the new Science items show increasing difficulty (such as increasing proportion of complex multiple choice) while some aspects show decrease in difficulty (lesser items with high cognitive demand – see section 7.2.5.1.).

affects O-NET. Further, Chapter Six shows that enrolment for Science schools are very low relative to OBEC schools. Hence, their score increase is unlikely to affect average score significantly.

Chapter Eight: Conclusion

This thesis aimed to answer the main research question of “What could explain the discrepancies in trends during 2011/2012-2015 between the three different assessments of O-NET, PISA, and TIMSS, and what implications do they have on education policy in Thailand?”. The study is multidisciplinary, drawing from education, social science, economics, and statistics literature. The thesis’s research question fits into answering a research problem of why learning outcomes are low in Thailand. Explaining why the trends differ helped us understand the nature of low learning outcomes and offered explanations why they remain low. The thesis also analysed its implications on Thai education policy to help policymakers make sense of the data.

Through mixed methods research, the thesis offers insights from both big-picture quantitative analysis and in-depth qualitative analysis. Quantitative analyses, utilising descriptive statistics and regression models, describe the nature of learning levels in Thailand, as well as identifying factors affecting variations in scores and how that changed over time. Qualitative analyses supplemented the quantitative findings by offering insights from practitioners (teachers and vice principals). Through semi-structured interviews, participants discuss factors affecting learning outcomes, their teaching practices, and what changed over time.

The research contributed to the literature explaining why learning outcomes are low in Thailand and how that can be improved. It is the first to employ mixed methods to address this research problem. While much research on learning levels in Thailand focus on cross-sectional results of one assessment, this research is the first to draw evidence from assessment results over time of PISA, TIMSS, and O-NET. This enriched our understanding of student performance, as well as in which ways the assessments are similar and different. It contributed to research critiquing reliability of assessment results as well as education research in understanding low learning levels in developing countries.

The summary of the findings can be found below.

8.1. Overall learning level is low and unequal

Throughout the thesis, results show that average learning levels in Thailand over the period studied (2011-2015) has been low and unsatisfactory. This confirms and adds to the results of many studies on Thailand (such as OECD and UNESCO (2016)) and Thai media that found learning to be low. Average scores in PISA and TIMSS have been at the lowest level of proficiency (Level 1 in PISA and Low in TIMSS). In O-NET, average scores have been consistently below 50% correct in core subjects of Mathematics and Science. This means that the policymakers' concerns are therefore justified. Thailand indeed has low learning levels and improvements can be made.

True learning levels seem to have not improved over time. Chapter Seven shows that O-NET trends can be influenced by many factors including changing exam difficulty. Hence, the positive changes found may not reflect improved levels of learning. PISA and TIMSS, by contrast, shows robust sampling, at least in the case of Thailand. PISA Science scores show decline while PISA Mathematics and TIMSS shows no change in learning. We can deduce from this that the learning levels in Thailand has remained roughly the same during 2011 – 2015. One explanation why PISA Science shows a decline is that Thai students may have become worse on higher level skills, while better at memorising facts. Hence, in a test that focuses more on higher-level application of knowledge such as PISA, results show a decline. This finding adds to research explaining different trends over time. For Thailand, change in scores is influenced by performance by skill level, instead of issues identified in literature such as sampling (Spaull, 2017), social class composition (Carnoy & Rothstein, 2013), and procedure change (Jerrim, 2013). Future research can investigate this aspect as a potential source of score discrepancies.

It is concerning that Thai students perform worse on skills such as interpreting and analysing information. This should not come as a surprise as both the national curriculum and assessment focuses more on memorising rather than application of knowledge. Nonetheless, the results prompt policymakers to rethink what skills are necessary and what students should learn. While memorisation skills are arguably foundational in learning other skills, it is no longer enough to have only such basic skills. To thrive in today's world, students should be able to apply knowledge and engage in critical thinking and problem-solving. They are skills reported by employers as essential, but also that Thai workers are lacking (ILO, 2019; World Bank, 2008). Thailand has been making attempts to move away from memorisation in recent years, with changes made to the curriculum and to O-NET. This change is in the right direction if Thailand wants to perform well in PISA. However, there remains improvements to be made as the new curriculum only focuses on basic application of knowledge.

While learning has been low, it is not uniformly low. OBEC2 schools perform around average. Schools such as Satit and Science schools perform at least one standard deviation above the mean. Their scores are as high as the top-performers in the OECD countries. At the same time, OBEC1, Local, and BMA schools perform more than 0.2 standard deviation below the mean. Similar to Phuaphansawat (2021), this shows that school effects are strong in Thailand and being in certain school types affect learning outcomes significantly. It seems that the gaps among school type also widened over time, with the top-performers showing more improvement than the bottom-performers. Reducing this gap can help bring up the average learning and reduce share of students performing below minimum standards. To address the gap, the thesis dived into the causes of inequality.

Both quantitative and qualitative data shows that much of the learning gap can be accounted for by student background, which is one of the important predictors of learning outcomes identified in the Education Production Function and Educational Effectiveness Research frameworks. In this sense, the findings reaffirm the relevance of these frameworks in explaining variations in learning outcomes. Additionally, the findings are similar to Phuaphansawat's (2021) study showing that student socioeconomic background is the main driver of learning outcomes. Satit and Science schools have selective admission policies, and as a result enrol students of much higher socioeconomic status and prior achievement. Students from other school types with high academic achievement seem to transition into Satit and Science schools. This situation is not unique to Thailand. In the UK for example, there is evidence of students of higher socioeconomic status aggregated in grammar schools[153] (Jerrim & Sims, 2019). OBEC1 and BMA schools, by contrast, enrol students from much lower socioeconomic status and those who did not perform well in O-NET at grade 6. There are also problems with foundational knowledge and motivation that stems from low socioeconomic status and prior achievement.

With self-selection, Science and Satit schools benefit from the aggregation of students with better prior scores and socioeconomic background. By contrast, poorer students enrol in low-performing schools. With this, simple policies such as increasing resources to low-performing schools alone or making every school adopt Science school curriculum cannot improve learning outcomes. The government can instead focus on addressing factors that make the disadvantaged schools not performing up to the standards. Qualitative results show major problems to be weak academic foundation, low motivation, and lack of resources. All these problems stem from systemic roots. While tutorials can help bring up students to standards, it is not a long-term solution. Weak foundational knowledge of students is a problem that has been accumulated from primary level, with students not learning the required content. Unpacking this, it is because the curriculum requires many contents to be taught and schools not having enough time to teach all of them. This, combined with low motivation to study and perform well on assessments, has resulted in poor learning outcomes. Lack of resources is another problem borne out of how funding is set up in schools. Small schools

[153] Students from lower income background who scored well were also admitted, but in lessor proportion than those with high income background.

get allocated less teachers as the Ministry controls the number of teachers per students. Additionally, teachers prefer to be allocated to schools in urban areas. With this, small schools such as OBEC1 in rural areas lack subject specialised teachers. These issues should be addressed to bring the performance level up.

In addition to knowing the problems of the low-performers, it is equally important to understand what sets the high-performers apart. It seems that advantage in socioeconomic background and student prior achievement only explain part of the advantage. This finding mirrors research that found achievement gap in Thailand to be largely unexplained by observable characteristics (Lathapipat & Sondergaard, 2015; Lounkaew, 2013; World Bank, 2012a). Qualitative results offer additional explanations that can explain the gap. Findings show that Satit and Science schools were able to teach deeper than the curriculum requires, making the students good in higher-level skills tested in PISA. They also engage in teaching practices that are more conducive to learning higher level skills, as opposed to giving lectures. As these schools enrol students with better academic background, it is easier for them to teach deeper and teach high level skills. Low-performing schools might need to prioritise foundational knowledge first before they can focus on teaching higher level skills.

8.2. Curriculum and national assessment dictate what skills are learnt

Participants mentioned the importance of following the national curriculum and scoring well on O-NET. In the current accountability system, schools are evaluated based on how well they perform on O-NET. This made them engage in intensive tutoring and teaching to the test to ensure that students perform well on O-NET. During 2012 – 2015, these practices led to schools teaching students mainly about memorising contents, as O-NET was heavily focused on testing memorisation. Though causal effects cannot be estimated, this probably plays a part in Thai students performing worse in PISA, which focuses less on memorising facts and more on critical thinking and problem-solving. Quantitative data also shows that Thai students perform worse in higher order skills over time in PISA, which can be because less time is spent on higher order skills as schools focus more on memorisation. In TIMSS, students improved in memorisation questions.

This shows the importance curriculum and assessment in affecting what is taught in classrooms and what skills are learnt. While the effect of schools and student characteristics can explain partially the variations in learning outcomes, the system-level factors affect significantly what are taught in classrooms, as stressed by studies like Mounier and Tangchuang (2010). When the curriculum and assessment focus heavily on memorisation of contents, these ended up being what students learned. At the same time, as teachers teach according to the curriculum, it can be a powerful tool for the government to push for change, as demonstrated in reforms in many countries (see the LEGO Foundation (2022)). After 2017, when the curriculum changed (and assessment subsequently), teachers adapted their teaching to align with the new standards. This poses significant decision for the Ministry on what kind of skills the students at the end of compulsory education should have. There is no right or wrong answer regarding which way Thailand should go. Nonetheless, if the goal is to perform well in international assessments such as PISA, a move towards knowledge application is a good start. Thailand can go further by integrating higher-level skills such as more complex problem-solving and critical thinking.

Another lesson that came out from the change in curriculum and assessment (post-2015) is regarding stakeholder buy-in. Participants implemented the change, but they also mentioned the challenges in adjusting to the new curriculum, as well as the lack of support that comes with it. This resulted in some low-performing schools not being able to implement the curriculum successfully. Hence, change needs to be carefully designed, lobbied, and carried out. This resonates with many of Hallinger’s research on school leadership support needed to implement change.

Part of why teachers teach according to the national curriculum and national assessment is because of the accountability system. In Thailand, schools are ranked against others based on O-NET performance. Good schools are those that perform better than the national average and show constant improvement. Chapter Seven shows many flaws related to these criteria. By definition, it is not possible for all schools to perform above the national average. Improvement every year is also unlikely. Arguably, this is not an effective nor equitable accountability system. This punishes schools with intakes of less advantaged students and leads to the schools engaging in intensive tutoring to perform well. The current situation is what Pritchett (2015) would describe as system incoherence. The skills beyond memorisation are not measured very well in O-NET, and even though the incentive system was set up to include learning improvement, it has not been done in such a way that incentivise equitable improvement.

8.3. Assessments are only useful when they are reliable

The thesis has used data from PISA, TIMSS, and O-NET to analyse learning levels both at one point in time and over time. For snapshot analyses, Chapter Six results show large degree of similarities among the assessments. When results are compared at one point in time, relative ranking of each subgroup is similar, with similar magnitude in gaps. This shows that all assessments are useful in providing information about gaps in learning outcomes among students in Thailand. They complement each other in explaining the nature of learning outcomes, as the tests differ in terms of actual contents tested as well as information available. PISA can offer insights on higher-level skills while O-NET on memorisation and simple applications of knowledge. Policymakers should draw insights from all the different assessments to understand more about learning outcomes.

When the results are compared over time, however, the trends differ. Chapter Seven shows there are changes over time in the tests themselves. One concerning finding is that contents tested on O-NET exams change each year. O-NET did not attempt to equate tests over time (OECD & UNESCO, 2016). This has serious implications on comparison of O-NET results over time. If test difficulty changed over time, the trends may not reflect true changes in learning levels. Consequently, using O-NET to infer changing level of learning outcomes is problematic. Claims such as the ones made by the OBEC secretary that O-NET's increase in scores are from certain successful policies (Manager Online, 2019) are not valid. Not only that the change in score cannot be causally attributed to said policies, but the increase in scores may not even reflect improvement in learning levels.

Additionally, as seen from the previous sub-section, O-NET results are used to evaluate schools and students. When O-NET scores are not reliable, it is unfair for schools and students. Students can be getting the same scores, but they can mean different things over time. Hence, the current system may not be rewarding the right schools. To incentivise schools to improve learning outcomes, O-NET results need to be reliable.

8.4. Policy recommendations

With the findings above, the thesis suggests policy recommendations below.

To help support small schools:

- Increase number of subject specialised teachers;
- Support best practice sharing among schools; and
- Adapt curriculum to be more relevant

To improve accountability system:

- Integrate school context

To improve data:

- Consider test equating measures in O-NET

Firstly, increasing number of subject specialised teachers can help small schools significantly. Chapter Five shows OBEC1 schools' lack of specialised teachers affecting learning. There have been attempts to relocate more teachers to rural areas (through policies such as Kru Kuen Tin[154], etc.). Nonetheless, many schools still lack subject specialised teachers. One of the solutions the Ministry can try is to allow and facilitate schools to share specialised teachers. One specialised teacher can work in multiple schools or act as a mentor teacher for non-specialised teachers. In the long run, arguably, the World Bank's arguments[155] that small schools are too economically costly to maintain makes sense, but the change should be done gradually with more support. For instance, students who live in remote areas with no alternative schools should be provided subsidies for transport or supporting schools to offer bussing and/or boarding. Schools with very small enrolment should not be allowed to accept more students, and gradually close as the last cohorts of students leave. This requires cooperation between parents, schools, and local authorities to implement. Increasing resources alone cannot improve learning outcomes, as seen from research investigating relationship between resources and learning outcomes such as Hanushek and Woessmann

[154] Policy to increase number of teachers in rural area. See section 6.5.
[155] Lathapipat and Sondergaard (2015)

(2007). Hence, this has to be done with improvement in other areas such as teaching quality.

Secondly, knowledge sharing among schools can help bring low-performing schools up to standards. Currently, schools have informal learning communities within themselves. Satit and Science schools mentioned that they informally help nearby schools in teaching and O-NET preparation. This can be expanded in scope by including collaboration from networks of schools. Schools that perform well can share their best practices to other schools. Perhaps within the same school type, schools that did exceptionally well can share what they did. We see from the transition of Science schools that having MWIT[156] mentoring them helped tremendously. A similar model can be piloted with other types of schools. In British Columbia, for example, a collaboration is facilitated by the local education authorities (S. Brown et al., 2017). In Thailand, the local authorities should be in the best position to do this as well, as they already work with many schools in the area.

Thirdly, to solve the problem at the root cause, the curriculum needs to become more relevant to students. While global trends are moving away from curriculum with heavy contents (Joynes et al., 2019), Thai curriculum still has many contents included. Many educators including Nakornthap (2018) had advocated for a more relevant curriculum. With more relevant curriculum, students might also be more motivated to learn. Lesson from curriculum change shows that stakeholders[157] should be supported as well as consulted during the change. In countries with successful curriculum change, stakeholder consultation and easing practitioners to the change are the common themes (LEGO Foundation, 2022). A support system needs to be built to aid teachers in the form of continual professional development for existing teachers and teacher training programmes for new teachers.

---

[156] Another science-focused school, see section 7.4.3.
[157] Teachers, principals, local authorities, and others involved in implementation.

Fourthly, to improve the accountability system, I argue that the system should take into account school context. In the UK, for example, school evaluation takes into account both absolute levels of academic performance (Attainment 8) and also performance accounting for pupil's prior scores (Progress 8) (DfE, 2020). By adding prior scores, the schools are compared with other schools with similar prior scores levels and are evaluated based on the progress made during the school year. The policy is not without its criticisms. Leckie and Goldstein (2019), for instance, argued for inclusion of both pupil background and prior scores in school evaluation. Nonetheless, this system compares more similar schools together, and is arguably fairer to the schools.

For Thailand, this system can be adapted. As grade 6 O-NET scores are readily available, they can be incorporated as part of school evaluation alongside looking at snapshots of learning outcomes. Additionally, instead of setting the same criteria for every school, the schools and local authorities should be able to set local level goals that are suited to their school context. An overarching goal can be set nationally (such as a minimum performance that corresponds to the minimum desired skill). Other than this goal, local authorities should be able to decide what constitutes success. What one school considers as significant improvement may not be the same for other schools. This change would allow schools to be more flexible and strive towards goals that are more relevant to them.

Fifthly, O-NET should introduce measures to keep test difficulty constant over time. PISA and TIMSS use anchor items or the same items in different rounds (OECD, 2017b). With this, scores can be compared over time and changes in terms of exam difficulty and student characteristics are controlled for. It might be possible for O-NET to attempt a similar technique. For example, non-equivalent groups anchor test (NEAT) design (Kolen & Brennan, 2014) can be implemented. This design can equate tests when there are two groups of test-takers who take two different exams, but both groups complete a common sets of anchor items. In this case, students of different years are considered different groups as they may differ in their characteristics. O-NET can introduce sets of common items that remain the same over time into the exam. These can be ungraded and unpublished items, so that they can be used to equate test results. With this, trends over time of O-NET will be more meaningful.

For the recommendations to be useful and practical, it is important to recognise the political constraints that affect policymaking in Thailand. These can be barriers to changes. There has been a constant political struggle between conservatives and those who want change (Mounier & Tangchuang, 2010). The conservatives value having a hierarchical society where knowledge is simply being reproduced through rote learning. Mounier and Tangchuang (2010) argue that historically, the conservatives prevail in deciding how education should be like. This resulted in many years of education that are focused on memorisation of knowledge and policies that are top-down in nature. Now that the military junta sits as the government, the country's power is in the hand of the conservatives.

This suggests that some of the changes may be difficult to implement. Historically, changes have happened when there is political will and the conservatives were not in power (Sangnapaboworn, 2018). Some changes that directly challenge the military government are likely to be deprioritised. The new curriculum and assessment clearly show attempts to move away from rote learning and towards knowledge application and critical thinking. However, there is tension as the military government is less likely to want the population to be more critical as they might be more critical towards the government's actions. Recently, a move towards a competency-based curriculum has been halted, with no concrete justifications (Prachachat, 2022). Hence, this poses challenges to change even though many educationists advocate for them.

Another related constraint is with the nature of the Thai culture itself. Jackson (2004) asserted that the culture is characterised by the concern with public image and respecting social order. This had led to many problems not being discussed in public. At a school level, Chapter Five shows that schools falsify documents during quality assurance and avoid discussing problems in the school. At a national level, this led to officials blaming the tests having flaws instead of acknowledging that students are not performing well. As seen in the news, the Minister at the time blamed PISA for having poor translation, resulting in low scores of Thai students (Mala, 2018). Going forward, it is important to acknowledge there is a systemic problem of low learning that needs to be addressed. Some schools perform much worse than others, and it is crucial that the problems are discussed frankly.

8.5. Limitations of the study and future research

There are some data restrictions I encountered in the quantitative research. In some cases, required data are simply not available. The research has worked mostly with student-level data. In PISA and TIMSS, simple item analysis was conducted such as looking at the percent correct of specific items. In future research, detailed item-level data analysis can be conducted, for example to look at the effect of item-response model choice on achievement. For this research, unfortunately, the permission to use multiple-years item-level data was not granted. However, NIETS may be able to provide a subset of item-level data for secondary analysis. This could be explored in future research.

For PISA and TIMSS, the biggest limitation is the lack of school type identifier in the microdata. TIMSS have always not included the school type, while PISA includes them up until 2015. IPST claimed that this exclusion is to keep the data anonymised (OECD & IPST, 2018). This choice to exclude school type hinders many of the possible secondary data analyses. For example, regression analysis cannot include school type as a predictor of scores beyond 2012. The anonymity concern is justified, as seen in PISA's sampling frame in Appendix C2. The frame shows the number of schools in each category and the number of schools actually sampled. Some categories are few in number and a school could be identified by including school type. For example, there is only one Satit school that has only lower secondary level. Nonetheless, I argue that making school type identifier available will be very beneficial with researchers doing secondary data analysis. School types can be grouped in some ways that make them not identifiable. For instance, Satit and Science schools can be grouped together as special-purpose schools to maintain anonymity.

The qualitative part of the study involved participants who are practitioners in the schools: teachers and management team. Focusing on this had offered many insights based on their perspectives, and how the schools view different assessments. The participants mentioned other stakeholders who also greatly influence assessment results and policies that influence school practice, including the local authorities, IPST, and the Ministry itself. One limitation of this study is that these stakeholders were not included in the research. Additionally, students are also important stakeholders. Interviewees have discussed about how they think students perceive assessments. While this offered explanations on the scores, interviewing students could yield a richer account of how assessments have affected them. Therefore, future studies could consider including these stakeholders as participants.

Another limitation of the study is that the time frame (2012 – 2015) could be perceived as having happened a long time ago by the participants. As of when the fieldwork was conducted, 2012 means 8 to 9 years ago for the participants. Some remarked that they may have not remembered that well what happened at that time. Changes over time that the participants perceived as important were changes that happened more recently, as opposed to those that happened 8 – 9 years ago. In future research, focusing on a more recent time frame may help participants remember more, or conducting a longitudinal study and to so being able to talk to the participants as the change unfolds. Additionally, by requiring that participants had been at the school from 2012 implied that the participants were practitioners who had worked at the school for relatively long time, as opposed to those who stayed only for a couple of years then moved to a different school. Due to the purpose of comparing practices over time, this requirement was correct for this study. Nonetheless, it is possible that those staying in one school for a long time may have different mindset or perspective than those who stayed for shorter amount of time.

Future studies can build upon the findings by utilising other data sources, including item-level data from O-NET. This can help pinpoint types of items that students excel at or need improvement. The research problem of low learning can be viewed from the perspectives of other stakeholders or using alternative methods of data collection. Additionally, a longitudinal study may be considered. Instead of asking participants to recall past practices, future study can have multiple data collection at different points in time.

References

ADB (Ed.). (2015). *Thailand, industrialization and economic catch-up: Country diagnostic study*. Asian Development Bank.

Aloisi, C., & Tymms, P. (2017). Pisa Trends, Social Changes, and Education Reforms. *Educational Research and Evaluation*, *23*(5–6), 180–220. https://doi.org/10.1080/13803611.2017.1455290

Anagnostopoulos, D. (2006). “Real Students” and “True Demotes”: Ending Social Promotion and the Moral Ordering of Urban High Schools: *American Educational Research Journal*, *43*(1), 5–42. https://doi.org/10.3102/00028312043001005

Anders, J., Jerrim, J., & McCulloch, A. (2016). How Much Progress Do Children in Shanghai Make Over One Academic Year? Evidence From PISA. *AERA Open*, *2*(4), 1–13. https://doi.org/10.1177/2332858416678841

Armor, D. J., Marks, G. N., & Malatinszky, A. (2018). The Impact of School SES on Student Achievement: Evidence From U.S. Statewide Achievement Data. *Educational Evaluation and Policy Analysis*, *40*(4), 613–630. https://doi.org/10.3102/0162373718787917

Atagi, R. (2002). *The Thailand Educational Reform Project: School Reform Policy*. Asian Development Bank. http://backoffice.onec.go.th/uploaded/Category/EngBook/SchoolReformPolicy-04-03-2011.pdf

Avvisati, F., & Keslair, F. (2020). Repest: Stata Module to Run Estimations with Weighted Replicate Samples and Plausible Values. In *Statistical Software Components* (Version S457918) [Computer software]. Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s457918.html

Baars, S., Shaw, B., Mulcahy, E., & Menzies, L. (2018). *School Cultures and Practices: Supporting the Attainment of Disadvantaged Pupils*. Department for Education.

Barber, B. (1989). Thailand Looking to Become Asia’s 5th Tiger. *The Globe and Mail; Toronto, Ont.*, B.9.

Barrera-Osorio, F., Garcia-Moreno, V., Patrinos, H. A., & Porta, E. (2011). *Using the Oaxaca-Blinder Decomposition Technique to Analyze Learning Outcomes Changes over Time: An Application to Indonesia’s Results in PISA Mathematics. Policy Research Working Paper 5584* (Policy Research Working Papers). Human Development Network Education.

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *Journal of Human Resources*, *8*(4), 436–455. https://doi.org/10.2307/144855

Bolotta, G. (2021). ‘Invisible Worldings’: Image and Reality in the Thai Seafood Industry’s Humanitarian Engagements. *European Journal of East Asian Studies*, 1–26. https://doi.org/10.1163/15700615-20211020

Booher-Jennings, J. (2005). Below the Bubble: “Educational Triage” and the Texas Accountability System. *American Educational Research Journal*, *4*(2), 231–268. https://doi.org/10.3102/00028312042002231

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology; London*, *3*(2), 77–101. http://dx.doi.org/10.1191/1478088706qp063oa

Brown, G., Micklewright, J., & Schnepf, S. V. (2007). International Surveys of Educational Achievement: How Robust Are the Findings? *Journal of the Royal Statistical Society*, *170*(3), 623–646.

Brown, G. T. L., & Harris, L. R. (2016). *Handbook of Human and Social Conditions in Assessment*. Routledge.

Brown, S., Hales, A., Kuehn, L., & Steffensen, K. (2017). *The State of Educators’ Professional Learning in British Columbia*. Learning Forward.

Bryman, A. (2012). *Social Research Methods* (4th edition). Oxford University Press.

Bualoy, P. (1998). *Operational Problems in Extended Primary Schools: A Case Study of Udon Thani School Groups*. Thammasat University.

Budget Bureau. (2019). *Thailand’s Budget in Brief Fiscal Year 2019*. Budget Bureau. http://www.bb.go.th/topic-detail.php?id=8319&mid=544&catID=1074

Carle, A. C. (2009). Fitting Multilevel Models in Complex Survey Data with Design Weights: Recommendations. *BMC Medical Research Methodology*, *9*, 49. https://doi.org/10.1186/1471-2288-9-49

Carnoy, M., Khavenson, T., & Ivanova, A. (2013). Using TIMSS and PISA Results to Inform Educational Policy: A Study of Russia and Its Neighbours. *Compare: A Journal of Comparative and International Education*, *45*(2), 248–271. https://doi.org/10.1080/03057925.2013.855002

Carnoy, M., & Rothstein, R. (2013). *What Do International Tests Really Show About U.S. Student Performance?* Economic Policy Institute.

Carr-Hill, R. (2012). Missing Millions and Measuring Progress towards the Millenium Development Goals with a focus on Central Asian States. *Central Asian Journal of Global Health*, *1*(1), Article 1. https://doi.org/10.5195/cajgh.2012.24

Chansri, S., & Kittichotipanit, N. (2014). Factors Influencing Scores in Learning Mathematics of Mathayomsuksa 6 in the Science and Mathematics Program Students under the Office of Phra Khanong Secondary Educational Area. *Journal of Science Ladkrabang*, *23*(1), 62–79.

Chapman, C., Muijs, D., Reynolds, D., Sammons, P., & Teddlie, C. (2015). *The Routledge International Handbook of Educational Effectiveness and Improvement: Research, Policy, and Practice*. Routledge. http://ebookcentral.proquest.com/lib/ucl/detail.action?docID=2166360

Chiengkul, P. (2019). Uneven Development, Inequality and Concentration of Power: A Critique of Thailand 4.0. *Third World Quarterly*, *40*(9), 1689–1707. https://doi.org/10.1080/01436597.2019.1612739

Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education* (6th edition). Routledge.

Coleman, J. S., Campbell, E., Hobson, C., McParttland, J., Mood, A., Weinfield, F., & York, R. (1966). *Equality of Educational Opportunity*. Office of Education, United States. https://eric.ed.gov/?id=ED012275

Creemers, B. P. M., & Kyriakides, L. (2008). *The Dynamics of Educational Effectiveness: A Contribution to Policy, Practice and Theory in Contemporary Schools*. Routledge.

Creswell, J. W., & Creswell, J. D. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (5th edition). SAGE.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research* (2nd edition). SAGE Publications.

Dailynews. (2017). IPST to Design Mathematics and Science O-NET Exams in 2018. *Dailynews*. https://www.dailynews.co.th/education/563459/

Day, C., Sammons, P., & Gu, Q. (2008). Combining Qualitative and Quantitative Methodologies in Research on Teachers' Lives, Work, and Effectiveness: From Integration to Synergy. *Educational Researcher*, *37*(6), 330–342. https://doi.org/10.3102/0013189X08324091

DeJaeghere, J., Duong, B.-H., & Dao, V. (2021). *Teaching Practices That Support Thinking and Promote Learning: Qualitative Evidence from High and Low Performing Classes in Vietnam* (RISE Insight Notes). RISE.

Della Porta, D., & Keating, M. (2008). How Many Approaches in the Social Sciences? An Epistemological Introduction. In M. Keating & D. Della Porta (Eds.), *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective* (pp. 19–39). Cambridge University Press. https://cadmus.eui.eu//handle/1814/10052

DfE. (2011). *Class Size and Education in England Evidence Report*. Department for Education.

DfE. (2020). *Secondary Accountability Measures: Guide for Maintained Secondary Schools, Academies and Free Schools*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/872997/Secondary_accountability_measures_guidance_February_2020_3.pdf

DFID. (2009). *Political Economy Analysis How To Note*. DFID. https://www.odi.org/sites/odi.org.uk/files/odi-assets/events-documents/3797.pdf

Doner, R. F. (2009). Puzzles of Thai Development in Comparative Perspective. In R. F. Doner (Ed.), *The Politics of Uneven Development: Thailand's Economic Growth in Comparative Perspective* (pp. 25–63). Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9780511819186.002

Duncan, K., & Sandy, J. (2013). Using the Blinder-Oaxaca Decomposition Method to Measure Racial Bias in Achievement Tests. *The Review of Black Political Economy*, *40*(2), 185–206. https://doi.org/10.1007/s12114-012-9146-2

EEF. (2021a). *Equitable Education Fund Annual Report 2021*. EEF. https://www.eef.or.th/wp-content/uploads/2022/03/EEF-AnnualReport2564.pdf

EEF. (2021b). *Teacher Information*. Information System for Equitable Education. https://isee.eef.or.th/screen/teacherdata/teacher.html

Egalite, A. J., & Kisida, B. (2016). School Size and Student Achievement: A Longitudinal Analysis. *School Effectiveness and School Improvement*, *27*(3), 406–417. https://doi.org/10.1080/09243453.2016.1190385

Eichengreen, B., Park, D., & Shin, K. (2011). When Fast Growing Economies Slow Down: International Evidence and Implications for China. *NBER Working Paper*, *16919*. https://www.nber.org/papers/w16919.pdf

El Masri, Y. H., Baird, J.-A., & Graesser, A. (2016). Language Effects in International Testing: The Case of PISA 2006 Science Items. *Assessment in Education: Principles, Policy & Practice*, *23*(4), 427–455. https://doi.org/10.1080/0969594X.2016.1218323

Ercikan, K., & Koh, K. (2005). Examining the Construct Comparability of the English and French Versions of TIMSS. *International Journal of Testing*, *5*(1), 23–35. https://doi.org/10.1207/s15327574ijt0501_3

Everson, K. C. (2017). Value-Added Modeling and Educational Accountability: Are We Answering the Real Questions? *Review of Educational Research*, *87*(1), 35–70. https://doi.org/10.3102/0034654316637199

Feskens, R., Fox, J.-P., & Zwitser, R. (2019). Differential Item Functioning in PISA Due to Mode Effects. In B. P. Veldkamp & C. Sluijter (Eds.), *Theoretical and Practical Advances in Computer-based Educational Measurement* (pp. 231–247). Springer International Publishing. https://doi.org/10.1007/978-3-030-18480-3_12

Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica*, *77*(3), 953–973.

Fry, G. W. (2018). *Education in Thailand: An Old Elephant in Search of a New Mahout*. Springer Singapore.

Fry, G. W., & Bi, H. (2013). The Evolution of Educational Reform in Thailand: The Thai Educational Paradox. *Journal of Educational Administration*, *51*(3), 290–319. https://doi.org/10.1108/09578231311311483

Fry, G. W., Bi, H., & Apahung, R. (2018). Regional Educational Disparities in Thailand. In G. W. Fry (Ed.), *Education in Thailand: An Old Elephant in Search of a New Mahout* (pp. 373–391). Springer Berlin Heidelberg.

FWAO. (2021). *Statistics of Migrant Workers with Permission to Work in Thailand as of September 2021*. Foreign Workers Administration Office. https://www.doe.go.th/prd/assets/upload/files/alien_th/e7ecaa17a00907e4eefe2a5a81462935.pdf

Gamage, D. T., & Sooksomchitra, P. (2004). Decentralisation And School-Based Management In Thailand. *International Review of Education*, *50*(3), 291–308. https://doi.org/10.1007/s11159-004-2624-4

Garza, J. E., Drysdale, L., Gurr, D., Jacobson, S., & Merchant, B. (2014). Leadership for School Success: Lessons from Effective Principals. *International Journal of Educational Management*, *28*(7), 798–811. https://doi.org/10.1108/IJEM-08-2013-0125

Gill, T., & Benton, T. (2013). *Investigating the Relationship Between Aspects of Countries' Assessment Systems and Achievement on the Programme for International Student Assessment (PISA) Tests* (Cambridge Assessment Research Report). Cambridge Assessment.

Glewwe, P., Chen, Q., & Katare, B. (2015). What Determines Learning among Kinh and Ethnic Minority Students in Vietnam? An Analysis of the Round 2 Young Lives Data. *Asia & the Pacific Policy Studies*, *2*(3), 494–516. https://doi.org/10.1002/app5.102

Glewwe, P., Hanushek, E., Humpage, S., & Ravina, R. (2011). *School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010* (No. w17554). National Bureau of Economic Research. http://www.nber.org/papers/w17554.pdf

Glewwe, P., Lee, J., & Vu, K. D. (2020). *What Explains Vietnam's Exceptional Performance in Education Relative to Other Countries? Analysis of the 2012 and 2015 PISA Data* (No. 20/036; RISE Working Paper Series).

Goldstein, H. (2017). Measurement and Evaluation Issues with PISA. In L. Volante (Ed.), *The PISA Effect on Global Educational Governance* (1st ed., pp. 49–58). Routledge. https://doi.org/10.4324/9781315440521

Goldstein, H., & Wood, R. (1989). Five Decades of Item Response Modelling. *British Journal of Mathematical and Statistical Psychology*, *42*(2), 139–167. https://doi.org/10.1111/j.2044-8317.1989.tb00905.x

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis*, *11*(3), 255–274. https://doi.org/10.3102/01623737011003255

Hadfield, M., & Chapman, C. (2015). Qualitative Methods in Educational Effectiveness and Improvement Research. In C. Chapman (Ed.), *The Routledge International Handbook of Educational Effectiveness and Improvement: Research, Policy, and Practice* (pp. 202–219). Routledge.

Hahs-Vaughn, D. L., McWayne, C. M., Bulotsky-Shearer, R. J., Wen, X., & Faria, A.-M. (2011). Methodological Considerations in Using Complex Survey Data: An Applied Example With the Head Start Family and Child Experiences Survey. *Evaluation Review*, *35*(3), 269–303. https://doi.org/10.1177/0193841X11412071

Halinen, I. (2018). The New Educational Curriculum in Finland. In M. Matthes, L. Pulkkinen, C. Clouder, & B. Heys (Eds.), *Improving the quality of childhood in Europe: Volume 7*. Alliance for Childhood European Network Foundation. https://converis.jyu.fi/converis/portal/detail/Publication/28916627

Hallinger, P. (2018). Thailand's Challenge of Systemic Education Reform: Where Are the Leaders with the "Right Stuff"? In G. W. Fry, *Education in Thailand: An Old Elephant in Search of a New Mahout* (pp. 555–577). Springer Singapore.

Hallinger, P., & Kantamara, P. (2000). Leading at the Confluence of Tradition and Globalization: The Challenge of Change in Thai Schools. *Asia Pacific Journal of Education*, *20*(2), 46–57. https://doi.org/10.1080/02188791.2000.10600182

Hallinger, P., & Kantamara, P. (2001). Exploring the Cultural Context of School Improvement in Thailand. *School Effectiveness and School Improvement*, *12*(4), 385–408. https://doi.org/10.1076/sesi.12.4.385.3446

Hallinger, P., & Lee, M. (2011). A Decade of Education Reform in Thailand: Broken Promise or Impossible Dream? *Cambridge Journal of Education*, *41*(2), 139–158. https://doi.org/10.1080/0305764X.2011.572868

Hanushek, E. A. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal*, *113*(485), F64–F98.

Hanushek, E. A. (2008). Education Production Functions. In S. N. Durlauf & L. Blume (Eds.), *The new Palgrave dictionary of economics* (2nd ed). Palgrave Macmillan.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, Labor-Force Quality, and the Growth of Nations. *The American Economic Review; Nashville*, *90*(5), 1184–1208.

Hanushek, E. A., & Rivkin, S. G. (2012). The Distribution of Teacher Quality and Implications for Policy. *Annual Review of Economics*, *4*(1), 131–157. https://doi.org/10.1146/annurev-economics-080511-111001

Hanushek, E. A., & Woessmann, L. (2007). *The Role of Education Quality in Economic Growth* (No. 4122; Policy Research Working Paper). World Bank Group.

Hanushek, E. A., & Woessmann, L. (2011). Sample Selectivity and the Validity of International Student Achievement Tests in Economic Research. *Economics Letters*, *110*(2), 79–82. https://doi.org/10.1016/j.econlet.2010.10.022

Hirosato, Y., & Kitamura, Y. (Eds.). (2009). *The Political Economy of Educational Reforms and Capacity Development in Southeast Asia* (Vol. 13). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9377-7

Hofstede, G. H. (1991). *Cultures and Organizations: Software of the Mind*. McGraw-Hill.

Hopkins, D., Pennock, D., Ritzen, J., Ahtaridou, E., & Zimmer, K. (2008). *External Evaluation of the Policy Impact of PISA*. OECD. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/PISA/GB(2008)35/REV1&docLanguage=En

Huang, F. L. (2010). The Role of Socioeconomic Status and School Quality in the Philippines: Revisiting the Heyneman–Loxley Effect. *International Journal of Educational Development*, *30*(3), 288–296. https://doi.org/10.1016/j.ijedudev.2009.10.001

Huang, F. L. (2018). Multilevel Modeling and Ordinary Least Squares Regression: How Comparable Are They? *The Journal of Experimental Education*, *86*(2), 265–281. https://doi.org/10.1080/00220973.2016.1277339

Hwb. (2021). *Introduction to Curriculum for Wales guidance*. https://hwb.gov.wales/curriculum-for-wales/introduction/

ILO. (2019). *Skills Shortages and Labour Migration in the Field of Information and Communication Technology in India, Indonesia and Thailand*. International Labour Organization. https://www.ilo.org/wcmsp5/groups/public/---ed_dialogue/---sector/documents/publication/wcms_710031.pdf

IPST. (2017). *TIMSS 2015 Thailand Report*. IPST.

IPST. (2020). *[INFOGRAPHICS] Steps in Implementing PISA in Thailand*. PISA THAILAND. https://pisathailand.ipst.ac.th/infographics-pisa-thailand-workflow/

IPST & MOE. (2017). *Manual in Using Mathematics Curriculum (revised Version 2017) Based on the National Core Curriculum 2008*. Ministry of Education.

Jackson, P. A. (2004). The Thai Regime of Images. *Sojourn: Journal of Social Issues in Southeast Asia*, *19*(2), 181–218.

Jerrim, J. (2012). The Socio-Economic Gradient in Teenagers' Reading Skills: How Does England Compare with Other Countries? *Fiscal Studies*, *33*(2), 159–184. https://doi.org/10.1111/j.1475-5890.2012.00157.x

Jerrim, J. (2013). The Reliability of Trends over Time in International Education Test Scores: Is the Performance of England's Secondary School Pupils Really in Relative Decline? *Journal of Social Policy*, *42*(2), 259–279. https://doi.org/10.1017/S0047279412000827

Jerrim, J. (2016). PISA 2012: How Do Results for the Paper and Computer Tests Compare? *Assessment in Education: Principles, Policy & Practice*, *23*(4), 495–518. https://doi.org/10.1080/0969594X.2016.1147420

Jerrim, J., & Micklewright, J. (2012). Children's Cognitive Ability and Changes over Age in the Socioeconomic Gradient. In J. Ermisch, M. Jäntti, & T. Smeeding (Eds.), *From Parents to Children: The Intergenerational Transmission of Advantage* (pp. 262–284). Russell Sage Foundation. https://johnjerrim.files.wordpress.com/2013/07/crita_website.pdf

Jerrim, J., & Shure, N. (2016). *Achievement of 15-Year-Olds in England: PISA 2015 National Report*. Department for Education.

Jerrim, J., & Sims, S. (2019). Why Do so Few Low- and Middle-Income Children Attend a Grammar School? New Evidence from the Millennium Cohort Study. *British Educational Research Journal*, *45*(3), 425–457. https://doi.org/10.1002/berj.3502

Jitsuchon, S. (2012). Thailand in a Middle-income Trap. *TDRI Quarterly Review*, *27*(2), 13–20.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14–26. https://doi.org/10.3102/0013189X033007014

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, *1*(2), 112–133. https://doi.org/10.1177/1558689806298224

Joynes, C., Rossignoli, S., & Amonoo-Kuofi, E. F. (2019). *21st Century Skills: Evidence of Issues in Definition, Demand and Delivery for Development Contexts* (K4D Helpdesk Report). Institute of Development Studies.

Kanchoochat, V., & Hewison, K. (2016). Introduction: Understanding Thailand's Politics. *Journal of Contemporary Asia*, *46*(3), 371–387. https://doi.org/10.1080/00472336.2016.1173305

Kantamara, P., Hallinger, P., & Jatiket, M. (2006). Scaling-up Educational Reform in Thailand: Context, Collaboration, Networks, and Change. *Planning and Changing; Normal*, *37*(1/2), 5–23.

Kharas, H., & Kohli, H. (2011). What Is the Middle Income Trap, Why Do Countries Fall into It, and How Can It Be Avoided? *Global Journal of Emerging Market Economies*, *3*(3), 281–289. https://doi.org/10.1177/097491011100300302

Kolen, M. J., & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking: Methods and Practices* (3rd Edition). Springer.

Kondakci, Y., & Sivri, H. (2014). Salient Characteristics of High-Performing Turkish Elementary Schools. *Journal of Educational Administration*, *52*(2), 254–272. https://doi.org/10.1108/JEA-12-2012-0136

Kreiner, S. (2011). *Is the Foundation Under PISA Solid? A Critical Look at the Scaling Model Underlying International Comparisons of Student Attainment.*

Kreuter, F., & Valliant, R. (2007). A Survey on Survey Statistics: What is Done and Can be Done in Stata. *The Stata Journal*, *7*(1), 1–21. https://doi.org/10.1177/1536867X0700700101

Krueger, A., & Whitmore, D. (2000). *The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project STAR* (Working Paper No. 7656). National Bureau of Economic Research. https://doi.org/10.3386/w7656

Kyriakides, L., & Creemers, B. P. M. (2008). A Longitudinal Study on the Stability Over Time of School and Teacher Effects on Student Outcomes. *Oxford Review of Education*, *34*(5), 521–545. https://doi.org/10.1080/03054980701782064

Lao, R. (2017). *Critical Study of Thailand's Higher Education Reforms: The Culture of Borrowing*. Routledge.

Lathapipat, D., & Sondergaard, L. M. (2015). *Thailand WANTED education for all*. World Bank Group.

Leckie, G., & Goldstein, H. (2019). The Importance of Adjusting for Pupil Background in School Value-Added Models: A Study of Progress 8 and School Accountability in England. *British Educational Research Journal*, *45*(3), 518–537. https://doi.org/10.1002/berj.3511

Lee, K. J., & Simpson, J. A. (2014). Introduction to Multiple Imputation for Dealing with Missing Data. *Respirology*, *19*(2), 162–167. https://doi.org/10.1111/resp.12226

LEGO Foundation. (2022). *Rebuilding Systems – National Stories of Social and Emotional Learning Reform*. LEGO Foundation. https://cms.learningthroughplay.com/media/lyxeeqvj/rebuilding-systems-full-report.pdf

Lei, W., Li, M., Zhang, S., Sun, Y., Sylvia, S., Yang, E., Ma, G., Zhang, L., Mo, D., & Rozelle, S. (2018). Contract Teachers and Student Achievement in Rural China: Evidence from Class Fixed Effects. *Australian Journal of Agricultural and Resource Economics*, *62*(2), 299–322. https://doi.org/10.1111/1467-8489.12250

Lopez-Agudo, L. A., Jerrim, J., Gutierrez, O. D. M., & Shure, N. (2017). To Weight or Not to Weight?: The Case of PISA Data. In *Investigaciones de Economía de la Educación volume 12* (Vol. 12, pp. 285–302). Asociación de Economía de la Educación. https://ideas.repec.org/h/aec/ieed12/12-13.html

Lounkaew, K. (2013). Explaining Urban–Rural Differences in Educational Achievement in Thailand: Evidence from Pisa Literacy Data. *Economics of Education Review*, *37*, 213–225. https://doi.org/10.1016/j.econedurev.2013.09.003

Macdonald, K. (2019). PV: Stata Module to Perform Estimation with Plausible Values. In *Statistical Software Components*. Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s456951.html

Mala, D. (2016). Thai Students Plunge in PISA Test Rankings. *Bangkok Post*. https://www.bangkokpost.com/thailand/general/1154113/thai-students-plunge-in-pisa-test-rankings

Mala, D. (2018). Thailand Threatens to Quit PISA Test. *Bangkok Post*. https://www.bangkokpost.com/thailand/general/1489970/thailand-threatens-to-quit-pisa-test

Manager Online. (2019). OBEC Cheers as O-NET Scores Improved in All Subjects. *Manager Online*. https://mgronline.com/qol/detail/9620000028863

Marks, G. N. (2013). *Education, Social Background and Cognitive Ability: The Decline of the Social*. Routledge.

Marks, G. N. (2017). Is SES Really That Important for Educational Outcomes in Australia? A Review and Some Recent Evidence. *The Australian Educational Researcher*, *44*(2), 191–211. https://doi.org/10.1007/s13384-016-0219-2

Matichon. (2018, April 10). MOE to Increase Private Schools' Per Head Subsidies. *Matichon Online*. https://www.matichon.co.th/education/news_911678

Mayne, J. (2008). Contribution Analysis: An Approach to Exploring Cause and Effect. *ILAC Brief 16*. https://cgspace.cgiar.org/handle/10568/70124

Mayne, J. (2012). Contribution Analysis: Coming of Age? *Evaluation*, *18*(3), 270–280. https://doi.org/10.1177/1356389012451663

Mc Namara, M. (2010). *Exploring the Impact of Standardised Assessment in the Primary School Classroom* [National University of Ireland, Galway]. https://www.dcu.ie/sites/default/files/carpe/Mc%20Namara%20%282010%29.pdf

McCaffrey, D. F., & Hamilton, L. S. (2007). *Value-Added Assessment in Practice: Lessons from the Pennsylvania Value-Added Assessment System Pilot Project*. Rand Corporation.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., M., & Hamilton, L., S. (2003). *Evaluating Value-Added Models for Teacher Accountability*. RAND.

McNeish, D. (2014). Analyzing Clustered Data with OLS Regression: The Effect of a Hierarchical Data Structure. *Multiple Linear Regression Viewpoints*, *40*, 11–16.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd edition). SAGE Publications, Inc.

MOE. (2008). *Basic Education Core Curriculum B.E. 2551 (A.D. 2008)*. MOE.

MOE. (2012). Kru Kuen Tin. *MOE*. https://www.moe.go.th/ครูคืนถิ่น/

MOE. (2021). *Regulations of Number, Characteristics, Criteria, Recruitment Methods, Selection of Principal and Vice Principal, Term of Service, and Termination of the Committees in the Basic Education Institutions*. MOE.

http://www.dla.go.th/upload/document/type2/2021/5/25487_1_1621925552989.pdf?time=1621926835981

Mons, N. (2009). *Theoretical and Real Effects of Standardised Assessment*. Eurydice Network.

Morsy, L., Khavenson, T., & Carnoy, M. (2018). How International Tests Fail to Inform Policy: The Unsolved Mystery of Australia's Steady Decline in PISA Scores. *International Journal of Educational Development*, *60*(C), 60–79. https://doi.org/10.1016/J.IJEDUDEV.2017.10.018

Mounier, A., & Tangchuang, P. (Eds.). (2010). *Education and Knowledge in Thailand: The Quality Controversy*. Silkworm Books.

Muijs, D., & Brookman, A. (2015). Quantitative Methods. In C. Chapman (Ed.), *The Routledge International Handbook of Educational Effectiveness and Improvement: Research, Policy, and Practice* (pp. 173–201). Routledge.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2013). *TIMSS 2015 Assessment Frameworks*. TIMSS & PIRLS International Study Center ; International Association for the Evaluation of Educational Achievement.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

MWIT. (2022). *About MWIT – Mahidol Wittayanusorn School*. Mahidol Wittayanusorn School. https://www.mwit.ac.th/en/about-mwit/

Nakornthap, A. (2018). Rethinking Thai Secondary Education. In G. W. Fry, *Education in Thailand: An Old Elephant in Search of a New Mahout* (pp. 125–161). Springer Singapore.

National Strategy Secretariat Office. (2018). *National Strategy 2018-2037*. National Strategy Secretariat Office. http://nscr.nesdb.go.th/wp-content/uploads/2019/10/National-Strategy-Eng-Final-25-OCT-2019.pdf

NCES. (2021). *TIMSS - Released Assessment Questions*. NCES. https://nces.ed.gov/timss/released-questions.asp

NESDC. (2019). *Social and Quality of Life Database System—Social Indicators*. Social and Quality of Life Database System. https://www.nesdc.go.th/main.php?filename=social

NESDC. (2020). *Transition Rates at Secondary and Tertiary Levels, Academic Year 2004-2017, Yearly Data*. Social and Quality of Life Database System. http://social.nesdc.go.th/SocialStat/StatReport_Final.aspx?reportid=3768&template=2R1C&yeartype=M&subcatid=19

Nieto, S., & Ramos, R. (2015). Educational Outcomes and Socioeconomic Status: A Decomposition Analysis for Middle-Income Countries. *PROSPECTS*, *45*(3), 325–343. https://doi.org/10.1007/s11125-015-9357-y

NIETS. (2017). *NIETS Annual Report 2017*. NIETS.

NIETS. (2021). *Exam Download System: O-NET*. Exam Download System. https://www.niets.or.th/examdownload/

NSO. (2019). *Sampling Procedure of Household Socioeconomic Survey 2561*. NSO Micro Data Services System. http://ddi.nso.go.th/index.php/catalog/220/sampling

NSO. (2021). *Education Statistics*. Education Branch, Key Statistical Data. http://statbbi.nso.go.th/staticreport/page/sector/en/03.aspx

NSO & UNICEF. (2016). *Thailand Multiple Indicator Cluster Survey 2015-2016 Final Report*. NSO and UNICEF. https://mics-surveys-prod.s3.amazonaws.com/MICS5/East%20Asia%20and%20the%20Pacific/Thailand/2015-2016/Final/Thailand%202015-16%20MICS_English.pdf.pdf

NSO & UNICEF. (2020). *Multiple Indicator Cluster Survey 2019 Survey Findings Report*. NSO and UNICEF. https://www.unicef.org/thailand/media/5146/file/Multiple%20Indicator%20Cluster%20Survey%202019.pdf

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, *14*(3), 693. https://doi.org/10.2307/2525981

OBEC. (2010). *Guidelines for Curriculum Management Based on the 2008 Curriculum*. Bureau of Academic Affairs and Educational Standards.

OBEC. (2018a). *Education Management Information System: EMIS*. Education Management Information System: EMIS. https://data.bopp-obec.info/emis/index.php

OBEC. (2018b). *Proposed Roles and Responsibilities of Education Service Areas*. Ministry of Education. http://www.moe.go.th/moe/th/news/detail.php?NewsID=50284&Key=news_act

OBEC & IPST. (2017). *Indicators and Contents of Mathematics (revised Version 2017) According to the National Curriculum 2008*. MOE. http://academic.obec.go.th/newsdetail.php?id=75

O'Dwyer, L. M., & Parker, C. E. (2014). *A Primer for Analyzing Nested Data: Multilevel Modeling in SPSS using an Example from a REL Study*. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands.

OEC. (2017). *National Education Plan 2017-2036*. OEC.

OECD. (2013). *PISA 2012 Released Mathematics Items*. OECD. https://www.oecd.org/pisa/pisaproducts/pisa2012-2006-rel-items-maths-ENG.pdf

OECD. (2014). *PISA 2012 Results in Focus: What 15-Year-Olds Know and What They Can Do with What They Know*. OECD. https://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

OECD. (2016a). *PISA 2015 Results (Volume I)*. OECD Publishing. https://doi.org/10.1787/9789264266490-en

OECD. (2016b). *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. OECD. https://doi.org/10.1787/9789264267510-en

OECD. (2017a). *PISA 2015 Technical Report*. OECD. https://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf

OECD. (2017b). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*. OECD. https://doi.org/10.1787/9789264281820-en

OECD & IPST. (2018). *PISA 2015 Thailand Report: Science, Reading, and Mathematics*. IPST.

OECD & UNESCO (Eds.). (2016). *Education in Thailand: An OECD-UNESCO Perspective*. OECD.

ONESQA. (2012). *Manual for the Third External Quality Assurance (2011-2015) at Basic Education Level, School Copy (Amended November 2011)*. ONESQA.

ONESQA. (2017). *Executive Summary Results of the Third Round of External Quality Assurance (2011-2015) for Basic Schools*. ONESQA. http://www.onesqa.or.th/upload/download/201704271505239.pdf

OPEC. (2013). *Number of Schools, Students, and Teachers in Private Schools*. https://sites.google.com/site/prachasampan56/canwn-rongreiyn-nakreiyn-laea-khru-rongreiyn-xekchn

OPEC. (2021). *History of the Office of the Private Education Commission*. https://opec.go.th/%e0%b8%9b%e0%b8%a3%e0%b8%b0%e0%b8%a7%e0%b8%b1%e0%b8%95%e0%b8%b4%e0%b8%84%e0%b8%a7%e0%b8%b2%e0%b8%a1%e0%b9%80%e0%b8%9b%e0%b9%87%e0%b8%99%e0%b8%a1%e0%b8%b2/

OTEPC. (2021). *Manual for Indicators and Evaluation Methods of Performance Appraisal and Progression for Teaching Civil Servants and Civil Servants with Teaching Duties*. OTEPC. https://otepc.go.th/images/00_YEAR2564/03_PV1/1Mv9-2564.pdf

P21. (2019). *Framework for 21st Century Learning*. P21. http://static.battelleforkids.org/documents/p21/P21_Framework_Brief.pdf

Patrinos, H. A., Arcia, G., & Macdonald, K. (2015). School Autonomy and Accountability in Thailand: Does the Gap Between Policy Intent and Implementation Matter? *Prospects*, *45*(4), 429–445. http://dx.doi.org/10.1007/s11125-015-9368-8

PCSHS. (2018). *PCCPL - Princess Chulabhorn's College Phitsanulok*. PCCPL - Princess Chulabhorn's College Phitsanulok. http://www.pccpl.ac.th/pccpl2012/index.php/en/

Peddell, L., Lynch, D., Waters, R., Boyd, W., & Willis, R. (2020). How do Principals of High Performing Schools Achieve Sustained Improvement Results? *IAFOR Journal of Education*, *8*(4), 133–149. https://doi.org/10.22492/ije.8.4.08

Perera, L. D. H., & Asadullah, M. N. (2019). Mind the Gap: What Explains Malaysia's Underperformance in PISA? *International Journal of Educational Development*, *65*, 254–263. https://doi.org/10.1016/j.ijedudev.2018.08.010

Pholphirul, P. (2017). Pre-Primary Education and Long-Term Education Performance: Evidence from Programme for International Student Assessment (PISA) Thailand. *Journal of Early Childhood Research*, *15*(4), 410–432. https://doi.org/10.1177/1476718X15616834

Phuaphansawat, V. (2021). *Tackling Educational Inequity in Thailand: Wealth Effects on Schooling Choice, Learning Process and Learning Outcome*. ASEAN and Global Connections. https://www.asean-global.info/sites/default/files/2021-10/Phuaphansawat.pdf?fbclid=IwAR2MRYka5ZkadfvlfaPO5hY0cWW9azkvJSw94GuD5ZRD-LeKma7szIvf2Ns

Pitiyanuwat, S., Phanphruk, S., & Pitiyanuwat, T. (2018). Educational Testing, Assessment, and Quality Assurance. In G. W. Fry (Ed.), *Education in Thailand: An Old Elephant in Search of a New Mahout* (pp. 597–623). Springer Singapore. https://doi.org/10.1007/978-981-10-7857-6_24

Pongwat, A., & Mounier, A. (2010). Education Act 1999: A Workable Education Reform? In A. Mounier & P. Tangchuang (Eds.), *Education and Knowledge in Thailand: The Quality Controversy* (Centre IRD de Bondy; pp. 63–89). Silkworm Books.

Pra Pariyat Schools. (2022). *Hub for Religious Schools in Thailand*. Hub for Religious Schools in Thailand. http://www.prapariyat.com/index.php

Prachachat. (2022, April 29). OBEC Board Clarifies No Consensus on Using Comeptency-Based Curriculum in Schools. *Prachachat Online*. https://www.prachachat.net/education/news-921319

Pritchett, L. (2015). *Creating Education Systems Coherent for Learning Outcomes: Making the Transition from Schooling to Learning* (No. 15/005; RISE Working Paper Series). RISE.

Punyasavatsut, C. (2019). *Teacher Career Reforms Inthailand*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000370862/PDF/370862eng.pdf.multi

Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel Modelling of Complex Survey Data. *Journal of the Royal Statistical Society: Series A (Statistics in*

Society), *169*(4), 805–827. https://doi.org/10.1111/j.1467-985X.2006.00426.x

Ratchapat, A., Kessung, P., & Bhoomkhokrak, P. (2012). Factors Affecting Thai Achievement of the Students Under the Office of the Basic Education Commission: A Multilevel Analysis. *Research and Development Journal Loei Rajabhat University*, *7*(22), 13–24.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd edition). Sage Publications.

Robson, K., & Pevalin, D. J. (2016). *Multilevel Modeling in Plain Language*. SAGE.

Rowley, K. J., McNeill, S. M., Dufur, M. J., Edmunds, C., & Jarvis, J. A. (2019). Trends in International PISA Scores over Time: Which Countries Are Actually Improving? *Social Sciences*, *8*(8), Article 8. https://doi.org/10.3390/socsci8080231

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

Saavedra, A. R., & Opfer, V. D. (2012). *Teaching and Learning 21st Century Skills: Lessons from the Learning Sciences*. RAND Corporation. https://asiasociety.org/sites/default/files/T/teaching-and-learning-21st-century-skills-rand.pdf

Sakellariou, C. (2008). Peer Effects and the Indigenous/Non-Indigenous Early Test-Score Gap in Peru. *Education Economics*, *16*(4), 371–390. https://doi.org/10.1080/09645290802133065

Sakellariou, C. (2016). *The "True" Private School Effect Using PISA 2012-Mathematics: Evidence from 40 Countries* (SSRN Scholarly Paper ID 2791302). Social Science Research Network. https://doi.org/10.2139/ssrn.2791302

Sammons, P., Davis, S., Day, C., & Gu, Q. (2014). Using Mixed Methods to Investigate School Improvement and the Role of Leadership: An Example of a Longitudinal Study in England. *Journal of Educational Administration*, *52*(5), 565–589. https://doi.org/10.1108/JEA-10-2013-0121

Sangbuaphuen, R. (2020). *Towards a Competency-based Thai Basic Education Curriculum: A Conceptual Framework*. OBEC.

Sangnapaboworn, W. (2018). The Evolution of Education Reform in Thailand. In G. W. Fry, *Education in Thailand: An Old Elephant in Search of a New Mahout* (pp. 517–554). Springer Singapore.

Satimanon, T. (2017). Thailand's Labor Mismatch: Contemporary Situations and Solutions. *NIDA Case Research Journal*, *9*(1), 38.

Savin-Baden, M., & Major, C. H. (2012). *Qualitative Research: The Essential Guide to Theory and Practice*. Routledge.

SCB EIC. (2015). *Bridging Thailand's Labor Gap*. SCB. https://www.scbeic.com/en/detail/file/product/1251/e22mxi3krw/ENG_labor_insight_Q1_2015.pdf

Scheerens, J., Witziers, B., & Steen, R. (2013). A Meta-Analysis of School Effectiveness Studies. *Revista de Educación*, *361*, 619–645. https://doi.org/10.4438/1988-592X-RE-2013-361-235

Seidel, T., & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research*, *77*(4), 454–499. JSTOR.

Shaeffer, S. (2018). Preschool and Primary Education: Thailand's Progress in Achieving Education for All. In G. W. Fry (Ed.), *Education in Thailand: An Old Elephant in Search of a New Mahout* (pp. 93–124). Springer Singapore. https://doi.org/10.1007/978-981-10-7857-6_5

Siribanpitak, P. (2018). Rethinking Teacher Education. In G. W. Fry, *Education in Thailand: An Old Elephant in Search of a New Mahout* (pp. 461–476). Springer Singapore.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453. JSTOR.

Spaull, N. (2017). *Who Makes It Into PISA?: Understanding the Impact of PISA Sample Eligibility Using Turkey as a Case Study (PISA 2003—PISA 2012)* (OECD Education Working Papers No. 154). https://doi.org/10.1787/41d175fc-en

Sripahol, S. (2014). *Development of a Distance Training Package on Organizing Instruction in the Social Studies, Religion and Culture Learning Area for Teachers Teaching at the Secondary Education Level in Education Expansion Schools*. STOU. http://ird.stou.ac.th/dbresearch/otheruploads/5/%E0%B8%9A%E0%B8%97%E0%B8%97%E0%B8%B5%E0%B9%88%201.pdf

Suter, K. D. (1991). Thailand: The Fifth Asian Tiger. *The Contemporary Review; London*, *258*(1502), 120–125.

Tangkijvanich, S., & Sasiwuttiwat, S. (2012). Revamping the Thai Education System: Quality for All. *TDRI Quarterly Review*, *27*(2), 3–12.

Tashakkori, A., & Teddlie, C. (2010). *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. SAGE Publications, Inc. https://doi.org/10.4135/9781506335193

Thairath. (2018). Who says Thai PISA results are bad? Help restore faith. *Thairath*. https://www.thairath.co.th/news/local/1434856

TIMSS & PIRLS International Study Center. (2017). *TIMSS 2015 Assessment Items*. TIMSS & PIRLS. http://timss.bc.edu/timss2015/international-database/assessment-items.html

Todd, P. E., & Wolpin, K. I. (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital*, *1*(1), 91–136. https://doi.org/10.1086/526401

Tongliemnak, P. (2018). Thai Students and PISA Tests: Illusion and Reality. *The101.World I Creative Knowledge Media for Social Change*. https://www.the101.world/thai-student-and-pisa-myth/

Townsend, T. (Ed.). (2007). *International Handbook of School Effectiveness and Improvement* (Vol. 17). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5747-2

UNESCO-UIS. (2022). *Indicator 4.1.1 – Global Alliance to Monitor Learning*. Indicator 4.1.1 – Global Alliance to Monitor Learning. https://gaml.uis.unesco.org/4-1-1/

UNICEF. (2017). *Public Expenditure Tracking Survey (PETS) on the 15-Year Free Education Program: Kingdom of Thailand*. UNICEF.

Valant, J. (2019). What Are Charter Schools and Do They Deliver? *Brookings*. https://www.brookings.edu/policy2020/votervital/what-are-charter-schools-and-do-they-deliver/

Varavarn, K. (2006). *Thailand's Experience in Addressing the Challenges of Secondary Education*. Ministry of Education.

Warr, P. G. (Ed.). (2005). *Thailand Beyond the Crisis*. Routledge.

West, A., & Wolfe, D. (2018). *Academies, the School System in England and a Vision for the Future* (Clare Market Papers No. 23). LSE Academic Publishing. http://eprints.lse.ac.uk/88240/1/Academies%20Vision%20Report%202%20JUNE.pdf

West, B. T., Beer, L., Gremel, G. W., Weiser, J., Johnson, C. H., Garg, S., & Skarbinski, J. (2015). Weighted Multilevel Models: A Case Study. *American Journal of Public Health*, *105*(11), 2214–2215. https://doi.org/10.2105/AJPH.2015.302842

World Bank. (2008). *Thailand—Investment Climate Assessment Update*. World Bank. http://documents.worldbank.org/curated/en/268141468120847586/Thailand-Investment-climate-assessment-update

World Bank. (2012a). *Learning Outcomes in Thailand: What Can We Learn from International Assessments?* World Bank Group. http://documents.worldbank.org/curated/en/284911468132280900/Learning-outcomes-in-Thailand-what-can-we-learn-from-international-assessments

World Bank. (2012b). *School Autonomy and Accountability SABER Country Report: Thailand*. World Bank. http://wbgfiles.worldbank.org/documents/hdn/ed/saber/supporting_doc/CountryReports/SAA/SABER_SAA_Thailand_CR_Final_2012.pdf

World Bank. (2018). *World Development Report 2018: Learning to Realize Education's Promise* [Text/HTML]. World Bank Group. http://www.worldbank.org/en/publication/wdr2018

World Bank. (2019a). *Income Share Held by Highest 10%*. Poverty and Inequality Platform, World Bank. https://data.worldbank.org/indicator/si.dst.10th.10

World Bank. (2019b). *School Enrollment, Secondary (% Gross)*. Poverty and Inequality Platform, World Bank. https://data.worldbank.org/indicator/SE.SEC.ENRR?locations=TH

World Bank. (2020). *Thailand Economic Monitor: Productivity for Prosperity* (No. 145281). The World Bank. http://documents.worldbank.org/curated/en/394501579357102381/Thailand-Economic-Monitor-Productivity-for-Prosperity

Wu, M. (2010). *Comparing the Similarities and Differences of PISA 2003 and TIMSS* (No. 32; OECD Education Working Papers). OECD. https://www.oecd-ilibrary.org/education/comparing-the-similarities-and-differences-of-pisa-2003-and-timss_5km4psnm13nx-en

Yang, F. M., & Kao, S. T. (2014). Item Response Theory for Measurement Validity. *Shanghai Archives of Psychiatry*, *26*(3), 171–177. https://doi.org/10.3969/j.issn.1002-0829.2014.03.010

Appendix A: Appendix to Chapter One

Table A-1 and A-2 shows differences in learning units and standards between the 2008 and 2017 curriculum. The descriptions are taken directly from the curriculum documents of 2008 (MOE, 2008) and 2017 (IPST & MOE, 2017). As there are no English version of the curriculum available at the time of writing, the learning descriptions in the 2017 curriculum were translated from Thai to English by the researcher. To maintain consistency, the translation was derived from the 2008 English version of the curriculum when possible.

Table A- 1 Learning units and standards in Mathematics of the 2008 and 2017 curriculum

2008 curriculum units	2017 curriculum units
Numbers and Operations Standard 1: Understanding diverse methods of presenting numbers and their application in real life Standard 2: Understanding results of operations of numbers, relationships of operations, and application of operations for problem-solving Standard 3: Use of estimation in calculation and problem-solving Standard 4: Understanding of numerical system and application of numerical properties	**Numbers and Algebra** Standard 1: Understanding diverse methods of presenting numbers, number systems, operations of numbers, results of operations, properties of operations and their application Standard 2: Understanding and analysing pattern, relation and function, sequences and series and their application Standard 3: Ability to apply algebraic expressions, equations, and inequalities to represent relationships or application for problem-solving
Algebra Standard 1: Understanding and ability to analyse pattern, relation and function Standard 2: Ability to apply algebraic expressions, equations, inequalities, graphs and other mathematical models to represent various situations, as well as interpretation and application for problem-solving	

2008 curriculum units	2017 curriculum units
Measurement Standard 1: Understanding the basics of measurement; ability to measure and estimate the size of objects to be measured Standard 2: Solving measurement problems	**Measurement and Geometry** Standard 1: Understanding the basics of measurement; ability to measure and estimate the size of objects to be measured and application Standard 2: Understanding and analysing geometric figures, properties of geometric figures, relationships between geometric figures, and geometric theories, and application
Geometry Standard 1: Ability to explain and analyse two-dimensional and three-dimensional geometric figures Standard 2: Ability for visualisation, spatial reasoning and application of geometric models for problem-solving	
Data Analysis and Probability Standard 1: Understanding and ability to apply statistical methodology for data analysis Standard 2: Application of statistical methodology and knowledge of probability for valid estimation	**Statistics and Probability** Standard 1: Understanding statistical methodology and using statistical knowledge to solve problems Standard 2: Understanding basic counting principles, probability, and application
Mathematical Skills and Processes Standard 1: Capacity for problem-solving, reasoning, and communication; communication and presentation of mathematical concepts; linking various bodies of	N/A

2008 curriculum units	2017 curriculum units
mathematical knowledge and linking mathematics with other disciplines; and attaining ability for creative thinking	

Sources: Thai national curricula (MOE, 2008; OBEC & IPST, 2017)

Table A- 2 Learning units and standards in Science of the 2008 and 2017 curriculum

2008 curriculum units	2017 curriculum units
Living Things and Processes of Life Standard 1: Understanding basic units of living things; relationship between structures and functions of living things, which are interlinked; investigative process for seeking knowledge; ability to communicate acquired knowledge that could be applied to one's life and care for living things Standard 2: Understanding of process and importance of genetic transmission; evolution of living things; biodiversity; application of biotechnology affecting humans and the environment; investigative process for seeking knowledge and scientific mind; communicating knowledge that could be applied for useful purposes	**Biological Science** Standard 1: Understanding diversity of the eco-system, the relationship between living things and non-living things, and the relationship between living things in the eco-system, energy transmission, replacement processes in the eco-system, meaning of population, problems and impacts on natural resources and the environment, ways in preserving the natural resources and solving environmental problems, including application of knowledge Standard 2: Understanding properties of living things, basic units of living things, transportation of substance into and out of cells. The relationship between structure and

2008 curriculum units	2017 curriculum units
	responsibilities of interrelated systems in animals and humans, relationship of structures and responsibilities of various interrelated organs in plants, including application of knowledge Standard 3: Understanding of process and importance of genetic transmission, genetic material, genetic mutations that affect living things, biodiversity and evolution of living things, including application of knowledge
Life and the Environment Standard 1: Understanding of local environment; relationship between the environment and living things; relationship between living things in the eco-system; investigative process for seeking knowledge and scientific mind; and communicating acquired knowledge that could be applied for useful purposes Standard 2: Appreciating the importance of natural resources; utilization of natural resources at local, national and global levels; and application of knowledge for management of natural resources and local environment on a sustainable basis	
Substances and Properties of Substances Standard 1: Understanding of properties of substances; relationship between properties of substances and structures and binding forces between particles; investigative process for seeking knowledge and scientific	**Physical Science** Standard 1: Understanding of properties of substances, components of substances, relationship between properties of substances and structures and binding forces between particles, principles and nature of change in the state of

2008 curriculum units	2017 curriculum units
mind; and communicating acquired knowledge for useful purposes Standard 2: Understanding of principles and nature of change in the state of substances; solution formation; reaction; investigative process for seeking knowledge and scientific mind; and communication of acquired knowledge that could be applied for useful purposes	substances, solution formation and chemical reaction formation Standard 2: Understanding the nature of force in everyday life, the effect of force on objects, types of motion of natural objects, including application of knowledge Standard 3: Understanding the meaning of energy, energy transformation and transfer, interrelationship between substances and energy, energy in everyday life, nature of waves, phenomenon related to sound, light, and electromagnetic, including application of knowledge
Forces and Motion Standard 1: Understanding of the nature of electromagnetic, gravitational and nuclear forces; investigative process of seeking knowledge and applying acquired knowledge for useful and ethical purposes Standard 2: Understanding of characteristics and various types of motion of natural objects; investigative process for seeking knowledge and scientific mind; and communication of acquired knowledge for useful purposes	
Energy Standard 1: Understanding of relationship between energy and life; energy transformation; interrelationship between	

2008 curriculum units	2017 curriculum units
substances and energy; effects of energy utilization on life and the environment; investigative process for seeking knowledge; and communication of acquired knowledge that could be applied for useful purposes	
Change Process of the Earth Standard 1: Understanding of various processes on the Earth's surface and inside the Earth; relationship between various processes causing changes in climate, topography and form of the Earth; investigative process for seeking knowledge and scientific mind; and communication of acquired knowledge that could be applied for useful purposes	**Earth and Space Science** Standard 1: Understanding of components, characteristics, the process of birth and evaluation of the universe, galaxies, stars, and the solar system, including interrelationships within the solar system that affect living things on Earth, and application of space technology Standard 2: Understanding of components and relationships of Earth systems, change process inside and on Earth's surface, earthquakes, changes in atmosphere and climate, including their effects on living things and the environment
Astronomy and Space Standard 1: Understanding of evolution of the solar system, galaxies and the universe; interrelationships within the solar system and their effects on living things on Earth; investigative process for seeking knowledge and scientific mind; and communication of acquired knowledge for useful purposes	

2008 curriculum units	2017 curriculum units
Standard 2: Understanding of importance of space technology utilized for space exploration and natural resources for agriculture and communication; investigative process for seeking knowledge and scientific mind; and communication of acquired knowledge that could be ethically applied to life and the environment	
Nature of Science and Technology Standard 1: Application of scientific process and scientific mind in investigation for seeking knowledge and problem-solving; knowing that most natural phenomena assume definite patterns that are explainable and verifiable within limitations of data and instruments available during particular periods of time; and understanding that science, technology, society and the environment are interrelated	N/A
N/A	**Technology** Standard 1: Understanding the main principles of technology for living in a fast-changing society, using knowledge and skills in Science, Mathematics, and other disciplines to solve problems or improve work creatively

2008 curriculum units	2017 curriculum units
	using engineering design process, choose appropriate technology considering its effects on life, society, and environment Standard 2: Understanding and using computational thinking to solve everyday problems systematically, using information technology in learning, working, and problem-solving effectively, smart, and morally sound

Sources: Thai national curricula (MOE, 2008; OBEC & IPST, 2017)

Appendix B: Appendix to Chapter Two

This Appendix provides detailed descriptions of cognitive domains, content domains, and the proficiency levels of the three assessments. PISA's domains and proficiency levels are taken from PISA reports of 2012 (OECD, 2014) and 2015 (OECD, 2017b). TIMSS's domains and proficiency levels are from its released assessment framework (Mullis & Martin, 2013). O-NET's domains are from the curriculum of 2008 (MOE, 2008). There are no cognitive domains and proficiency levels in O-NET.

Appendix B1: Cognitive domains

PISA: Mathematics

PISA assesses students' performance in mathematics through questions related to processes, content, and context. Process describes what students do to connect the context of a problem with the mathematics involved and thus solve the problem.

1. **Formulating situations mathematically**
    - Identifying the mathematical aspects of a problem situated in a real-world context and identifying the significant variables
    - Recognising mathematical structure (including regularities, relationships and patterns) in problems or situations
    - Simplifying a situation or problem in order to make it amenable to mathematical analysis
    - Identifying constraints and assumptions behind any mathematical modelling and simplifications gleaned from the context

- Representing a situation mathematically, using appropriate variables, symbols, diagrams and standard models
- Representing a problem in a different way, including organising it according to mathematical concepts and making appropriate assumptions
- Understanding and explaining the relationships between the context-specific language of a problem and the symbolic and formal language needed to represent it mathematically
- Translating a problem into mathematical language or a representation
- Recognising aspects of a problem that correspond with known problems or mathematical concepts, facts or procedures
- Using technology (such as a spreadsheet or the list facility on a graphing calculator) to portray a mathematical relationship inherent in a contextualised problem.

2. Employing mathematical concepts, facts, procedures and reasoning

- Devising and implementing strategies for finding mathematical solutions
- Using mathematical tools, including technology, to help find exact or approximate solutions
- Applying mathematical facts, rules, algorithms and structures when finding solutions
- Manipulating numbers, graphical and statistical data and information, algebraic expressions and equations, and geometric representations
- Making mathematical diagrams, graphs and constructions, and extracting mathematical information from them
- Using and switching between different representations in the process of finding solutions
- Making generalisations based on the results of applying mathematical procedures to find solutions
- Reflecting on mathematical arguments and explaining and justifying mathematical results.

3. Interpreting, applying and evaluating mathematical outcomes

- Interpreting a mathematical result back into the real-world context
- Evaluating the reasonableness of a mathematical solution in the context of a real-world problem
- Understanding how the real world impacts the outcomes and calculations of a mathematical procedure or model in order to make contextual judgements about how the results should be adjusted or applied
- Explaining why a mathematical result or conclusion does, or does not, make sense given the context of a problem
- Understanding the extent and limits of mathematical concepts and mathematical solutions
- Critiquing and identifying the limits of the model used to solve a problem.

PISA: Science

Scientific literacy is the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen. A scientifically literate person is willing to engage in reasoned discourse about science and technology, which requires the competencies to:

1. **Explain phenomena scientifically** – recognise, offer and evaluate explanations for a range of natural and technological phenomena.
2. **Evaluate and design scientific enquiry** – describe and appraise scientific investigations and propose ways of addressing questions scientifically.
3. **Interpret data and evidence scientifically** – analyse and evaluate data, claims and arguments in a variety of representations and draw appropriate scientific conclusions.

In addition, each item can also be mapped using a third dimension based on a depth-of-knowledge taxonomy. This provides a means of operationalising **cognitive demand** as each item can be categorised as making demands that are:

1. **Low**: Carry out a one-step procedure, for example recall a fact, term, principle or concept, or locate a single point of information from a graph or table.
2. **Medium**: Use and apply conceptual knowledge to describe or explain phenomena, select appropriate procedures involving two or more steps, organise/display data, interpret or use simple data sets or graphs.
3. **High**: Analyse complex information or data; synthesise or evaluate evidence; justify; reason, given various sources; develop a plan or sequence of steps to approach a problem.

TIMSS: Mathematics

The cognitive domains are the following.

1. **Knowing**
    - **Recall**: Recall definitions, terminology, number properties, units of measurement, geometric properties, and notation (e.g., a × b = ab, a + a + a = 3a).
    - **Recognize**: Recognize numbers, expressions, quantities, and shapes. Recognize entities that are mathematically equivalent (e.g., equivalent familiar fractions, decimals, and percents; different orientations of simple geometric figures).
    - **Classify/Order:** Classify numbers, expressions, quantities, and shapes by common properties.
    - **Compute:** Carry out algorithmic procedures for +, –, ×, ÷, or a combination of these with whole numbers, fractions, decimals, and integers. Carry out straightforward algebraic procedures.
    - **Retrieve**: Retrieve information from graphs, tables, texts, or other sources. Measure Use measuring instruments; and choose appropriate units of measurement.
2. **Applying**
    - **Determine**: Determine efficient/appropriate operations, strategies, and tools for solving problems for which there are commonly used methods of solution.
    - **Represent/Model**: Display data in tables or graphs; create equations, inequalities, geometric figures, or diagrams that model problem situations; and generate equivalent representations for a given mathematical entity or relationship.
    - **Implement**: Implement strategies and operations to solve problems involving familiar mathematical concepts and procedures
3. **Reasoning**

- **Analyze**: Determine, describe, or use relationships among numbers, expressions, quantities, and shapes.
- **Integrate/Synthesize**: Link different elements of knowledge, related representations, and procedures to solve problems.
- **Evaluate**: Evaluate alternative problem solving strategies and solutions. Draw Conclusions Make valid inferences on the basis of information and evidence.
- **Generalize**: Make statements that represent relationships in more general and more widely applicable terms.
- **Justify:** Provide mathematical arguments to support a strategy or solution.

TIMSS: Science

The cognitive domains are the following.

1. **Knowing**
   - **Recall/Recognize:** Identify or state facts, relationships, and concepts; identify the characteristics or properties of specific organisms, materials, and processes; identify the appropriate uses for scientific equipment and procedures; and recognize and use scientific vocabulary, symbols, abbreviations, units, and scales.
   - **Describe**: Describe or identify descriptions of properties, structures, and functions of organisms and materials, and relationships among organisms, materials, and processes and phenomena.
   - **Provide Examples**: Provide or identify examples of organisms, materials, and processes that possess certain specified characteristics; and clarify statements of facts or concepts with appropriate examples.
2. **Applying**

- **Compare/Contrast/ Classify**: Identify or describe similarities and differences between groups of organisms, materials, or processes; and distinguish, classify, or sort individual objects, materials, organisms, and process based on given characteristic and properties.
- **Relate**: Relate knowledge of an underlying science concept to an observed or inferred property, behavior, or use of objects, organisms, or materials.
- **Use Models**: Use a diagram or other model to demonstrate knowledge of science concepts, to illustrate a process cycle relationship, or system, or to find solutions to science problems.
- **Interpret Information**: Use knowledge of science concepts to interpret relevant textual, tabular, pictorial, and graphical information.
- **Explain**: Provide or identify an explanation for an observation or a natural phenomenon using a science concept or principle.

3. **Reasoning**
    - **Analyze**: Identify the elements of a scientific problem and use relevant information, concepts, relationships, and data patterns to answer questions and solve problems.
    - **Synthesize**: Answer questions that require consideration of a number of different factors or related concepts.
    - **Formulate Questions/ Hypothesize/Predict**: Formulate questions that can be answered by investigation and predict results of an investigation given information about the design; formulate testable assumptions based on conceptual understanding and knowledge from experience, observation, and/or analysis of scientific information; and use evidence and conceptual understanding to make predictions about the effects of changes in biological or physical conditions.
    - **Design Investigations**: Plan investigations or procedures appropriate for answering scientific questions or testing hypotheses; and describe or recognize the characteristics of well-designed investigations in terms of

variables to be measured and controlled and cause-and-effect relationships.

- **Evaluate**: Evaluate alternative explanations; weigh advantages and disadvantages to make decisions about alternative processes and materials; and evaluate results of investigations with respect to sufficiency of data to support conclusions.
- **Draw Conclusions**: Make valid inferences on the basis of observations, evidence, and/or understanding of science concepts; and draw appropriate conclusions that address questions or hypotheses, and demonstrate understanding of cause and effect.
- **Generalize**: Make general conclusions that go beyond the experimental or given conditions; apply conclusions to new situations.
- **Justify**: Use evidence and science understanding to support the reasonableness of explanations, solutions to problems, and conclusions from investigations.

Appendix B2: Content domains

PISA: Mathematics

The content domains are not written based on the outlined curriculum knowledge. However, the parallel can still be found between PISA and curriculum contents. These are underlined below.

1. **Change and relationships**

Being more literate about change and relationships involves understanding fundamental types of change and recognising when they occur in order to use suitable mathematical models to describe and predict change. Mathematically this means modelling the change and the relationships with appropriate functions and equations, as well as creating, interpreting, and translating among symbolic and graphical representations of relationships. Aspects of the traditional mathematical content of <u>functions and algebra, including algebraic expressions, equations and inequalities, tabular and graphical representations</u>, are central in describing, modelling and interpreting change phenomena. Representations of data and relationships described using <u>statistics</u> also are often used to portray and interpret change and relationships, and a firm grounding in the <u>basics of number and units</u> is also essential to defining and interpreting change and relationships. Some interesting relationships arise from geometric measurement, such as the way that changes in perimeter of a family of shapes might relate to changes in area, or the relationships among lengths of the sides of triangles.

2. **Space and shape**

<u>Geometry</u> serves as an essential foundation for space and shape, but the category extends beyond traditional geometry in content, meaning and method, drawing on elements of other mathematical areas such as <u>spatial visualisation, measurement and algebra</u>. For instance, shapes can change, and a point can move along a locus, thus requiring function concepts. <u>Measurement formulas</u> are central in this

area. The manipulation and interpretation of shapes in settings that call for tools ranging from dynamic geometry software to Global Positioning System (GPS) software are included in this content category. PISA assumes that the understanding of a set of core concepts and skills is important to mathematical literacy relative to space and shape. Mathematical literacy in the area of space and shape involves a range of activities such as understanding perspective (for example in paintings), creating and reading maps, transforming shapes with and without technology, interpreting views of three-dimensional scenes from various perspectives and constructing representations of shapes.

3. Quantity

It incorporates the quantification of attributes of objects, relationships, situations and entities in the world, understanding various representations of those quantifications, and judging interpretations and arguments based on quantity. To engage with the quantification of the world involves understanding measurements, counts, magnitudes, units, indicators, relative size, and numerical trends and patterns. Aspects of quantitative reasoning – such as number sense, multiple representations of numbers, elegance in computation, mental calculation, estimation and assessment of reasonableness of results – are the essence of mathematical literacy relative to quantity. Quantification is a primary method for describing and measuring a vast set of attributes of aspects of the world. It allows for the modelling of situations, for the examination of change and relationships, for the description and manipulation of space and shape, for organising and interpreting data, and for the measurement and assessment of uncertainty. Thus mathematical literacy in the area of quantity applies knowledge of number and number operations in a wide variety of settings.

4. Uncertainty and data

The uncertainty and data content category includes recognising the place of variation in processes, having a sense of the quantification of that variation, acknowledging uncertainty and error in measurement, and knowing about chance. It also includes forming, interpreting and evaluating conclusions drawn in situations where uncertainty is central. The traditional curricular areas of <u>probability and statistics</u> provide formal means of describing, modelling and interpreting a certain class of uncertainty phenomena, and for making inferences. In addition, knowledge of <u>number and of aspects of algebra, such as graphs and symbolic representation</u>, contribute to facility in engaging in problems in this content category. The focus on the interpretation and presentation of data is an important aspect of the uncertainty and data category.

PISA: Science

The content domains are the following.

1. **Physical systems**
   - Structure of matter (e.g. particle model, bonds)
   - Properties of matter (e.g. changes of state, thermal and electrical conductivity)
   - Chemical changes of matter (e.g. chemical reactions, energy transfer, acids/bases)
   - Motion and forces (e.g. velocity, friction) and action at a distance (e.g. magnetic, gravitational and electrostatic forces)
   - Energy and its transformation (e.g. conservation, dissipation, chemical reactions)
   - Interactions between energy and matter (e.g. light and radio waves, sound and seismic waves)
2. **Living systems**
   - Cells (e.g. structures and function, DNA, plant and animal)
   - The concept of an organism (e.g. unicellular and multicellular)
   - Humans (e.g. health, nutrition, subsystems such as digestion, respiration, circulation, excretion, reproduction and their relationship)
   - Populations (e.g. species, evolution, biodiversity, genetic variation)
   - Ecosystems (e.g. food chains, matter and energy flow)

- Biosphere (e.g. ecosystem services, sustainability)

3. **Earth and space systems** Structures of the Earth systems (e.g. lithosphere, atmosphere, hydrosphere)
   - Energy in the Earth systems (e.g. sources, global climate)
   - Change in Earth systems (e.g. plate tectonics, geochemical cycles, constructive and destructive forces)
   - Earth's history (e.g. fossils, origin and evolution)
   - Earth in space (e.g. gravity, solar systems, galaxies)
   - The history and scale of the universe and its history (e.g. light year, Big Bang theory)

TIMSS: Mathematics

The content domains are the following.

1. **Numbers**: Whole numbers, Fractions, decimals, and integers; and Ratio, proportion, and percent.
2. **Algebra**: Expressions and operations, Equations and inequalities, and Relationships and functions
3. **Geometry**: Geometric shapes; Geometric measurement; and Location and movement
4. **Data and chance**: Characteristics of data sets, Data interpretation, and Chance

TIMSS: Science

The content domains are the following.

1. **Biology:** Characteristics and life processes of organisms, Cells and their functions, Life cycles, reproduction, and heredity, Diversity, adaptation, and natural selection, Ecosystems, and Human health.
2. **Chemistry**: Composition of matter, Properties of matter, and Chemical change
3. **Physics**: Physical states and changes in matter, Energy transformation and transfer, Light and sound, Electricity and magnetism, and Forces and motion
4. **Earth Science**: Earth's structure and physical features, Earth's processes, cycles, and history, Earth's resources, their use and conservation, and Earth in the solar system and the universe.

O-NET: Mathematics

The learning areas outlined in the national basic curriculum are as follows;

1. **Numbers and Operations**: the numerical concepts and sense of perception; real number system; the properties of real numbers; the operation of numbers; ratio; percentage; problem-solving involving numbers; and the application of numbers for real life
2. **Measurement**: length; distance; weight; area; volume and capacity; money and time; measuring units; estimation for measurement; trigonometric ratio; problem-solving regarding measurement; and application of measurement for various situations
3. **Geometry**: the geometric figures and the properties of one-dimensional geometric figures; visualisation of geometric models; geometric theories; and geometric transformation through translation, reflection and rotation
4. **Algebra**: pattern; relationship; function; sets and their operations; reasoning; expression; equation; equation system; inequality; graph; arithmetic order; geometric order; arithmetic series; and geometric series
5. **Data Analysis and Probability**: determining an issue; writing questions; determining methods of study; data collection; systematisation and presentation; central tendency and data distribution; data analysis and interpretation; opinion polling; probability; application of statistical knowledge and probability; application of probability for explaining various situations as well as for facilitating decision-making for real life
6. **Mathematical Skills and Processes**: problem-solving through diverse methods; reasoning; communication; presentation of mathematical concepts; linking mathematics with other disciplines; and attaining ability for creative thinking

O-NET: Science

The learning area of science is aimed at enabling learners to link knowledge with processes, acquire essential skills for investigation, build up knowledge through investigative processes, seek knowledge and solve various problems. Learners are allowed to participate in all stages of learning, with activities organized through diverse practical work suitable to their levels. The main content areas are prescribed as follows;

1. **Living Things and Processes of Life**: living things; basic units of living things; the structures and functions of various systems of living things and the processes of life; the biodiversity; genetic transmission; the functioning of various systems of living things, the evolution and diversity of living things and the biotechnology
2. **Life and the Environment**: diverse living things in the environment; the relationship between living things and the environment; the relationships among living things in the eco-system; the importance of natural resources, utilization and management of natural resources at local, national and global levels; the factors affecting survival of living things in various environments
3. **Substances and Properties of Substances**: the properties of materials and substances; binding forces between particles; changes in the state of substances; the solution formation and chemical reaction of substances, chemical equations and separation of substances
4. **Forces and Motion**: the nature of electromagnetic, gravitational and nuclear forces; forces acting on objects; motion of objects; frictional forces; moment of variety of motions in daily life
5. **Energy**: energy and life; energy transformation; the properties and phenomena of light, sound, electrical circuits, electromagnetic waves, radioactivity and nuclear reactions; the interrelationship between

substances and energy; energy conservation; the effects of utilization of energy on life and the environment

6. **Change Process of the Earth**: the structure and components of the Earth; geological resources; the physical properties of soil, rock, water and air; the properties of the earth surface and atmosphere; change processes of the earth's crust; the geological phenomena; the factors affecting atmospheric change
7. **Astronomy and Space**: evolution of the solar system; galaxies; the universe; interrelationship and effects on living things on the earth; the relationship between the sun, the moon and the earth; the importance of space technology
8. **Nature of Science and Technology**: the scientific processes; investigation for seeking knowledge, problem-solving, and scientific mind

Appendix B3: IRT criticisms

The item-response theory is not without its critiques. The model relies on assumptions of monotonicity (as the underlying trait level increases, the probability of getting a correct response increases), unidimensionality (there is one underlying trait the instrument is measuring), local independence (responses to different items are mutually independent), and item invariance (the trait is independent of sample characteristics) (Yang & Kao, 2014). There are contestations mainly on the assumptions of item invariance and unidimensionality. Kreiner (2011) raised concerns on whether the items were equally difficult across countries, or whether there is Differential Item Functioning (DIF) in PISA 2006's Reading items. When there is DIF, the items may not have the same level of difficulty across countries, which creates concerns over between-country comparisons. The study shows that there is evidence of DIF, and this affects the ranking of countries significantly. Changing the mode of test administration (from paper-based to computer-based) is also found to create Differential Item Functioning in the Netherlands (Feskens et al., 2019). Differential item functioning, if happens over time, could also bias comparison over time as the change in scores over time may be from either test takers becoming more able or the items becoming easier (Goldstein & Wood, 1989). Additionally, Goldstein (2017) critiques the notion of unidimensionality. He asserted that by assuming a single underlying trait to be measured, it can bias the test construction process since the items that correlate weakly with this trait tends to be rejected. Hence, the results tend to be biased against countries that may perform well in those items due probably to cultural or linguistic idiosyncrasies.

Assessment results may be dependent on the item-response model used in the analysis. G. Brown et al. (2007) re-analysed data of TIMSS 1995 using a different specification of the item-response model (one- and three-parameter models). Generally, the mean scores remain similar, except for some extreme case such as South Africa, where the score decreased significantly after guessing was accounted for in the three-parameter model. The authors hypothesised that accounting for guessing reveals the previously masked low performers. The measure of dispersion (P95 – P5), however, changed significantly for all countries. There is zero correlation between the one- and three-parameter models. This means that the spread of scores is sensitive to the model choice.

Appendix B4: Proficiency level descriptions

Table B4- 1 PISA Mathematics proficiency level descriptions

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.

Level	**Lower score limit**	**Characteristics of tasks**
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.

Level	Lower score limit	Characteristics of tasks
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.

Source: OECD (2016a, p. 191)

Note that level 1a and 1b was only made in 2015.

Table B4- 2 PISA Science proficiency level descriptions

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

Level	Lower score limit	Characteristics of tasks
		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

Level	**Lower score limit**	**Characteristics of tasks**
		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.

Source: OECD (2016a, p. 60)

Table B4- 3 TIMSS Mathematics proficiency level descriptions

Level	Lower score limit	Characteristics of tasks
4 Advanced	625	Can apply and reason in a variety of problem situations, solve linear equations, and make generalisations.
3 High	550	Can apply understanding and knowledge in a variety of relatively complex situations.
2 Intermediate	475	Can apply basic Mathematical knowledge in a variety of situations.
1 Low	400	Have some knowledge of whole numbers and basic graphs.

Source: Mullis and Martin (2013)

Table B4- 4 TIMSS Science proficiency level descriptions

Level	Lower score limit	Characteristics of tasks
4 Advanced	625	Students communicate understanding of complex concepts related to biology, chemistry, physics, and Earth science in practical, abstract, and experimental concepts.
3 High	550	Students apply and communicate understanding of concepts from biology, chemistry, physics, and Earth science in everyday and abstract situations.
2 Intermediate	475	Students demonstrate and apply their knowledge of biology, chemistry, physics, and Earth science in various contexts.
1 Low	400	Students show some basic knowledge of biology, chemistry, physics, and Earth science.

Source: Mullis and Martin (2013)

Appendix C: Appendix to Chapter Four

Appendix C1: O-NET data details

Table C1-1 shows number of schools and students participating in the examination each year from 2008-2017. The data is very rich, comprising of all schools in Thailand. Here, it can be observed that there are significantly greater number of primary schools (at grade six) comparing to the other grade levels. At grade nine, less students are in the education system. They are all concentrated in only around 4,000 schools across the country.

Table C1- 1 Number of Schools and Students taking O-NET Exams from 2008-2017

Years	Grade 6		Grade 9		Grade 12	
	# of schools	# of students	# of schools	# of students	# of schools	# of students
2008	31,660	N/A[158]	11,508	N/A	3,323	N/A
2009	31,652	N/A	11,590	N/A	3,408	N/A
2010	31,626	805,073	11,706	804,895	3,507	357,050
2011	31,619	780,370	11,817	790,325	3,582	372,662
2012	31,537	773,337	11,841	754,501	3,684	394,664
2013	31,422	734,820	11,838	680,652	3,755	414,984
2014	31,333	728,480	11,835	664,952	3,807	431,792
2015	31,306	716,784	11,865	656,817	3,846	423,654
2016	N/A	724,349	N/A	637,491	N/A	379,064
2017	N/A	704,705	N/A	643,904	N/A	372,853

Source: Own calculation using O-NET microdata

[158] Data not publicly reported

Following the sample restriction criteria mentioned in section 4.3.3.3, the final sample has the following characteristics.

Table C1- 2 Number of observations before and after data cleaning and restrictions

Year	Initial sample	Excluding those without unique ID	Excluding those with missing scores	Percentage dropped
2010	805,392	805,392	725,200	10%
2011	792,837	792,837	728,293	8%
2012	755,365	755,365	692,506	8%
2013	681,800	681,800	625,999	8%
2014	666,398	666,398	621,875	7%
2015	657,625	657,625	628,302	4%
2016	690,365	659,833	603,601	13%
2017	698,083	668,610	605,435	13%

Source: Own calculation using O-NET microdata

Table C1- 3 Summary statistics of O-NET continuous variables

Variables	Obs.	Mean	Std. Dev.	Min	Max
2012					
Mathematics score	692,506	27.11	10.75	0	100
Science score	692,506	35.60	11.80	0	98
Grade 6 Mathematics score	692,506	37.29	18.13	0	100
Grade 6 Science score	692,506	40.15	14.76	0	100
Grade 6 Thai score	692,506	39.38	10.29	10	97
2015					
Mathematics score	628,302	32.64	15.10	0	100
Science score	628,302	37.87	13.45	0	100
Grade 6 Mathematics score	628,302	37.43	17.09	0	100
Grade 6 Science score	628,302	38.70	12.76	0	95
Grade 6 Thai score	628,302	47.25	14.51	0	100

Source: Own calculation using O-NET microdata

Table C1- 4 Summary statistics of O-NET categorical variables

Variables	Proportion 2012		Percent 2012	Proportion 2015		Percent 2015
School type	OBEC1	148,490	21.44%	OBEC1	127,207	20.25%
	OBEC2	385,109	55.61%	OBEC2	349,599	55.64%
	Private	98,739	14.26%	Private	95,115	15.14%
	BMA	8,582	1.24%	BMA	7,984	1.27%
	Local	45,309	6.54%	Local	42,869	6.82%
	Satit	4,292	0.62%	Satit	1,181	0.19%
	Science	1,985	0.29%	Science	4,347	0.69%
	Total	692,506		Total	628,302	
School size	Small	74,786	10.80%	Small	76,843	12.23%
	Medium	229,360	33.12%	Medium	190,428	30.31%
	Large	149,366	21.57%	Large	132,857	21.15%
	Extra-large	238,994	34.51%	Extra-large	228,174	36.32%
	Total	692,506		Total	628,302	
Urban	Rural	494,080	71.35%	Rural	391,663	62.34%
	Urban	198,426	28.65%	Urban	236,639	37.66%
	Total	692,506		Total	628,302	
Region	BKK and	123,161	17.78%	BKK &	113,169	18.01%
	perimeter		6.45%	perimeter		6.46%
	Central	44,684	8.94%	Central	40,562	8.40%
	Upper	61,934	8.20%	Upper	52,786	7.93%
	north		16.48%	north		15.98%
	Lower	56,770	16.79%	Lower	49,831	16.39%
	north		13.69%	north		14.7%
	Upper NE	114,127	7.71%	Upper NE	100,390	8%
	Lower NE	116,259	3.96%	Lower NE	102,969	4.13%
	South	94,773		South	92,371	
	East	53,359		East	50,244	

Variables	**Proportion 2012**		**Percent 2012**	**Proportion 2015**		**Percent 2015**
	West	27,439		West	25,980	
	Total	692,506		Total	628,302	
Gender	Female	373,525	53.94%	Female	339,656	54.06%
	Male	318,981	46.06%	Male	288,646	45.94%
	Total	692,506		Total	628,302	

Source: Own calculation using O-NET microdata

Even though the exams are based on the curriculum, the actual subjects the students are tested on has changed over time as shown in Table C1-4. Starting from 2009, all students took exams in eight core subjects, including 1) Thai Language, 2) Social Science, Religion, and Culture, 3) Foreign Languages (English), 4) Mathematics, 5) Science, 6) Health and Physical Education, 7) Art, and 8) Career and Technology (NIETS, 2017). Then from 2015, the subjects tested were reduced to five – Thai, Social Studies, English, Mathematics, and Science. In 2017, Social Studies was excluded from grade nine tests as NIETS deemed the subject to be heavily contextualised and should be assessed by the schools themselves. Therefore, the core subjects that are tested every year include Thai, Mathematics, and Science.

Table C1- 5 Subjects included in O-NET Exams from 2007-2017

Grade	Year	Thai	Social Studies	English	Mathematics	Science	Other subjects[159]
Grade 6	2007	/			/	/	
	2008	/			/	/	
	2009	/	/	/	/	/	/
	2010	/	/	/	/	/	/
	2011	/	/	/	/	/	/
	2012	/	/	/	/	/	/
	2013	/	/	/	/	/	/
	2014	/	/	/	/	/	/
Grade 9	2010	/	/	/	/	/	/
	2011	/	/	/	/	/	/
	2012	/	/	/	/	/	/
	2013	/	/	/	/	/	/
	2014	/	/	/	/	/	/
	2015	/	/	/	/	/	
	2016	/	/	/	/	/	
	2017	/		/	/	/	
Grade 12	2013	/	/	/	/	/	/
	2014	/	/	/	/	/	/
	2015	/	/	/	/	/	
	2016	/	/	/	/	/	
	2017	/	/	/	/	/	

Source: Own calculation using O-NET microdata

[159] Health and physical education, Arts, and Career

Appendix C2: PISA data details

Table C2- 1 PISA school sampling frame

School type	2012			2015		
	School population	Sampled school	%	School population	Sampled school	%
OBEC schools with LS only	6,993	37	1%	6,979	42	1%
OBEC schools with both US and LS levels	2,417	88	4%	2,433	92	4%
OBEC schools with US only	10	2	20%	8	2	25%
Private schools with LS only	619	3	0.5%	539	4	1%
Private schools with US and LS levels, and vocational degree	501	13	3%	547	13	2%
Private schools with US and vocational degree	385	13	3%	408	10	2%
BMA schools with LS only	94	14	15%	98	23	23%
BMA schools with both US and LS levels	8	4	50%	9	7	78%
Local schools with LS only	297	4	1%	288	6	2%

School type	2012			2015		
	School population	**Sampled school**	**%**	**School population**	**Sampled school**	**%**
Local schools with US and LS levels, and vocational degree	314	12	4%	331	18	5%
Local schools with US and vocational degree	10	2	20%	11	2	18%
Satit schools with LS only	1	1	100%	1	1	100%
Satit schools with both US and LS levels	25	15	60%	26	21	81%
Vocational schools	411	19	5%	418	19	5%
Science schools	13	12	92%	13	13	100%
Total	**12,098**	**239**		**12,109**	**273**	

LS is lower secondary, US is upper secondary
Source: OECD and IPST (2018)

Table C2- 2 PISA's student sampling frame

School type	2012			2015		
	Student population	**Sampled student**	**%**	**Student population**	**Sampled student**	**%**
OBEC schools with LS only	53,324	256	0.5%	43,646	244	1%
OBEC schools with both US and LS levels	440,249	3,072	1%	441,320	3,740	1%
OBEC schools with US only	4,495	70	2%	4,172	84	2%
Private schools with LS only	10,599	39	0.4%	7,546	57	1%
Private schools with US and LS levels, and vocational degree	64,795	430	1%	63,051	522	1%
Private schools with US and vocational degree	62,638	431	1%	48,324	348	1%
BMA schools with LS only	2,453	283	12%	2,155	544	25%
BMA schools with both US and LS levels	997	140	14%	1,180	294	25%
Local schools with LS only	6,509	97	1%	5,233	128	2%
Local schools with US and LS levels, and vocational degree	28,616	420	1%	30,210	714	2%

School type	**2012**			**2015**		
	Student population	**Sampled student**	**%**	**Student population**	**Sampled student**	**%**
Local schools with US and vocational degree	959	66	7%	768	84	11%
Satit schools with LS only	2	4	200%[160]	2	2	100%
Satit schools with both US and LS levels	4,246	525	12%	4,471	837	19%
Vocational schools	93,813	665	1%	93,198	793	1%
Science schools	2,079	420	20%	1,995	546	27%
Total	**775,774**	**6,918**		**747,271**	**8,937**	

LS is lower secondary, US is upper secondary
Source: OECD and IPST (2018)

[160] This figure is taken from the official IPST report. The population number is probably incorrect as microdata shows four students in one Satit school matching the description.

Table C2- 3 Number of observations before and after data cleaning and restrictions

Year	Initial sample	Excluding vocational schools	Excluding missing data (for regression)	Percentage dropped
2006	6,192	5,554	-	-
2009	6,225	5,160	-	-
2012	6,606	5,646	5,382	5%
2015	8,249	7,268	6,783	7%
2018	8,633	7,299	-	-

Source: Own calculation using PISA microdata

Table C2- 4 Summary statistics of PISA continuous variables

Variables	Obs.	Mean	Std. Dev.	Min	Max
2012					
Mathematics score	5,382	437.77	82.92	173.68	783.12
Science score	5,382	454.50	75.95	200.35	747.61
2015					
Mathematics score	6,783	424.23	81.41	122.66	765.44
Science score	6,783	430.57	78.50	172.67	724.52

Source: Own calculation using PISA microdata

Table C2- 5 Summary statistics of PISA categorical variables

Variables	Proportion 2012		Proportion 2015	
School type	OBEC1	7.98%	OBEC1	N/A[161]
	OBEC2	72.86%	OBEC2	N/A
	Private	11.55%	Private	N/A
	BMA	0.48%	BMA	N/A

[161] Identifier not available from the data

Variables	Proportion 2012		Proportion 2015	
	Local	6.02%	Local	N/A
	Satit	0.75%	Satit	N/A
	Science	0.36%	Science	N/A
School size	Small	16.36%	Small	20.03%
	Medium	29.57%	Medium	25.56%
	Large	23.30%	Large	20.08%
	Extra-large	30.77%	Extra-large	34.33%
Urban	Rural	68.67%	Rural	79.30%
	Urban	31.32%	Urban	20.70%
Gender	Female	60.80%	Female	59.71%
	Male	39.20%	Male	40.29%
Parental education	Some primary or lower	33.15%	Some primary or lower	24.46%
	Lower secondary	15.66%	Lower secondary	19.14%
	Upper secondary or above	51.19%	Upper secondary or above	56.40%
Books at home	0-10 books	21.25%	0-10 books	19.37%
	11-25 books	34.10%	11-25 books	33.40%
	26-100 books	30.54%	26-100 books	31.29%
	101-200 books	8.28%	101-200 books	9.36%
	More than 200 books	5.83%	More than 200 books	6.58%

Source: Own calculation using PISA microdata

Appendix C3: TIMSS data details

Unlike PISA, TIMSS does not reveal a sampling frame. The only data available is the number of students and schools sampled.

Table C3- 1 TIMSS's sampled schools and students

School types	2011		2015	
	Sampled school	**Sampled student**	**Sampled school**	**Sampled student**
OBEC1 schools	82	3,172	42	907
OBEC2 schools	38	1,225	98	3,446
Private schools	20	572	22	751
BMA schools	10	349	10	311
Local schools	12	384	12	454
Science schools	0	0	10	236
Satit schools	10	422	10	377
Total	**172**	**6,124**	**204**	**6,482**

Source: Own calculation using TIMSS microdata

Table C3- 2 Number of observations before and after data cleaning and restrictions

Year	Initial sample	Excluding cases with any missing data	Excluding cases with missing data, but impute parental education (for regression)	Percentage dropped
2007	5,412	-	-	-
2011	6,124	4,152	5,967	3%
2015	6,482	4,368	6,465	0%

Source: Own calculation using TIMSS microdata

Table C3- 3 Summary statistics of TIMSS continuous variables

Variables	Obs.	Mean	Std. Dev.	Min	Max
2011					
Mathematics score	6,124	427.11	90.95	144.13	764.30
Science score	6,124	450.89	83.74	203.17	753.52
2015					
Mathematics score	6,482	431.42	98.83	172.35	842.44
Science score	6,482	455.84	89.13	211.69	825.19

Source: Own calculation using TIMSS microdata

Table C3- 4 Summary statistics of TIMSS categorical variables

Variables	Proportion 2011		Proportion 2015	
School size	Small	16.23%	Small	N/A
	Medium	31.87%	Medium	N/A
	Large	26.14%	Large	N/A
	Extra-large	25.75%	Extra-large	N/A
Urban	Rural	78.3%	Rural	76.1%
	Urban	21.1%	Urban	23.9%

Variables	Proportion 2011		Proportion 2015	
Gender	Female	54.8%	Female	53.7%
	Male	45.2%	Male	46.3%
Mother's education	Some primary or lower	27.3%	Some primary or lower	10.3%
	Lower secondary	25.4%	Lower secondary	34.3%
	Upper secondary or above	24.9%	Upper secondary or above	31.4%
	Don't know	22.5%	Don't know	24.1%
Father's education	Some primary or lower	21.6%	Some primary or lower	8.3%
	Lower secondary	23.6%	Lower secondary	30.9%
	Upper secondary or above	28.7%	Upper secondary or above	32.6%
	Don't know	26.2%	Don't know	28.2%
Books at home	0-10 books	28.1%	0-10 books	25.2%
	11-25 books	45.2%	11-25 books	48.1%
	26-100 books	19.7%	26-100 books	19.4%
	101-200 books	4.5%	101-200 books	4.3%
	More than 200 books	2.5%	More than 200 books	2.9%

Source: Own calculation using TIMSS microdata

Appendix C4: Variable details

Table C4- 1 Summary of questions asked in the independent variables

Variable	Questions	Possible values	Availability
Student-level variables			
Gender	TIMSS: Are you a girl or a boy?	- Girl (Female) - Boy (Male)	PISA, TIMSS, and O-NET
Prior score	N/A	From 0 – 100 (percent correct)	O-NET
Parental education	PISA: What is the <highest level of schooling> completed by your father/mother?	- <ISCED level 3A> - <ISCED level 3B, 3C> - <ISCED level 2> - <ISCED level 1> - He/She did not complete <ISCED level 1>[162]	PISA and TIMSS
	TIMSS: What is the highest level of education completed by your father/mother (or stepfather/stepmother or male/female legal guardian)?”	- Less than high school - Some high school - High school graduate - Associate’s degree (2-year college program) - Bachelor’s degree (4-year college program) - Master’s degree or professional degree (MD, DDS, lawyer, minister) - Doctorate (Ph.D., or Ed.D.)	

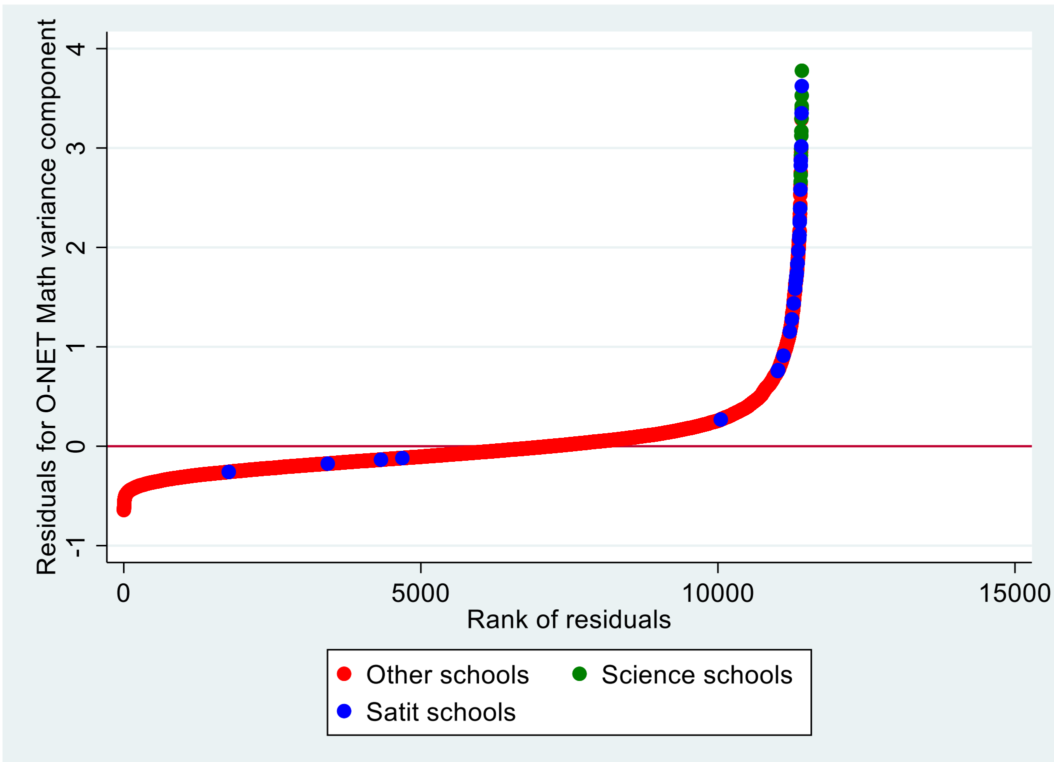
[162] ISCED level 3A is general track upper secondary, while 3B is vocational track upper secondary and 3C is an upper secondary designed for learners to go straight to labour market after completion. ISCED level 2 is lower secondary and ISCED level 1 is primary level of education.

Variable	Questions	Possible values	Availability
		- I don't know	
Books at home	PISA: How many books are there in your home? There are usually about 40 books per metre of shelving. Do not include magazines, newspapers, or your schoolbooks.	- 0-10 books - 11-25 books - 26-100 books - 101-200 books - 201-500 books - More than 500 books	PISA and TIMSS
	TIMSS: About how many books are there in your home? (Do not count magazines, newspapers, or your school books.)	- None or very few (0–10 books) - Enough to fill one shelf (11–25 books) - Enough to fill one bookcase (26–100 books) - Enough to fill two bookcases (101–200 books) - Enough to fill three or more bookcases (more than 200)	
School-level variables			
School type	N/A	- OBEC1 - OBEC2 - Private - BMA - Local - Satit	PISA, TIMSS, and O-NET

Variable	**Questions**	**Possible values**	**Availability**
		- Science	
School size	<u>PISA</u>: As at <February 1, 2015 (2012)>, what was the total school enrolment (number of students)?	- Values greater than 0	PISA, TIMSS, and O-NET
	<u>TIMSS</u>: What is the total enrolment of students in your school as of March 1, 2015 (April 1, 2011)?	- Values greater than 0	
Urban area	<u>PISA</u>: Which of the following definitions best describes the community in which your school is located?	- A village, hamlet or rural area (fewer than 3,000 people) - A small town (3,000 to about 15,000 people) - A town (15,000 to about 100,000 people) - A city (100,000 to about 1,000,000 people) - A large city (with over 1,000,000 people)	PISA, TIMSS, and O-NET
	<u>TIMSS</u>: Which best describes the immediate area in which your school is located?	- Urban–Densely populated - Suburban–On fringe or outskirts of urban area - Medium size city or large town - Small town or village - Remote rural	

Appendix C5: Residual plot

Figure C5- 1 Estimates of school residuals by rank based on a null model using Mathematics score 2015



Source: Own calculation using O-NET 2015 microdata

Appendix C6: Interview questions

The sample questions for the teachers are as follows.

- **Personal information**
    - What subject do you teach?
    - How long have you been teaching?
    - How many periods do you teach per week?
    - What other responsibilities do you have besides teaching?
    - When did the school take PISA/TIMSS exams?
- **Awareness of the assessments**
    - When was your first experience of either PISA or TIMSS; and the most recent?
    - Can you describe what PISA, TIMSS, and O-NET measure?
    - What skills or knowledge should students have to perform well on these tests? Has this changed over time? Especially between your first and most recent experience of PISA or TIMSS?
    - What are the pros and cons of these assessments?
- **Pressure from the assessments**
    - How is the quality assurance system like at this school?
    - How are learning outcomes indicators include in quality assurance?
    - How do assessment results affect you? How have these changed over time? From when you first started teaching? Especially between your first and most recent experience of PISA or TIMSS?
    - What is the process of career progression and salary progression? What learning outcome indicators are involved? How important are these?
    - How do you prepare your students for the assessments? (Prompt for examples of activities) How long do you prepare in advance? Who decides how the preparation is done? Has this changed over time?

Especially between your first and most recent experience of PISA or TIMSS?

- Who sets learning goals/standards at the school? What are they? What assessments are they based on?
- Were you able to accomplish those set goals? Are there any consequences for you if the goals were not accomplished? What are factors that affect learning outcomes at this school?
- How have the assessment results been used at school? Formatively – to change teaching practices? Or summatively – to regrade students?
- What kind of skills or content knowledge do you think the students should have after finishing compulsory education?

- **Assessments and practices**
  - Can you describe formative and summative assessments of your subjects? Who designs these exams? What is the process of exam design? How has this changed over time? Especially between your first and most recent experience of PISA or TIMSS?
  - What knowledge or skills do you focus on in your assessment?
  - Can you give an example of how you would teach a new topic to students? Has your teaching methods changed over time? Especially between your first and most recent experience of PISA or TIMSS?
  - Do you have enough time to cover all topics in a given school year? If not, how do you prioritise the contents to teach students? How do the prioritised contents change over time? Especially between your first and most recent experience of PISA or TIMSS?
  - How have your teaching and assessment practice changed over time?
  - What are other changes that happened that affect your teaching and assessment?

  - What are some of the challenges in improving learning outcomes? (Probe for student background, school resources, parental expectations, students' aspirations)

For vice principals/principals/academic teachers, the sample questions are as follows.

- **Introduction**
  - How long have you been a vice principal/principal/academic teachers at this school?
  - Do you have teaching duties? If so, what subject do you teach and how many periods per week do you teach?
  - When did the school take PISA/TIMSS exams?
- **Awareness of the assessments**
  - When was your first experience of either PISA or TIMSS; and the most recent?
  - Can you describe what PISA, TIMSS, and O-NET measure?
  - What skills or knowledge should students have to perform well on these tests? Has this changed over time? Especially between your first and most recent experience of PISA or TIMSS?
  - What are the pros and cons of these assessments?
- **Pressure from the assessments**
  - How is the quality assurance system like at this school?
  - How are learning outcomes indicators include in quality assurance?
  - How do assessment results affect you? How have these changed over time? From when you first started teaching? Between your first and most recent experience of PISA or TIMSS?
  - What is the process of career progression and salary progression? What learning outcome indicators are involved? How important are these?
  - Who sets learning goals/standards at the school? What are they? What assessments are they based on?

  - Were you able to accomplish those set goals? Are there any consequences for you if the goals were not accomplished? What are factors that affect learning outcomes at this school?
  - How have the assessment results been used at school? Formatively – to change teaching practices? Or summatively – to regrade students?
  - What kind of skills or content knowledge do you think the students should have after finishing compulsory education?

- **Assessments and practices**
  - What are the schools' values, goals, and policies on learning outcomes? How are these developed? How are they measured?
  - Can you give examples of some policies you have implemented to improve student learning? Who sets those policies? How successful are those policies?
    - Do you separate the students into classes based on ability? Or on results in any of the three assessments?
  - How has the school's academic focus changed over time? and especially between your first and most recent experience of PISA or TIMSS?
  - What kind of feedback do you receive from participating in the assessments? In what ways do you use this data?
  - What are some of the challenges in improving learning outcomes? (Probe for student background, school resources, parental expectations, students' aspirations)

Appendix D: Appendix to Chapter Five

Appendix D1: O-NET estimation results

Table D1- 1 Multilevel model results for O-NET Mathematics 2015

	(1)	**(2)**	**(3)**	**(4)**
Fixed Part				
Constant	0.06***	0.11***	0.46***	0.35***
	(0.00)	(0.01)	(0.01)	(0.01)
School type (Reference: OBEC2)				
OBEC1		-0.11***	0.09***	0.15***
		(0.01)	(0.01)	(0.01)
Private		0.06***	0.14***	0.12***
		(0.01)	(0.01)	(0.01)
BMA		-0.14***	0.00	-0.04
		(0.04)	(0.03)	(0.03)
Local		-0.15***	-0.02	0.01
		(0.02)	(0.01)	(0.01)
Satit		**1.72*** **	**0.85*** **	**0.80*** **
		(0.07)	(0.06)	(0.06)
Science		**3.23*** **	**1.63*** **	**1.67*** **
		(0.11)	(0.09)	(0.09)
Gender-female			-0.08***	0.08***
			(0.00)	(0.00)
Prior scores				
Grade 6 Math			**0.44*** **	**0.44*** **
			(0.00)	(0.00)
Grade 6 Science			**0.22*** **	**0.22*** **
			(0.00)	(0.00)
Grade 6 Thai			0.13***	0.13***

	(1)	**(2)**	**(3)**	**(4)**
			(0.00)	(0.00)
School size (Reference: S)				
M				0.00
				(0.01)
L				0.03***
				(0.01)
XL				**0.23*** **
				(0.01)
Urban area				0.03***
				(0.01)
Region (Reference: Lower NE)				
Bangkok and its perimeter				0.08***
				(0.01)
Central				0.05***
				(0.01)
Upper North				0.12***
				(0.01)
Lower North				0.06***
				(0.01)
Upper NE				0.05***
				(0.01)
South				0.06***
				(0.01)
East				0.07***
				(0.01)
West				0.10***
				(0.02)
ICC	16.23%	13.73%	12.75%	12.00%

	(1)	(2)	(3)	(4)
Model fit				
Log-likelihood	-853412	-852600	-743555	-743315
AIC	1706826	1705214	1487131	1486677
BIC	1706838	1705293	1487256	1486938
Observations	628,302	628,302	628,302	628,302
No. of groups	11,413	11,413	11,413	11,413

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using O-NET 2015 microdata

Table D1- 2 Multilevel model results for O-NET Science 2012

	(1)	(2)	(3)	(4)
Fixed Part				
Constant	-0.11***	-0.09***	0.13***	0.03**
	(0.00)	(0.01)	(0.01)	(0.01)
School type (Reference: OBEC2)				
OBEC1		-0.03***	0.11***	0.17***
		(0.01)	(0.01)	(0.01)
Private		0.02	0.11***	0.10***
		(0.01)	(0.01)	(0.01)
BMA		-0.13***	-0.05	-0.08**
		(0.04)	(0.04)	(0.04)
Local		-0.00	0.11***	0.15***
		(0.02)	(0.02)	(0.02)
Satit		**1.49*** **	**0.84*** **	**0.82*** **
		(0.08)	(0.08)	(0.07)
Science		**1.41*** **	**0.79*** **	**0.82*** **
		(0.12)	(0.11)	(0.11)
Gender-female			-0.02***	0.02***
			(0.00)	(0.00)

	(1)	**(2)**	**(3)**	**(4)**
Prior scores				
Grade 6 Math			**0.22*** **	**0.29*** **
			(0.00)	(0.00)
Grade 6 Science			**0.29*** **	**0.22*** **
			(0.00)	(0.00)
Grade 6 Thai			0.12***	0.12***
			(0.00)	(0.00)
School size (Reference: S)				
M				0.02**
				(0.01)
L				0.09***
				(0.02)
XL				**0.28*** **
				(0.02)
Urban area				-0.00
				(0.01)
Region (Reference: Lower NE)				
Bangkok and its perimeter				0.05***
				(0.02)
Central				-0.03
				(0.02)
Upper North				0.15***
				(0.01)
Lower North				0.07***
				(0.01)
Upper NE				0.01
				(0.01)
South				0.02*
				(0.01)
East				-0.04**

	(1)	**(2)**	**(3)**	**(4)**
				(0.02)
West				-0.00
				(0.02)
ICC	19.89%	18.90%	22.89%	22.06%
Model fit				
Log-likelihood	-853048	-852778	-754623	-754412
AIC	1706098	1705570	1509269	1508870
BIC	1706109	1705650	1509395	1509133
Observations	692,506	692,506	692,506	692,506
No. of groups	11,383	11,383	11,413	11,383

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using O-NET 2012 microdata

Table D1- 3 Multilevel model results for O-NET Science 2015

	(1)	**(2)**	**(3)**	**(4)**
Fixed Part				
Constant	0.03***	0.08***	0.42***	0.36***
	(0.00)	(0.01)	(0.01)	(0.01)
School type (Reference: OBEC2)				
OBEC1		-0.07***	0.15***	0.18***
		(0.01)	(0.01)	(0.01)
Private		-0.02	0.07***	0.07***
		(0.01)	(0.01)	(0.01)
BMA		-0.13***	0.02	0.04
		(0.04)	(0.03)	(0.03)
Local		**-0.21*** **	-0.06***	-0.04**
		(0.02)	(0.02)	(0.02)

	(1)	**(2)**	**(3)**	**(4)**
Satit		**1.45*** **	**0.50*** **	**0.49*** **
		(0.074)	(0.06)	(0.06)
Science		**2.66*** **	**0.93*** **	**0.96*** **
		(0.11)	(0.09)	(0.09)
Gender-female			-0.071***	0.07***
			(0.00)	(0.00)
Prior scores				
Grade 6 Math			**0.36*** **	**0.33*** **
			(0.00)	(0.00)
Grade 6 Science			**0.33*** **	**0.36*** **
			(0.00)	(0.00)
Grade 6 Thai			0.20***	0.20***
			(0.00)	(0.00)
School size (Reference: S)				
M				-0.00
				(0.01)
L				0.02
				(0.01)
XL				0.17***
				(0.02)
Urban area				0.00
				(0.00)
Region (Reference: Lower NE)				
Bangkok and its perimeter				0.01
				(0.02)
Central				-0.00
				(0.01)
Upper North				0.13***
				(0.01)
Lower North				0.01

	(1)	**(2)**	**(3)**	**(4)**
				(0.01)
Upper NE				0.02**
				(0.01)
South				0.05***
				(0.01)
East				0.03*
				(0.01)
West				0.06***
				(0.02)
ICC	16.07%	14.32%	14.92%	14.48%
Model fit				
Log-likelihood	-856877	-856340	-719626	-719497
AIC	1713756	1712693	1439274	1439039
BIC	1713767	1712773	1439399	1439300
Observations	628,302	628,302	628,302	628,302
No. of groups	11,413	11,413	11,413	11,413

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using O-NET 2015 microdata

Appendix D2: PISA estimation results

Table D2- 1 OLS results for PISA Mathematics

	2012		2015	
	(1)	**(2)**	**(1)**	**(2)**
Constant	402.29***	401.91***	399.58***	392.98***
	(5.88)	(10.49)	(5.31)	(8.05)
Gender-female	4.21	-0.13	-1.83	-4.21
	(3.42)	(3.02)	(3.56)	(3.22)
Parental education (Reference: Primary or lower)				
Lower secondary	8.07*	4.12	-8.25*	-8.86**
	(4.84)	(4.50)	(4.57)	(4.45)
Upper secondary or above	**25.74*** **	11.40***	14.71***	4.90
	(5.03)	(4.27)	(5.18)	(4.46)
Books at home (Reference: 0-10 books)				
11-25 books	9.67**	8.08**	8.86**	6.04
	(4.10)	(3.94)	(3.91)	(3.99)
26-100 books	**25.06*** **	**20.72*** **	**28.07*** **	**22.94*** **
	(4.57)	(4.12)	(4.91)	(4.69)
101-200 books	**39.78*** **	**33.53*** **	**31.40*** **	**25.39*** **
	(7.46)	(7.10)	(6.22)	(5.75)
More than 200 books	**72.60*** **	**58.42*** **	**66.04*** **	**55.90*** **
	(7.34)	(7.00)	(9.69)	(8.91)
Urban area		**17.30*** **		9.44
		(6.65)		(8.57)
School size (Reference: S)				
M		-11.51		-.14
		(11.36)		(9.59)

	2012		2015	
	(1)	**(2)**	**(1)**	**(2)**
L		4.22		7.18
		(12.76)		(9.25)
XL		**36.36*****		**41.07*****
		(11.08)		(9.42)
R-squared	0.085	0.159	0.066	0.124
Observations	5,382	5,382	6,783	6,783

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using PISA microdata

Table D2- 2 OLS results for PISA Science

	2012		2015	
	(1)	**(2)**	**(1)**	**(2)**
Constant	418.62***	416.20***	396.48***	383.35***
	(5.40)	(9.66)	(4.28)	(5.72)
Gender-female	10.87***	6.65**	4.25**	0.93
	(3.43)	(3.11)	(3.01)	(2.53)
Parental education (Reference: Primary or lower)				
Lower secondary	5.62	2.09	-7.91**	-8.73***
	(4.42)	(4.15)	(3.38)	(3.25)
Upper secondary or above	**22.18*****	9.09**	**19.52*****	6.21**
	(4.32)	(3.78)	(4.09)	(3.33)
Books at home (Reference: 0-10 books)				
11-25 books	10.85***	9.376***	13.20***	9.48***
	(3.63)	(3.60)	(3.28)	(3.31)
26-100 books	**21.76*****	**17.79*****	**33.59*****	**26.69*****
	(4.36)	(4.06)	(3.87)	(3.53)

	2012		2015	
	(1)	**(2)**	**(1)**	**(2)**
101-200 books	**41.73***	**35.95***	**36.74***	**29.16***
	(6.53)	(6.37)	(5.69)	(5.12)
More than 200 books	**55.46***	**42.56***	**56.14***	**43.37***
	(6.28)	(6.00)	(8.50)	(6.98)
Urban area		15.05**		15.19**
		(6.56)		(7.50)
School size (Reference: S)				
M		-6.94		5.15
		(9.74)		(7.75)
L		6.49		**24.14***
		(10.53)		(7.36)
XL		**35.79***		**53.66***
		(9.96)		(7.49)
R-squared	0.077	0.148	0.075	0.171
Observations	5,382	5,382	6,783	6,783

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using PISA microdata

Appendix D3: TIMSS OLS results without missing data imputation

Table D3- 1 OLS results for TIMSS Mathematics without imputation

	2011			2015	
	(1)	**(2)**	**(3)**	**(1)**	**(2)**
Constant	401.88***	400.59***	390.56***	409.67***	405.18***
	(7.21)	(7.38)	(1.94)	(7.75)	(7.21)
Gender-female	16.12***	15.08***	15.36***	9.92*	9.48**
	(4.47)	(4.37)	(.24)	(5.20)	(4.53)
Parental education (Reference: Primary or lower)					
Lower secondary	-16.66***	-16.72***	-17.81***	**-23.88*****	**-23.98*****
	(4.77)	(4.79)	(1.43)	(6.62)	(5.87)
Upper secondary or above	**21.15*****	17.66**	8.19***	15.66**	6.89
	(6.60)	(6.74)	(1.12)	(7.38)	(6.23)
Books at home (Reference: 0-10 books)					
11-25 books	13.60***	12.98***	9.38**	17.44***	15.53***
	(4.68)	(4.63)	(2.91)	(4.29)	(4.21)
26-100 books	**40.00*****	**38.45*****	**31.85*****	**47.79*****	**40.12*****
	(6.95)	(6.93)	(3.51)	(6.12)	(6.03)

	2011			2015	
	(1)	**(2)**	**(3)**	**(1)**	**(2)**
101-200 books	**73.83*** **	**71.27*** **	**59.61*** **	**90.79*** **	**77.70*** **
	(11.45)	(11.19)	(3.90)	(10.14)	(9.76)
More than 200 books	**87.88*** **	**85.59*** **	**70.75*** **	**106.45*** **	**89.05*** **
	(15.42)	(14.32)	(3.88)	(16.91)	(13.26)
Urban area		18.62	5.63***		**50.93*** **
		(12.94)	(1.10)		(10.59)
School size (Reference: S)					
M			6.41**		
			(2.17)		
L			**29.63*** **		
			(1.59)		
XL			**49.71*** **		
			(1.56)		
R-squared	0.135	0.142	0.181	0.158	0.212
Observations	4,152	4,152	4,152	4,368	4,368

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using TIMSS microdata

Table D3- 2 OLS results for TIMSS Science without imputation

	2011			2015	
	(1)	**(2)**	**(3)**	**(1)**	**(2)**
Constant	430.07***	428.92***	419.36***	433.16***	429.32***
	(5.72)	(5.84)	(0.54)	(8.03)	(8.02)
Gender-female	12.05**	11.11**	11.24***	12.46**	12.08***
	(4.67)	(4.67)	(2.07)	(4.74)	(4.26)
Parental education (Reference: Primary or lower)					
Lower secondary	**-19.90*** **	**-19.96*** **	**-21.00*** **	**-22.04*** **	**-22.13*** **
	(4.31)	(4.30)	(1.72)	(6.19)	(5.95)
Upper secondary or above	16.72***	13.60**	5.23**	13.49*	5.99
	(5.53)	(5.47)	(1.57)	(7.10)	(6.69)
Books at home (Reference: 0-10 books)					
11-25 books	18.04***	17.49***	14.16***	17.66***	16.03***
	(3.12)	(3.07)	(1.49)	(4.04)	(4.07)
26-100 books	**38.58*** **	**37.20*** **	**31.26*** **	**46.66*** **	**40.11*** **
	(5.31)	(5.30)	(0.87)	(5.92)	(5.89)
101-200 books	**59.40*** **	**57.10*** **	**46.83*** **	**79.83*** **	**68.65*** **
	(9.36)	(9.06)	(3.57)	(8.03)	(8.04)

	2011			**2015**	
	(1)	**(2)**	**(3)**	**(1)**	**(2)**
More than 200 books	**82.98*** **	**80.93*** **	**68.21*** **	**98.44*** **	**83.57*** **
	(13.50)	(12.69)	(2.43)	(12.16)	(9.80)
Urban area		16.65	5.83**		**43.54*** **
		(11.80)	(1.72)		(8.65)
School size (Reference: S)					
M			6.88*		
			(2.70)		
L			**28.58*** **		
			(1.33)		
XL			**43.03*** **		
			(0.81)		
R-squared	0.135	0.142	0.181	0.158	0.212
Observations	4,152	4,152	4,152	4,368	4,368

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using TIMSS microdata

Appendix D4: OLS VS Multilevel model results

Using the command PV in Stata, it is possible to produce multilevel model estimates that take into account student weights and plausible values. Replication weights were not specified. Hence, the results are likely to have underestimated standard errors. Additionally, there are no model fit measures and the results should be interpreted with caution.

The estimations focus on Mathematics and Science data in 2012 (2011 in TIMSS) as this is the year with more complete variables (school types data are available in PISA and school size is available in TIMSS). Results show similarities in the results between OLS and MLM. The coefficients are generally similar in terms of significance and sign. The magnitude of each coefficient varies by less than 10 points, which is around 0.1 standard deviation. In PISA, the effect of school type is slightly higher in MLM comparing to OLS and the effect of socioeconomic status is lessor in magnitude. In TIMSS, however, the results are more different, with the effects of socioeconomic measures being significantly stronger in OLS than in TIMSS.

Table D4- 1 OLS and Multilevel results for PISA Mathematics 2012

	(1) **OLS**	**(2)** **MLM**
Constant	378.93***	398.54***
	(10.97)	(10.33)
School type (Reference: OBEC2)		
OBEC1	**62.05*** **	**54.06*** **
	(22.84)	(12.06)
Private	-12.64	-15.33
	(8.66)	(12.00)
BMA	**-29.96** **	**-34.57*** **
	(13.65)	(12.17)
Local	-11.00	-14.16
	(7.21)	(11.65)
Satit	**76.31*** **	**86.27*** **
	(10.03)	(12.62)
Science	**142.89*** **	**147.30*** **
	(8.02)	(14.48)
Gender-female	0.61	-3.67*
	(3.13)	(1.91)
Parental education (Reference: Primary or lower)		
Lower secondary	5.86	5.84
	(3.95)	(3.53)
Upper secondary or above	13.24***	7.43***
	(3.62)	(2.55)
Books at home (Reference: 0-10 books)		
11-25 books	8.61**	3.77
	(3.88)	(3.13)
26-100 books	**20.44*** **	14.02***
	(3.95)	(2.69)
101-200 books	**31.99*** **	**16.97*** **

	(1)	**(2)**
	OLS	**MLM**
	(7.20)	(4.07)
More than 200 books	**55.49*** **	**35.88*** **
	(7.20)	(4.08)
Urban area	**17.81*** **	**18.09** **
	(6.66)	(7.97)
School size (Reference: S)		
M	6.92	1.17
	(12.45)	(10.64)
L	**27.23** **	19.97
	(12.40)	(12.22)
XL	**58.38*** **	**48.93*** **
	(11.74)	(12.60)
R-squared	0.213	N/A
Observations	5,382	5,382

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using PISA 2012 microdata

Table D4- 2 OLS and Multilevel results for PISA Science 2012

	(1)	**(2)**
	OLS	**MLM**
Constant	393.21***	408.42***
	(8.62)	(8.98)
School type (Reference: OBEC2)		
OBEC1	**60.97*** **	**54.45*** **
	(13.70)	(10.39)
Private	-11.03	-14.61
	(7.70)	(9.75)
BMA	**-32.14*** **	**-36.12*** **
	(12.11)	(9.83)
Local	-8.55	-10.19
	(6.33)	(9.44)
Satit	**63.10*** **	**71.18*** **
	(8.04)	(10.33)
Science	**120.70*** **	**122.25*** **
	(6.53)	(11.93)
Gender-female	7.41**	3.84
	(3.17)	(2.25)
Parental education (Reference: Primary or lower)		
Lower secondary	3.74	2.42
	(3.70)	(3.11)
Upper secondary or above	11.10***	5.27**
	(3.37)	(2.48)
Books at home (Reference: 0-10 books)		
11-25 books	9.96***	6.046**
	(3.56)	(3.03)
26-100 books	**17.72*** **	13.79***
	(3.91)	(2.80)

	(1)	(2)
	OLS	**MLM**
101-200 books	**34.80*** **	**24.94*** **
	(6.43)	(3.85)
More than 200 books	**40.30*** **	**28.73*** **
	(6.07)	(4.04)
Urban area	15.67**	14.99**
	(6.48)	(6.44)
School size (Reference: S)		
M	11.30	8.86
	(8.16)	(8.82)
L	**28.96*** **	**23.48** **
	(8.91)	(9.94)
XL	**57.31*** **	**50.34*** **
	(8.46)	(10.12)
R-squared	0.204	N/A
Observations	5,382	5,382

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using PISA 2012 microdata

Table D4- 3 OLS and Multilevel results for TIMSS Mathematics

	2011		2015	
	OLS	**MLM**	**OLS**	**MLM**
Constant	400.59***	422.16***	405.18***	437.54***
	(7.38)	(6.23)	(7.21)	(7.68)
Gender-female	15.08***	-1.55	9.48**	-4.29
	(4.37)	(2.05)	(4.53)	(2.92)
Parental education (Reference: Primary or lower)				
Lower secondary	-16.66***	-9.58***	**-23.98*** **	-12.88**
	(4.77)	(2.96)	(5.87)	(4.88)
Upper secondary or above	**21.15*** **	0.95	6.89	-3.42
	(6.60)	(2.94)	(6.23)	(4.46)
Books at home (Reference: 0-10 books)				
11-25 books	12.98***	4.17	15.53***	3.38
	(4.63)	(3.93)	(4.21)	(2.78)
26-100 books	**38.45*** **	13.63***	**40.12*** **	12.10***
	(6.93)	(4.41)	(6.03)	(3.15)
101-200 books	**71.27*** **	**23.08*** **	**77.70*** **	**20.52*** **
	(11.19)	(5.09)	(9.76)	(5.00)
More than 200 books	**85.59*** **	**23.72*** **	**89.05*** **	**20.85*** **
	(14.32)	(6.71)	(13.26)	(5.28)
Urban area	18.62	**40.27*** **	**50.93*** **	**52.52*** **
	(12.94)	11.48	(10.59)	(12.46)
R-squared	0.14	N/A	0.21	N/A
Observations	4,152	4,152	4,368	4,368

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using TIMSS microdata

Table D4- 4 OLS and Multilevel results for TIMSS Science

	2011		2015	
	OLS	**MLM**	**OLS**	**MLM**
Constant	428.92***	448.22***	429.32***	456.71***
	(5.84)	(5.63)	(8.02)	(6.70)
Gender-female	11.11**	-1.67	12.08***	-0.96
	(4.67)	(2.53)	(4.26)	(2.62)
Parental education (Reference: Primary or lower)				
Lower secondary	**-19.96*** **	-14.33***	**-22.13*** **	-14.02***
	(4.30)	(3.32)	(5.95)	(3.82)
Upper secondary or above	13.60**	-1.74	5.99	-3.75
	(5.47)	(3.18)	(6.69)	(4.06)
Books at home (Reference: 0-10 books)				
11-25 books	17.49***	8.79***	16.03***	5.55*
	(3.07)	(2.79)	(4.07)	(2.88)
26-100 books	**37.20*** **	13.93***	**40.11*** **	17.71***
	(5.30)	(2.99)	(5.89)	(4.07)
101-200 books	**57.10*** **	17.78***	**68.65*** **	**22.70*** **
	(9.06)	(5.27)	(8.04)	(4.48)
More than 200 books	**80.93*** **	**28.55*** **	**83.57*** **	**29.01*** **
	(12.69)	(5.25)	(9.80)	(4.79)
Urban area	16.65	**34.62*** **	**43.54*** **	**45.72*** **
	(11.80)	(10.09)	(8.65)	(10.55)
R-squared	0.134	N/A	0.216	N/A
Observations	4,152	4,152	4,368	4,368

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Own calculation using TIMSS microdata

Appendix D5: Qualitative findings regarding changes beyond 2015

This appendix presents additional qualitative findings that do not fit into the timeframe of the study. That is, they are the changes participants mentioned beyond 2015.

Changes in curriculum

One of the most important changes the participants mentioned is the adapted national curriculum of Science and Mathematics. All schools have been affected by the adjustment made to the National Curriculum in 2017 by IPST (OBEC & IPST, 2017). The curriculum of Mathematics and Science were adjusted so that it is competency-based, with skills such as critical thinking included (Sangbuaphuen, 2020). Note that this is not strictly the time period that this thesis focuses on and could not be used to explain why there are discrepancies during 2012-2015. Nonetheless, this aspect is worth discussing as it is one of the most important changes the participants mentioned. It also affects all schools.

It could be argued that the 2017 curriculum was born out of the direct influence of assessment outcomes of PISA and TIMSS. One of the justifications for the change is regarding low performance in national (National testing or NT and O-NET) and international tests (PISA and TIMSS). This shows the increase in interest put into learning outcomes that was not seen before in the previous curriculum of 2008. Other reasons for the revision are the global changes such as social and economic changes, changes in science and technology and innovation. With this, the curriculum needs to be adjusted to reflect these changes. 21st century skills were mentioned including critical thinking and problem-solving, communication, collaboration, and creativity and innovation (IPST & MOE, 2017). Knowing what had changed allow us to assess the direction Thailand is moving towards to solving the problems of low learning outcomes.

Comparing the two curricula, we can see the goals of the curriculum changed significantly as seen in Table F-1. Both curricula have overall goals students should achieve when completing compulsory education. The revised version has more specific goals of what students should be able to do with regards to Mathematics and Science. Notably, knowledge and skills are more emphasised and detailed in the new curriculum. In 2008, this aspect is described simply as "Knowledge and skills for communication, thinking, problem-solving, technological know-how, and life skills (MOE, 2008)". In 2017, there is more elaboration on what students should be able to do. There is also less focus on the aspect of citizenship. This affects how Mathematics and Science should be taught in classrooms.

Table F- 1 Goals of the 2008 and 2017 national curriculum

2008 curriculum goals	2017 curriculum goals
1. Morality, ethics, desirable values, self-esteem, self-discipline, observance of Buddhist teachings or those of one's faith, and guiding principles of Sufficiency Economy 2. Knowledge and skills for communication, thinking, problem-solving, technological know-how, and life skills 3. Good physical and mental health, hygiene, and preference for physical exercise 4. Patriotism, awareness of responsibilities and commitment as Thai citizens and members of the world community, and adherence to a democratic way of	**<u>Mathematics</u>** 1. Able to understand and apply necessary concepts, principals, and theories in Mathematics 2. Able to problem-solve, communicate and convey meaning in Mathematics, connect, reason, and have creativity 3. Have positive mindset regarding Mathematics, see value and importance of Mathematics, able to use knowledge in Mathematics as learning tools in higher levels of education as well as in occupation 4. Able to select appropriate media, tools, technology, and information sources as tools for learning,

2008 curriculum goals	2017 curriculum goals
life and form of government under constitutional monarchy 5. Awareness of the need to preserve all aspects of Thai culture and Thai wisdom, protection and conservation of the environment, and public-mindedness with dedication to public service for peaceful and harmonious co-existence	communication, work, and correct and effective problem-solving **<u>Science</u>** 1. Able to understand basic concepts, principals, and theories in Science 2. Able to understand the scope of the nature of Science and limitations of Science 3. Have important skills in researching information and develop technology 4. Realise the mutual effects among Science, technology, humanities, and environment 5. Able to apply concept knowledge and skills in Science and technology in a way that benefits the society and living 6. Develop thinking process and imagination, ability in problem-solving, management, communication skills, and ability in evaluating and making decisions 7. Have scientific mind, morality, ethics, and value in using science and technology creatively

Sources: Thai national curricula (MOE, 2008; OBEC & IPST, 2017)

Other than changes in the goals, the new curriculum also involved changes in contents. In both subjects, as seen in Table F-2, the learning units and core learning standards[163] were consolidated to be less than before (see full changes in Appendix A). In the 2008 curriculum, parts of the learning units are "Mathematical Skills and Processes" and "Nature of Science and Technology" in Mathematics and Science respectively. These describe the skills students should possess, such as problem-solving, creativity, scientific mind, etc. They were taken off the units and moved to separate skill sections.

Table F- 2 Comparison of number of learning units in the curriculum of 2008 and 2017

Learning units	Mathematics		Science	
	2008	2017	2008	2017
Units	6	3	8	4
Learning standards	13	7	13	10
Learning indicators	60	33	100	174

Source: Own calculation based on the curricula (MOE, 2008; OBEC & IPST, 2017)

Some teachers reacted positively towards a smaller number of indicators and units to teach. They perceive this as an improvement to the curriculum. The curriculum is more streamlined and repetitive contents were cut. This matches IPST's intentions in the curriculum revision (IPST & MOE, 2017).

> *From the 2008 curriculum to the 2017 revision, I see that there are adjustments of the indicators. There are a lot less indicators. … I think it's more concise than the 2007 curriculum, from 8 units for Science, now there are only 4, just the main units … Some repetitive contents were cut. For*

[163] The learning standards are core learning outcomes students should be able to demonstrate, while learning indicators are smaller units of learning outcomes. For example, in the unit "Numbers and Operations", a standard is "Understanding diverse methods of presenting numbers and their application in real life", and one of the indicators is "Explain and specify square roots and cube roots of real numbers".

> *example, … molecules, in grade 8, they study nutrients, we have to teach [molecules] again and they have to learn again. But for the 2017 curriculum, these contents on Chemistry were cut and put in Biology instead. They study this only in Biology (Science teacher, OBEC2 school (school 2)).*

> *The greatest change [in the curriculum] is the content. IPST cut out the contents that are not important. They cut out chapters. Say biology or physics already have those contents, chemistry doesn't have to repeat that again … It makes it easier for me to teach. Some unimportant contents they cut out. Sometimes they cut the whole chapter. (Science teacher, Private school (school 5)).*

While this suggests that the 2017 curriculum addressed the flaws in the 2008 curriculum, it seems that the new curriculum remains chunky. In Mathematics, a smaller number of learning is mostly a result of many indicators being combined into a single, lengthier indicator. Arguably, the contents students are required to learn are not necessarily less than before. The learning outcomes remain very similar to the ones in 2008. However, there is greater emphasis on the application of knowledge. Almost all indicators specify students need to be able to apply their knowledge in real life. For example, an indicator in the Measurement unit;

> *"Find the surface area of prisms and cylinders* (MOE, 2008, p. 81).*"*

was changed to

> *"Adapt the knowledge about surface area of prisms and cylinders to solve mathematical problems and real-world problems* (IPST & MOE, 2017, p. 23).*"*

Hence, students need to be able to both solve for surface areas, as well as apply the knowledge related to everyday life. In Science, by contrast, the number of indicators increased significantly. This is primarily because the addition of a new unit "Technology" and new learning indicator of awareness of environmental issues. It seems there has been more structural changes to how the learning units are structured in Science. The units were re-sequenced to help students learn from concrete to abstract topics. There is also an emphasis on scientific inquiry and scientific investigation, with a framework on the scientific inquiry process that starts participating in with asking scientific question, evidence gathering, and relating findings to existing evidence. With this, even though repetitive contents were cut, students are demanded to learn more than previously.

This may contribute to many teachers, especially those from low-performing schools, not welcoming this curriculum change. They see it as making the curriculum more difficult for the students. For example, teachers said that the topics that were usually in the upper levels were brought down to be studied in the lower levels. Additionally, some teachers complain that the re-sequencing of contents makes it more difficult for some grades more than others.

> *Some contents previously present in grade 10 is now in grade 7. Like some of the upper secondary contents were changed to lower secondary. And some of the lower secondary contents were changed to primary. Some of the contents in grade 4-6 were changed to grade 1-3. Everything is more difficult (Mathematics teacher, OBEC2 school (school 2)).*

> *From when the curriculum changes, I feel like it's hard work for grade 9. I think if separated by the topics, grade 9 is tough, but grade 8 is light. For grade 8, the topics that are difficult are Force and Moments, Vectors, Scalar, something like this. But for grade 9, there are Genetics, which is already difficult, and there are also Electricity, Electronics, and Energy, all the difficult topics, and Astronomy. They are all in grade 9. I don't understand why they are changing the curriculum, when it's even more difficult! (Science teacher, OBEC1 school).*

Arguably, the teachers' reflections on the curriculum are justified. Many topics have been moved down to lower grade levels than in the 2008 curriculum. For example, rational and irrational numbers and operations of percentages and fractions was in grade 8 in the 2008 curriculum. In the 2017 curriculum, these topics are in grade 7. Additionally, there are topics from Additional Mathematics course (optional course offered to students in some schools) that have been moved to core compulsory Mathematics such as polynomials. Further, another way the new curriculum can be perceived as difficult is that it is no longer sufficient for students being able to “understand” or “explain” things, they must also be able to apply the knowledge to the real-world context.

The change in curriculum reflected the effort of the system moving away from memorisation, and towards a focus on application of knowledge and skills more relevant to a globalised world. While some teachers viewed the change as positive, many teachers in low-performing schools have expressed difficulties in adjusting to the new curriculum. Particularly, they expressed concern that the curriculum is too difficult for their students. Despite this reluctance, it is compulsory for the schools to adopt the curriculum.

Thailand still has a long way to go before scores on international assessments improved, as is part of the goal in the National Education Plan. As discussed, changes in the curriculum and O-NET involve relatively simple applications of knowledge compared to PISA. Students are still not required to show some written work in O-NET. Teachers have remarked that PISA is much more difficult than O-NET. For the curriculum itself, even though there is more focus on skills, the skills are not yet fully integrated into each learning units or indicators. There are also no explanations on what level of skills students should attain, or what they should be able to do at a given grade level. Hence, there remains much improvement to be done.

Changes in assessment

The shift in curriculum is accompanied by changes in assessment. In addition to helping the rewriting of curriculum, IPST was brought in in 2017 to lead changes in O-NET exams in the subjects of Mathematics and Science starting from 2018 (Dailynews, 2017). One teacher speculates that this change is brought about by the changes in interest in PISA.

> *Previously, frankly, O-NET is evaluating just knowing and understanding. There are some applications, but it doesn't go deep into analysing or creating, but lately, with PISA, like when interest in PISA increased, O-NET exam changed, and the organisation that designs O-NET exam, especially Science and Math, changed too. From before, NIETS was the one responsible doing all subjects' exams. But recently, they gave [the responsibility] to IPST to supervise in Science and Math. Hence, the format and style of the exam changed. There are more questions about events or news, for students to read and analyse (Science teacher, Science school).*

The motivation for the change is probably the same one that drove changes in Mathematics and Science curriculum. This shows that there is alignment between changes in curriculum and national assessment, with both moving towards higher-order skills. When there is alignment in the system, the chance of teachers changing their teaching practices increased.

From the teachers' perspectives, the new O-NET exam is more PISA-like and requires more analysis than the previous exams. When prompted further, teachers illustrated that prior to this, O-NET exams emphasised heavily on knowing and understanding things. In some questions, students can simply read the question and answer immediately. With the recent O-NET exams, however, the questions include more context. There may be scenarios, graphs, or additional information related to the contents tested. Students need to be able to interpret the information given before coming to an answer. This clearly made the exam more difficult. One teacher reflected that after the IPST designed the exams, there had been no students scoring 100% on the test.

To illustrate further, one of the questions in O-NET 2018 is as follows.

> Villagers in one village like to use chemical fertiliser and insect repellent in great quantity for a long time when growing crops. This results in a build-up of chemical substances and lowering soil quality. Even though now the villagers stopped using chemicals and improved soil quality, there remains build-up of chemicals in the soil and the ecosystem.
>
> If we want to prevent transferring the chemical build-up in the soil to other living organisms, how should villagers choose to utilise the crops?
>
> 1. Grow rice for export
> 2. Grow potatoes to be transformed to animal feed
> 3. Growing casava to become fuel
> 4. Growing corn to produce corn flour

As the teachers described, the question provides contextual information, linking to the application of knowledge in real life. Students need to analyse the information given to be able to answer the question, as opposed to being asked to recall facts. This is arguably a question that is similar to PISA (see example questions in Appendix E2). In this respect, the new exams showed a move away from memorisation and asked students to think more before they can answer. Nonetheless, there is still large gaps between O-NET and PISA exams. The question shown above is arguably one of the most difficult ones on the test. There are no short answer questions, or questions that require students to defend their viewpoints. Many teachers acknowledged that the new O-NET exams are more complex, but it is not on par with PISA. Many teachers use the word ‘critical thinking’ when they refer to skill tested in the exams. However, critical thinking here means when the question requires more than one-step of the thinking process, or when there is application of knowledge.

> *I think the exams are difficult, and the exams test critical thinking. To solve the problem, it’s not like they see the question and they can answer straight*

> *away. They need to think, think to be able to solve problems. … There are some requiring calculation. But it's not like putting in formulas or numbers and you get the answer right away, no. … like Moments and Force … sometimes they don't ask directly about Moments, they ask about division or angles, how far is this from that, how far is this from the fulcrum (Science teacher, OBEC1 school).*

In some respect, this gradual change can be beneficial in helping stakeholders adjust to the new exams. It also explains how the PISA scores remain low even after shifting the focus to be more PISA-like.

Other than the changes made from IPST stepping in, teachers mention changes in exam difficulty as driving the year-to-year score change. Participants speculated that each year's exam is different in terms of difficulty. Consequently, scores went up when the exams are easier, or when there was more emphasis on topics students are good at.

> *The fluctuations of the scores are from the exams too. Some years, from the exam that I see, some years they are difficult, some years the contents are quite complex, you cannot think with only one step, you need to think more, 2-3 steps. … Some years the exams are easy and O-NET scores would increase, … on average, the scores increased, for us but also for other schools. Some years the exams are difficult, then the scores decreased. Because … for each content, there is unequal number of questions, if coincidentally, the exam contents happen to be contents that many students like and can learn a lot, they would get high scores … These affect O-NET scores directly (Mathematics teacher, OBEC2 school (school 2)).*

This is similar with quantitative result that shows O-NET's exam format, content focus, and skill requirements changed from 2012 to 2015. When the exams vary in difficulty over time, there is less confidence that trends over time reflect changes in education quality. This calls into question of O-NET's positive changes during our time period of interest, as it may have just been reflecting changes in exam difficulties.

Changes in pedagogy

The previous sections discuss changes in curriculum and assessment that participants reported as affecting them significantly. Participants reported both curriculum and assessment changes as driving changes in how teachers teach in classrooms and how they prepare for examinations. They mentioned changes in O-NET as the main driving force for their changes in practice. This highlights the importance of system alignment and using assessment as a force to drive change. This section summarises practices that changed over time.

From the participants' perspectives, there are changes in what they think students should know and be able to do over time. This has been influenced largely by changes in the curriculum and assessment. Previously, the emphasis of teaching has been previously on memorising contents. Students who can score well on examinations are those who are able to remember taught contents. They perceive teachers' jobs are to ensure students learn all the contents. Now, however, simply knowing and memorising contents are not enough for students to perform well on examinations. Participants mention that students should also be able to apply their knowledge in the real-world context.

> *10 years ago, I was a teacher here. Before, in terms of teaching and learning, we emphasise knowledge and we are quite rigorous in giving knowledge and dive deep into each topic. … But at present, we don't emphasise so much on only knowledge. We need to teach them other skills too, like more applying. There are many skills. The exams today are quite difficult … We need to teach them other skills. Only knowledge is not enough (Vice principal, OBEC2 school (school 2)).*

With this shift in mindset, the participants were ready to teach beyond memorisation. To help teachers make this adjustment, IPST provided some resources to support teachers in the form of a curriculum guide and textbooks. Not only that the 2017 curriculum provides specific suggested activities (for each indicator in Science and for selected indicators in Mathematics), they also support the use of technology more than the previous curriculum. The manual provides online resources and promotes the use of calculators in learning Mathematics. In theory, these should help teachers in changing their pedagogies to match curriculum requirements.

The textbooks show alignment with the curriculum. Teachers noted the differences from the previous textbooks. Notably, the questions and activities encourage students to think more and apply their knowledge, in contrast to simply memorising the contents. This alignment is critical in ensuring that teachers teach according to the curriculum, as well as helping teachers adjust to the changes.

> *In the IPST textbooks now, the books for the new curriculum, there are many questions that stimulate thinking, like have the students think, say if it's this situation, what should be done. [The textbook] incorporates this in quite a lot for students to learn (Mathematics teacher, OBEC2 school (school 2)).*

> *In the new curriculum, there are questions that are interesting to do, that may be aligned with PISA or something that supports thinking process. There is more critical thinking in the questions in the book (Mathematics teacher, Private school (school 5)).*

As a result of the change in curriculum, assessment, and required textbooks, teachers changed the way they teach and what they teach in class. Many participants mentioned that they shifted from lecturing to allowing more participation in classrooms. One teacher illustrated how she allowed students to think and form their own knowledge rather than lecturing them the contents. With this, students exercise their thinking skills as opposed to waiting for teachers to tell them the answers.

> *For example, I was teaching additional Science subject in grade 8 about behaviours, about social communication. There are sounds, for example. Instead of telling them what's available, I showed them a video of animals expressing different kinds of behaviours, and I designed the worksheet for them to answer what is it that they saw, what are the animals trying to communicate, and the resulting behaviours, what did they use to communicate, like gestures and so on … And teaching like this, students like them, because they get to think and do, and it's not boring like how it was before. Before this we teach lecture-based, we just teach and students need to write down. Previously, I think it's good, but when we change, this is better. (Science teacher, Private school (school 5)).*

The change in pedagogy is met with varying levels of success. This teacher noted that by teaching this way, students are able to answer applied questions in the examinations. Other teachers said that students show more interest in learning but are still unable to correctly answer questions that require application of knowledge in the examinations. Nonetheless, even though the pedagogy does not guarantee that students will learn the higher-order skills, it is arguably an improvement from lecturing alone.

Even with support provided by IPST, some schools still struggle to implement the new curriculum and keep up with changes in O-NET. Some claimed that the textbooks are too difficult, and the skills students are required to have are beyond their abilities.

> *Previously, the questions in the books are easy, and the questions in the exams are difficult. But now, they changed. The books are more difficult. But it's too difficult. They adjusted the curriculum in 2018, 2 years ago. And the kids cannot follow (Mathematics teacher, OBEC2 school (school 3)).*

This is probably the result of weak foundational skills and literacy problems. The schools were already struggling with O-NET exams that focus on memorisation. It is more difficult for them to teach students how to apply knowledge as well. In this sense, the teacher's guide is less useful for these schools. There should be targeted support to help the school keep up with the curriculum.

Nonetheless, it is compulsory for teachers to adopt the new curriculum and textbooks. Many teachers in low-performing schools adapted by teaching easier contents to students, even though more complex contents are listed in the curriculum. They revert to using previous versions of textbooks that have easier contents.

> *I teach according to the books. But I need to adjust all of it. I need to find the easy questions in the previous books. … I analyse the kids. … I might not teach exactly like the books. I need to adjust based on the students' context. The kids are like this, and I need to adjust downwards. Write new exercise problems, new worksheets, by myself, and train the students on these (Mathematics teacher, OBEC2 school (school 3)).*

When this happened, students do not learn all contents as intended by the curriculum, and schools ended up with extensive tutoring at the end of the year to keep students up to speed before examination. This might suggest that the curriculum was not designed with students of all academic levels in mind. As a result, gaps in implementation exists when low-performing schools struggle to catch up. At the very least, schools with students with poorer academic backgrounds should be supported with implementation. For example, it might benefit them to have more time adjusting to the curriculum compared to other schools.

When asked about when the shift came about, schools vary in their answers. For most schools, they changed when they perceive O-NET to be more difficult. As the participants struggle to perform well on O-NET that includes more application questions, they started teaching more of applied questions in class. This imply that the changes are recent (from 2017 – 2018). From up to 2015, most schools are still focusing heavily on memorising contents. At that time, O-NET itself also emphasises on knowing, rather than application of knowledge. This can partially explain why students perform so poorly on questions requiring higher-level skills, and specifically on PISA, which places emphasis on problem-solving and critical thinking.

The new curriculum does not affect the high-performing schools as much as the low-performing ones. This may be because they had already gone above the curriculum knowledge in their school curriculum. The participants from the high-performing schools did not mention curriculum or textbook difficulties.

> *When the curriculum from the MOE adjusted, we followed, but just that …, the MOE adjusted moving this and that, we did similarly, but the contents and rigor, it's the same for us. … We have our core curriculum, which is the national curriculum. Our contents are equal to the national curriculum, and also go over the national curriculum. Ours is higher, greater, and deeper (Vice principal, Science school).*

Appendix E: Appendix to Chapter Six

Some of the tables shows in Chapter Six use standardised values of PISA and TIMSS. This Appendix provides the same tables with unstandardised scores.

Appendix E1: O-NET exam format

Proportion of types of O-NET exam items are presented in tables E1-1 and E1-2. The bars are visual representations of coverage percentages. Examples of constructed response and complex multiple choices are provided in Appendix D2. Note that constructed response are calculation problems in which students are required to fill in a number. This may be different from some of PISA constructed response questions, which requires students to write simple explanations. Information here are taken from exam blueprints and actual released exam papers.

Table E1- 1 O-NET Mathematics exam format from 2010-2020

Years	Multiple choices	Constructed response	Exam time
2010	15 questions, 2 points each 60%	5 questions, 4 points each 40%	45 minutes
2011	15 questions, 2 points each 60%	5 questions, 4 points each 40%	45 minutes
2012	25 questions, 3 points each 75%	5 questions, 5 points each 25%	1 hour
2013	25 questions, 3.2 points each 80%	5 questions, 4 points each 20%	1.5 hours
2014	25 questions, 3.2 points each 80%	4 questions, 5 points each 20%	1.5 hours
2015	25 questions, 3.2 points each 80%	4 questions, 5 points each 20%	1.5 hours
2016	25 questions, 3.2 points each 80%	4 questions, 5 points each 20%	1.5 hours
2017	18 questions, 4 points each 72%	7 questions, 4 points each 28%	1.5 hours
2018	20 questions, 4 points each 80%	5 questions, 4 points each 20%	1.5 hours
2019	20 questions, 4 points each 80%	5 questions, 4 points each 20%	1.5 hours
2020	20 questions, 4 points each 80%	5 questions, 4 points each 20%	1.5 hours

Source: NIETS (2021)

Table E1- 2 O-NET Science exam format from 2010-2020

Years	Multiple choices	Complex multiple choices	Constructed response	Exam time
2010	12 questions, 2.5 points each 75%		3 questions, 3-4 points each 25%	25 minutes
2011	N/A	N/A	N/A	N/A
2012	40 questions, 2 points each 80%	5 questions, 4 points each 20%		1.5 hours
2013	40 questions, 2 points each 80%	5 questions, 4 points each 20%		1.5 hours
2014	40 questions, 2 points each 80%	5 questions, 4 points each 20%		1.5 hours
2015	40 questions, 2 points each 80%	5 questions, 4 points each 20%		1.5 hours
2016	40 questions, 2 points each 80%	5 questions, 4 points each 20%		1.5 hours
2017	40 questions, 2 points each 80%	4 questions, 5 points each 20%		1.5 hours
2018	40 questions, 2 points each 80%	4 questions, 5 points each 20%		1.5 hours
2019	40 questions, 2 points each 80%	4 questions, 5 points each 20%		1.5 hours
2020	36 questions, 2.2 points each 79.2%	4 questions, 5.2 points each 20.8%		1.5 hours

Source: NIETS (2021)

Table E1- 3 Proportion of each content domain in Mathematics

	Numbers and Operations	Measurem-ent	Geometry	Algebra	Data Analysis and Probability	Mathematical Skills and Processes
2012	16%	13%	16%	16%	19%	20%
2013	19%	13%	16%	16%	16%	20%
2014	16%	13%	16%	19%	16%	20%
2015	16%	10%	19%	19%	16%	20%
2016	17%	16%	22%	23%	22%	0%
2017	24%	12%	20%	24%	20%	0%
2018	24%	12%	20%	24%	20%	0%
2019	24%	12%	20%	24%	20%	0%

Source: Mapping based on examinations found at NIETS (2021)

Table E1- 4 Proportion of each content domain in Science

	Living Things and Processes of Life	Life and the Environm-ent	Substances and Properties of Substances	Forces and Motion	Energy	Change Process of the Earth	Astronomy and Space	Nature of Science and Technology
2012	14%	4%	18%	20%	16%	10%	12%	6%
2013	10%	14%	24%	12%	20%	8%	8%	4%
2014	10%	14%	24%	12%	20%	8%	8%	4%
2015	14%	16%	12%	30%	14%	0%	14%	0%
2016	18%	12%	12%	30%	16%	2%	10%	0%
2017	20%	8%	22%	14%	16%	14%	6%	0%
2018	22%	8%	20%	12%	18%	14%	6%	0%
2019	23%	6%	21%	12%	19%	10%	9%	0%

Source: Mapping based on examinations found at NIETS (2021)

Appendix E2: Item examples from the assessments

This section presents examples of exam questions used in the three assessments. O-NET questions (in Thai), can be found directly from the NIETS website (2021). PISA and TIMSS released some items to the public (See (NCES, 2021; OECD, 2013, 2016a)). Questions from each cognitive domain are presented here.

Mathematics items

PISA

PISA 2012 Released item	PISA 2012 Released item
Mount Fuji is only open to the public for climbing from 1 July to 27 August each year. About 200 000 people climb Mount Fuji during this time. On average, about how many people climb Mount Fuji each day? A. 340 B. 710 C. 3 400 D. 7 100 E. 7 400 **Item type:** Simple multiple choices **Content domain**: Formulating situations mathematically **Cognitive domain**: Change and relationships	Toshi wore a pedometer to count his steps on his walk along the Gotemba trail. His pedometer showed that he walked 22 500 steps on the way up. Estimate Toshi's average step length for his walk up the 9 km Gotemba trail. Give your answer in centimetres (cm). **Item type:** Constructed response **Content domain**: Employing mathematical concepts, facts, procedures and reasoning **Cognitive domain**: Change and relationships

PISA 2012 Released item

For a homework assignment on the environment, students collected information on the decomposition time of several types of litter that people throw away:

Type of litter	Decomposition time
Banana peel	1-3 years
Orange peel	1-3 years
Cardboard boxes	0.5 years
Chewing gum	20-25 years
Newspapers	A few days
Polystyrene cups	Over 100 years

A student thinks of displaying the results in a bar graph.

Give one reason why a bar graph is unsuitable for displaying these data

Item type: Constructed response

Content domain: Interpreting, applying and evaluating mathematical outcomes

Cognitive domain: Uncertainty and data

Sources: OECD (2013, 2016a)

TIMSS

TIMSS 2011 Released item

Which of these shows how 36 can be expressed as a product of prime factors?

A. 6 × 6 B. 4 × 9

C. 4 × 3 × 3 D. 2 × 2 × 3 × 3

Item type: Simple multiple choices

Content domain: Numbers

Cognitive domain: Knowing

TIMSS 2011 Released item

Ann and Jenny divide 560 zeds between them. If Jenny gets of the money, how many zeds will Ann get?

Item type: Constructed response

Content domain: Numbers

Cognitive domain: Applying

TIMSS 2011 Released item

Place the four digits 3, 5, 7, and 9 into the boxes below in the positions that would give the greatest result when the two numbers are multiplied.

Item type: Constructed response

Content domain: Numbers

Cognitive domain: Reasoning

Source: NCES (2021)

O-NET

O-NET 2015 Question 2

Which is false?

1. (-1) + (-2) + (-3) = -6
2. (-1) – (-2) – (-3) = -6
3. (-1) x (-2) x (-3) = -6
4. {(-1) ÷ (-2)} ÷ (-3) = $-\frac{1}{6}$

Item type: Simple multiple choices

Content domain: Number and Operations

Cognitive domain: Knowing

O-NET 2015 Question 1

If we want to separate 25 boys and 20 girls into equal-sized groups, what is the smallest possible number of groups?

1. 4 groups
2. 5 groups
3. 9 groups
3. 13 groups

Item type: Simple multiple choices

Content domain: Number and Operations

Cognitive domain: Applying

O-NET 2012 Question 26

Write numbers 1,2,3,4,5 in each square such that the numbers in the same row, column, and diagonally are not the same. Some numbers are already filled. What is A?

3	4			5
2				
		A		
				4

Item type: Constructed response

Content domain: Number and Operations

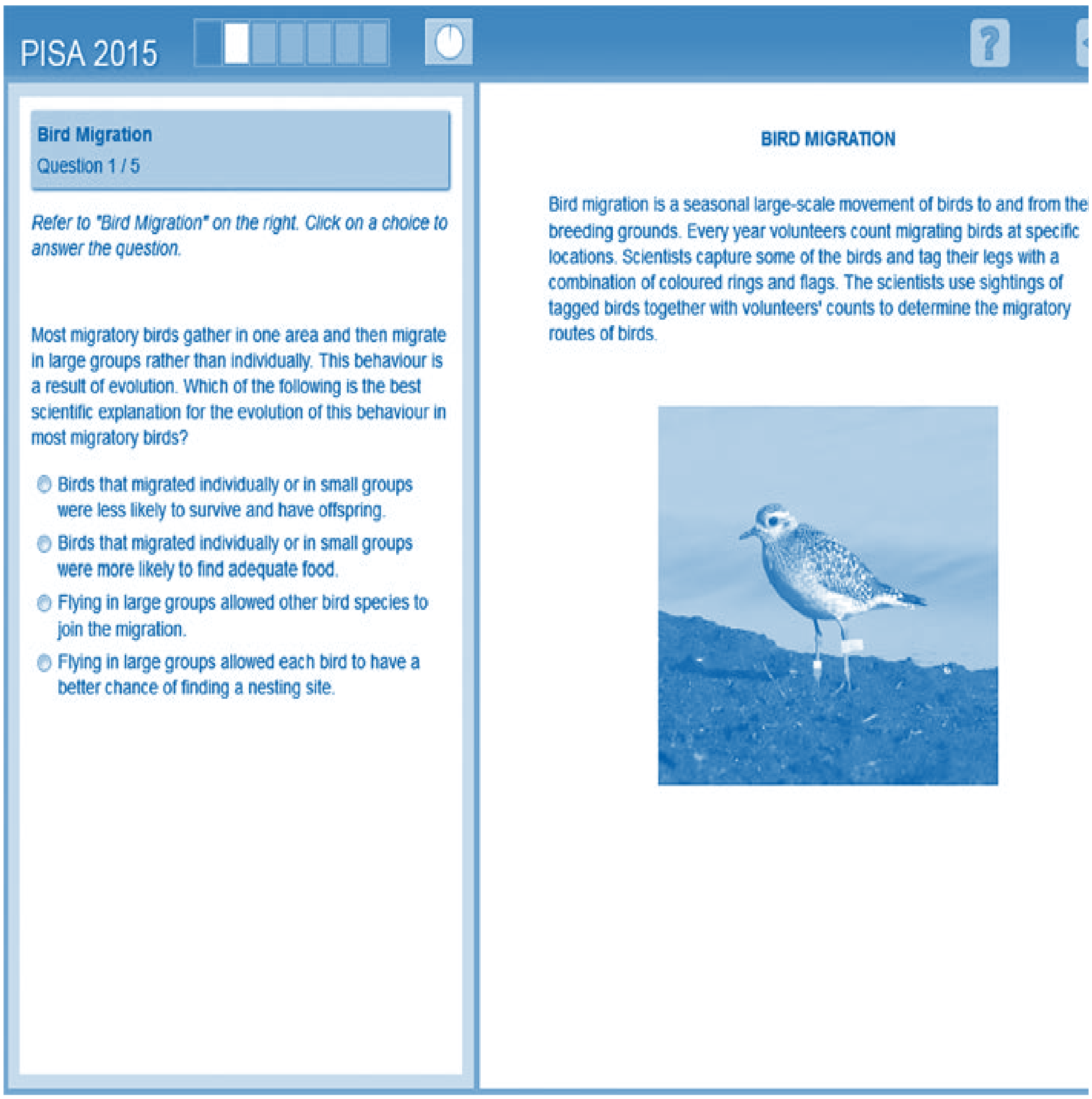
Cognitive domain: Reasoning

Source: NIETS (2021)

Science items

<u>PISA</u>

<u>PISA 2015 Released item</u>

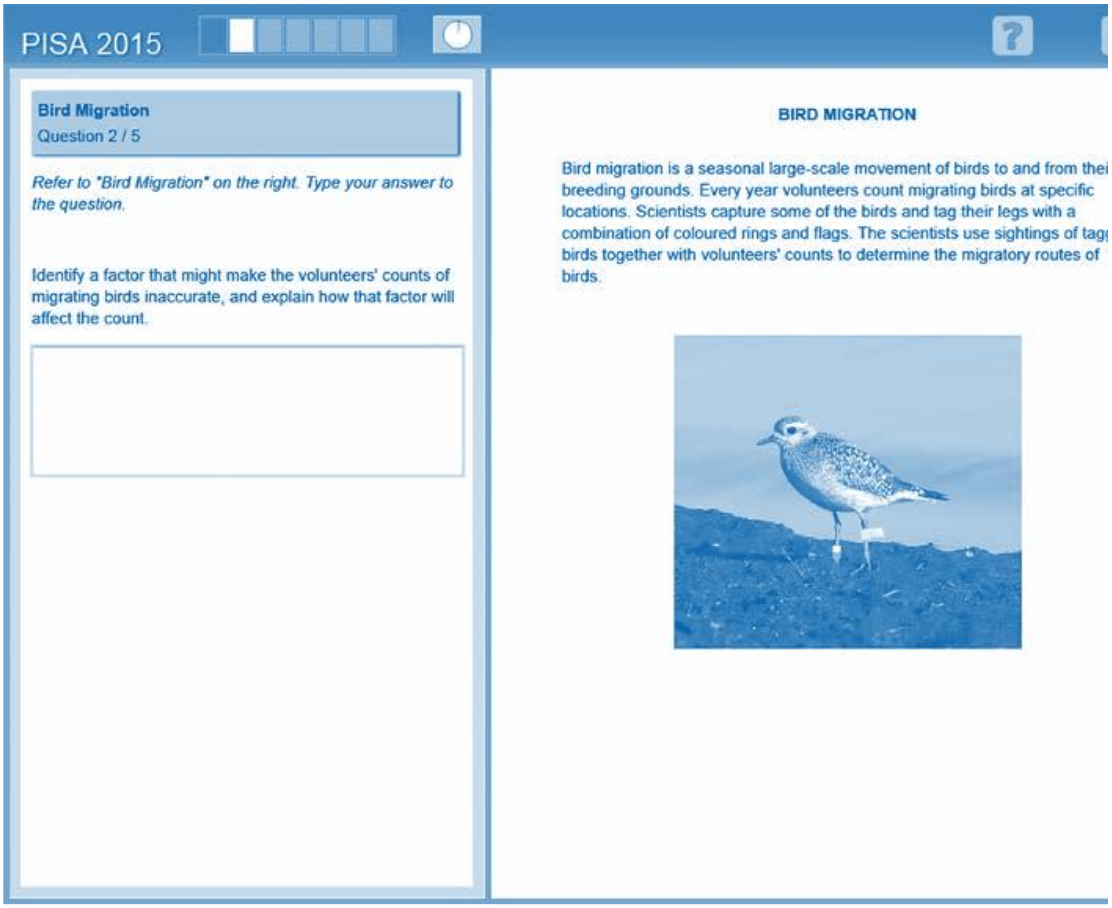


Item type: Simple multiple choices

Content domain: Living systems

Cognitive domain: Explain Phenomena Scientifically

<u>PISA 2015 Released item</u>

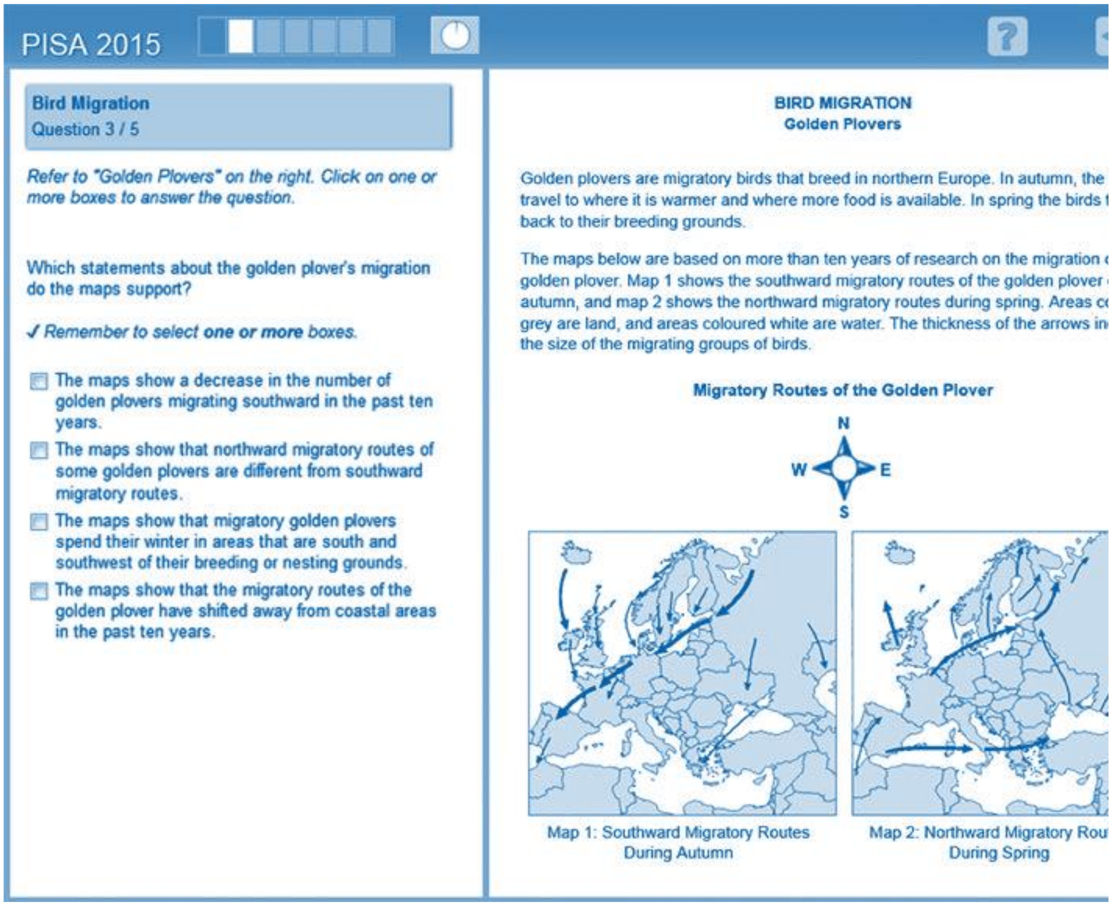


Item type: Constructed responses

Content domain: Living systems

Cognitive domain: Evaluate and design scientific enquiry

PISA 2015 Released item



Item type: Complex multiple choices

Content domain: Living systems

Cognitive domain: Interpret data and evidence scientifically

Sources: OECD (2013, 2016a)

TIMSS

TIMSS 2011 Released item

Bacteria that enter the body are destroyed by which type of cells?

A. white blood cells

B. red blood cells

C. kidney cells

D. lung cells

Item type: Simple multiple choices

Content domain: Biology

Cognitive domain: Knowing

TIMSS 2011 Released item

The following table shows the classification of some animals into two categories.

Category 1	Category 2
Rabbit Giraffe Elephant	Frog Spider Lion

Which of the following was used to classify these animals?

A. organs used in breathing B. food source

C. method of reproduction D. pattern of movement

Item type: Simple multiple choices

Content domain: Biology

Cognitive domain: Applying

<u>TIMSS 2011 Released item</u>

Susie has a potted plant. She sets up an experiment that shows that water travels through a plant into the air.

Which experiment would show this?

A. Put water in a container under the pot; water will disappear from the container.

B. Cover one of the stems of the plant with a plastic bag and water the plant; drops of water will be seen in the bag.

C. Place a cut stem from the plant in a plastic bag; water will be seen in the bag.

D. Place a cut stem from the plant in a glass of colored water; the plant's leaves will change color.

Item type: Simple multiple choices

Content domain: Biology

Cognitive domain: Reasoning

Source: NCES (2021)

O-NET

O-NET 2012 Question 1

Which organelle control amount and type of substance passing in and out of the cell?

1. Cell wall 2. Nucleus 3. Cell membrane 4. Nucleus membrane

Item type: Simple multiple choices

Content domain: Living Things and Processes of Life

Cognitive domain: Knowing

O-NET 2015 Question 4

One man has a homozygous black skin. He married a white woman and have 2 children. What are the chances of his two children's skins?

1. All black 2. All white 3. Either black or white 4. One black and another one white

Item type: Simple multiple choices

Content domain: Living Things and Processes of Life

Cognitive domain: Applying

O-NET 2015 Question 2

Mr. Dam used sweet mango to breed, when the fruits grew, he brought one fruit from each tree, tasted, and found out that:

The fruit from the first three tasted bland. The fruit from the second tree tasted sweet. The fruit from the third tree tasted sour.

Mr. Dam concludes the method in breeding the mangos as follows:

A. The first tree was bred using budding

B. The second tree was bred using grafting

C. The third tree was bred using seeds

D. All three trees were bred using layering

Which is a credible conclusion?

1. A and B   2. B and C   3. C and D   4. D and A

Item type: Simple multiple choices

Content domain: Living Things and Processes of Life

Cognitive domain: Reasoning

O-NET 2015 Question 41

Breeding beans with purple and white flowers gives the first-generation children (F1) in all purples. Breeding F1 together gives the second generation children (F2) with the proportion of dominant:recessive characteristics = 3:1. Which is a correct conclusion?

1. Purple flower is a recessive characteristic, while white flower is a dominant characteristic

2. The first-generation children (F1) all have dominant characteristics

3. F1 have characteristics of purple and white flowers

4. F2 have characteristics of all purple flowers

5. F2 have characteristics of purple and white flowers

6. F2 are white:purple flowers = 3:1

Item type: Complex multiple choices

Content domain: Living Things and Processes of Life

Cognitive domain: Reasoning

Source: NIETS (2021)

Appendix E3: Tables with unstandardised PISA and TIMSS scores

Table E3- 1 Gaps in learning outcomes by each subgroup

Subgroups	Subjects	PISA	TIMSS	O-NET
School types	Math	166	165	3.33
	Science	178	140	2.79
School types – excluding Satit and Science schools	Math	41	43	0.40
	Science	49	41	0.40
Regions	Math	55	64	0.52
	Science	52	59	0.40
Urban-rural	Math	32	66	0.46
	Science	42	58	0.40
School size	Math	54	42	0.74
	Science	69	39	0.66
Performance percentile (P90-P50)	Math	108	123	1.69
	Science	109	103	1.61
Performance percentile (P50-P10)	Math	100	103	0.93
	Science	95	104	0.97
Number of books at home	Math	73	107	N/A
	Science	66	99	N/A
Parental education	Math	29	54	N/A
	Science	32	49	N/A
Gender	Math	1	18	0.21
	Science	6	20	0.23

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table E3- 2 Mathematics ranking by school type

School types	PISA Rank	PISA Score	PISA Dif from mean	TIMSS Rank	TIMSS Score	TIMSS Dif from mean	O-NET Rank	O-NET Score	O-NET Dif from mean
Science	1	556	**132**	1	633	**202**	1	3.33	**3.03**
Satit	2	503	**79**	2	567	**136**	2	2.09	**1.79**
OBEC2	3	431	7	4	438	7	4	0.39	0.09
Private	4	405	**-19**	3	445	14	3	0.40	0.10
OBEC1	5	403	**-21**	7	402	**-29**	6	0.00	**-0.31**
BMA	6	391	**-33**	6	431	0	7	-0.01	**-0.31**
Local	7	390	**-34**	5	433	2	5	0.03	**-0.28**
Average		424			431			0.30	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

Table E3- 3 Science ranking by school type

School types	PISA Rank	PISA Score	PISA Dif from mean	TIMSS Rank	TIMSS Score	TIMSS Dif from mean	O-NET Rank	O-NET Score	O-NET Dif from mean
Science	1	567	**137**	1	626	**170**	1	2.73	**2.48**
Satit	2	510	**80**	2	569	**113**	2	1.77	**1.52**
OBEC2	3	438	8	4	461	5	3	0.34	0.10
Private	4	410	**-20**	3	470	14	4	0.28	0.04
OBEC1	5	397	**-33**	7	429	**-27**	5	0.00	**-0.25**
BMA	6	391	**-39**	5	458	2	7	-0.06	**-0.30**
Local	7	389	**-41**	6	458	2	6	-0.03	**-0.28**
Average		430			456			0.24	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

Table E3- 4 Mathematics ranking by region

	PISA			TIMSS			O-NET		
Regions	Rank	Score	Dif from mean	Rank	Score	Dif from mean	Rank	Score	Dif from mean
BKK and perimeter	1	448	**33**	4	435	4	1	0.62	**0.32**
Lower north	2	423	8	9	400	**-31**	7	0.22	-0.08
Central	3	421	6	5	434	3	6	0.24	-0.06
East	4	418	3	1	464	**33**	3	0.34	0.03
Upper NE	5	415	0	7	413	-18	8	0.15	-0.15
Upper north	6	407	-8	3	440	9	2	0.41	0.11
South	7	405	-10	2	450	**19**	5	0.28	-0.02
West	8	405	-10	6	429	-2	4	0.31	0.00
Lower NE	9	393	**-22**	8	402	**-29**	9	0.10	-0.20
Average		415			431			0.30	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: PISA and TIMSS taken from official Thailand reports (IPST, 2017; OECD & IPST, 2018) and own calculation using O-NET 2015 microdata

Table E3- 5 Science ranking by region

	PISA			TIMSS			O-NET		
Regions	Rank	Score	Dif from mean	Rank	Score	Dif from mean	Rank	Score	Dif from mean
BKK and perimeter	1	454	**33**	4	461	5	1	0.50	**0.26**
Lower north	2	428	7	9	428	**-28**	6	0.17	-0.08
Central	3	425	4	5	459	3	7	0.15	-0.09
East	4	422	1	1	487	**31**	3	0.27	0.03
Upper NE	5	419	-2	3	464	8	8	0.11	-0.13
Upper north	6	418	-3	7	438	-18	2	0.38	0.14
South	7	412	-9	6	452	-4	5	0.21	-0.03
West	8	410	-11	2	472	16	4	0.23	-0.01
Lower NE	9	402	**-19**	8	430	**-26**	9	0.10	-0.14
Average		421			456			0.24	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: PISA and TIMSS taken from official Thailand reports (IPST, 2017; OECD & IPST, 2018) and own calculation using O-NET 2015 microdata

Table E3- 6 Mathematics ranking by urban-rural location

	PISA			TIMSS			O-NET		
Location	Rank	Score	Dif from mean	Rank	Score	Dif from mean	Rank	Score	Dif from mean
Urban	1	450	**16**	1	482	**51**	1	0.59	**0.29**
Rural	2	418	-6	2	416	-15	2	0.13	-0.17
Average		424			431			0.30	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

Table E3- 7 Science ranking by urban-rural location

	PISA			TIMSS			O-NET		
Location	Rank	Score	Dif from mean	Rank	Score	Dif from mean	Rank	Score	Dif from mean
Urban	1	463	**33**	1	500	**44**	1	0.49	**0.25**
Rural	2	422	-8	2	442	-14	2	0.09	-0.15
Average		430			456			0.24	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

Table E3- 8 Mathematics ranking by school size

	PISA			TIMSS			O-NET		
Size	Rank	Score	Dif from mean	Rank	Score	Dif from mean	Rank	Score	Dif from mean
XL	1	456	**32**	1	440	9	1	0.72	**0.41**
Large	2	412	-12	2	437	6	2	0.17	-0.14
Medium	3	406	**-18**	3	411	**-20**	3	0.03	**-0.27**
Small	4	402	**-22**	4	398	**-33**	4	-0.03	**-0.33**
Average		424			431			0.30	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA and O-NET 2015, and TIMSS 2011 microdata

Table E3- 9 Science ranking by school size

	PISA			TIMSS			O-NET		
Size	Rank	Score	Dif from mean	Rank	Score	Dif from mean	Rank	Score	Dif from mean
XL	1	466	**36**	1	463	7	1	0.62	**0.38**
Large	2	430	0	2	462	6	2	0.11	-0.13
Medium	3	408	**-22**	3	436	**-20**	3	-0.01	**-0.25**
Small	4	397	**-33**	4	423	**-33**	4	-0.04	**-0.28**
Average		430			456			0.24	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA and O-NET 2015, and TIMSS 2011 microdata

Table E3- 10 Mathematics ranking by gender

	PISA			TIMSS			O-NET		
Gender	Rank	Score	Dif from mean	Rank	Score	Dif from mean	Rank	Score	Dif from mean
Male	1	425	1	2	422	-9	2	0.19	-0.11
Female	2	424	0	1	440	9	1	0.40	0.10
Average		424			431			0.30	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

Table E3- 11 Science ranking by gender

	PISA			TIMSS			O-NET		
Gender	Rank	Score	Dif from mean	Rank	Score	Dif from mean	Rank	Score	Dif from mean
Female	1	433	3	1	465	9	1	0.35	0.11
Male	2	427	-3	2	445	-11	2	0.12	-0.12
Average		430			456			0.24	

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET 2015 microdata

Table E3- 12 Change in gaps in learning outcomes by each subgroup

Subgroups	Subjects	PISA	TIMSS	O-NET
School types	Math	-4	**79**	**1.90**
	Science	**28**	**71**	**0.97**
School types – excluding Satit and Science schools	Math	-16	5	**0.22**
	Science	-8	3	0.07
Regions	Math	13	**-29**	**0.24**
	Science	12	**-19**	0.05
Urban-rural	Math	-14	**31**	**0.22**
	Science	0	**25**	0.07
School size	Math	-12	N/A	**0.40**
	Science	10	N/A	0.14
Performance percentile (P90-P50)	Math	-7	12	**0.70**
	Science	7	6	**0.32**
Performance percentile (P50-P10)	Math	5	-4	**0.23**
	Science	1	-4	0.00
Number of books at home	Math	-11	14	N/A
	Science	0	9	N/A
Parental education	Math	-7	4	N/A
	Science	1	2	N/A
Gender	Math	-7	1	0.19
	Science	-6	5	0.12

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table E3- 13 Score change over time by school type

School types	**Mathematics**			**Science**		
	PISA	**TIMSS**	**O-NET**	**PISA**	**TIMSS**	**O-NET**
OBEC1	**-54**	0	**0.22**	**-75**	3	0.12
OBEC2	-9	-2	**0.46**	-18	-3	**0.22**
Private	-13	**26**	**0.48**	**-27**	**29**	0.18
BMA	-9	-2	**0.24**	**-26**	1	0.18
Local	-16	9	**0.22**	**-35**	8	0.00
Satit	**-31**	13	**0.91**	**-23**	17	0.16
Science			**2.50**			**1.31**
Average	-13	4	**0.40**	**-24**	5	0.18

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table E3- 14 Score change over time by school region

Regions	**Mathematics**			**Science**		
	PISA	TIMSS	O-NET	PISA	TIMSS	O-NET
BKK and perimeter	4	**-44**	**0.54**	-8	**-41**	**0.23**
Central	15	**32**	**0.38**	3	**24**	0.15
Upper north	**-24**	**25**	**0.48**	**-30**	**23**	**0.20**
Lower north	-10	-14	**0.38**	**-26**	-14	0.18
Upper NE	-8	5	**0.30**	-19	8	0.13
Lower NE	-14	-18	**0.30**	**-26**	-12	0.18
South	-13	**40**	**0.42**	**-25**	**36**	0.18
East	-16	**-31**	**0.44**	**-30**	**-21**	**0.21**
West	**-43**	-7	**0.42**	**-45**	**-41**	0.19
Average	-11	4	**0.40**	**-23**	5	0.18

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: PISA and TIMSS taken from official Thailand reports (IPST, 2017; OECD & IPST, 2018) and,own calculation using O-NET microdata

Table E3- 15 Score change over time by school location

School location	Mathematics			Science		
	PISA	TIMSS	O-NET	PISA	TIMSS	O-NET
Rural	-6	-3	**0.24**	**-20**	-1	0.06
Urban	**-19**	**27**	**0.66**	**-18**	**24**	**0.37**
Average	-13	4	**0.40**	**-24**	5	0.18

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table E3- 16 Score change over time by school size

School size	Mathematics		Science	
	PISA	O-NET	PISA	O-NET
S	**-17**	**0.22**	**-36**	0.11
M	-4	**0.26**	**-23**	0.13
L	**-20**	**0.34**	**-23**	0.15
XL	**-21**	**0.62**	**-23**	**0.25**
Average	-13	**0.40**	**-24**	0.18

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA and O-NET microdata

Table E3- 17 Score change over time by performance percentile

Performance percentile	Mathematics			Science		
	PISA	TIMSS	O-NET	PISA	TIMSS	O-NET
P10	-15	4	0.00	**-26**	6	0.16
P50	-10	0	**0.23**	**-25**	2	0.16
P90	**-17**	12	**0.93**	**-18**	8	**0.48**
Gaps						
P90-P10	-2	8	**0.93**	3	0.03	**0.32**
P90-P50	-7	12	**0.70**	6	0.07	**0.32**
P50-P10	5	-4	**0.23**	-4	-0.04	0.00

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata

Table E3- 18 Score change over time by gender

School location	Mathematics			Science		
	PISA	TIMSS	O-NET	PISA	TIMSS	O-NET
Female	**-16**	**5**	**0.49**	**-16**	7	**0.24**
Male	**-19**	**4**	**0.30**	**-20**	2	0.12
Average	-13	4	**0.40**	**-24**	5	0.18

Meaningful coefficients with magnitude greater than 0.2 SD are in bold
Source: Own calculation using PISA, TIMSS, and O-NET microdata