UNIVERSIDADE CATÓLICA PORTUGUESA

Shadow schooling: an international comparative analysis using data from PISA

Rita Santos Silva

Católica Porto Business School
2021

UNIVERSIDADE CATÓLICA PORTUGUESA

Shadow schooling: an international comparative analysis using data from PISA

Master's Final Assignment in the modality of Dissertation
presented to Universidade Católica Portuguesa
to fulfill the requirements for the degree of MSc. in
Management

by

Rita Santos Silva

under the supervision of
Prof. Dr. Maria da Conceição Andrade Silva Portela

Católica Porto Business School, Universidade Católica Portuguesa
March 2021

Acknowledgments

I would like to thank Prof. Dr. Conceição Silva for the availability and tireless support throughout this entire process, for sharing with me her accumulated knowledge and allowing me to develop a better and more complete Master's Final Assignment, being much more than just my supervisor.

Also, a sincere thank you to my family, my boyfriend, friends and my psychologist for the support and advices shared during this period, for always standing by my side.

Resumo

Shadow schooling é um fenómeno que tem vindo a crescer em termos de importância e incidência nos últimos anos. A par do seu desenvolvimento ser generalizado no este asiático há já muito tempo, na Europa, por exemplo, o seu desenvolvimento é muito mais recente e menos explorado. Apesar da tendência de crescimento, a sua incidência e "forma" pode variar com o país.

Mesmo estando generalizado, não há muitos países a recolher dados oficiais neste tópico, principalmente devido à pouca importância dada pelas entidades políticas ao *shadow schooling*. No entanto, a OCDE recolhe dados que contêm informação sobre o apoio extra que pode ser importante no âmbito da comparação entre países, permitindo que o fenómeno seja mapeado e sejam evidenciadas diferenças e semelhanças nos sistemas educativos entre países.

Para atingir estes objetivos recolhemos dados do PISA 2009, 2012 e 2015, começando por uma análise descritiva e depois uma análise comparativa, avaliando o desempenho das escolas de acordo com as suas características.

Os resultados obtidos apontam para um crescimento considerável do fenómeno entre 2009 e 2015, acompanhado da redução dos *scores* do PISA. Tendo a conta a correlação entre as variáveis, o apoio extra parece ser "corretivo", exceto na Coreia do Sul que parece mais orientado para assegurar a vantagem competitiva dos alunos. Para além disso, através da análise comparativa de escolas, parece ser apenas na Coreia que o apoio extra contribui positivamente para o desempenho das escolas, já que noutros países intensos em apoio extra parece contribuir negativamente, como no Peru. Isto deve ser tido em conta pelas entidades nacionais ao nível da política educativa do país.

Palavras-chave: Explicações; PISA; Data Envelopment Analysis; Apoio extra.

Abstract

Shadow schooling is a phenomenon that has been growing both in importance and extent in the past years. Whereas its development in East Asia has been intense for many years now, in Europe, for example, its development is much recent and less explored. Even though there is a general growth trend, its extent and "form" may differ across countries.

Despite being more generalized, there are not many countries collecting official data on the topic, mainly due to the unimportance given by policy analysts to shadow education. However, there are data collected by OECD containing information on the topic that can be valuable for cross-country comparison, allowing us to map the phenomenon and try to understand differences and similarities between education systems around the world.

To achieve these goals, we collected data from PISA 2009, 2012 and 2015, starting by a descriptive analysis of the data and then, benchmarking countries to compare their schools' performances according to their characteristics.

The results obtained point out to a considerable growth of the phenomenon between 2009 and 2015, with a decrease in the PISA scores. According to the correlation between variables, additional instruction seems to be essentially remedial, except in South Korea that leans more towards student enrichment. Also, through the benchmarking analysis only in Korea does it seem that additional instruction is positively contributing to enhance the schools' performance, whereas in other countries where the phenomenon is intense, like Peru, it seems to negatively contribute to their performance. This should be considered by national entities regarding their education system policies.

Keywords: Shadow schooling; PISA; Data Envelopment Analysis; Additional instruction.

Index

Figure Index

Table Index

Introduction

Shadow education has been growing significantly in Europe, nowadays considered a "standard feature of daily life for many families" (Bray, 2017, p. 473), being well established in East Asia for many years now (Ireson, 2004; Manzon & Areepattamannil, 2014). Therefore, this topic is preponderant and should be in the sight of policymakers in order to legislate private tutoring, so that social inequalities are avoided, and reflection about the role and state of the current national education system is undertaken (Bray, 2020).

With the increase of shadow education some fragilities emerge. These may be general and applicable to most countries and others may be country dependent. Lacking data availability on the nature and extent of shadow education is referred by many researchers as a major concern. Bray (2020) highlighted that this may be due to the fact that private tutoring "has been barely on the agenda of either researchers or policy analysts" (Bray, 2020, p. 18) because they considered it to be temporary when it appeared. Whereas Ireson (2004) referred the difficulty in obtaining samples that are representative of students and tutors, because they work independently or do not want to expose themselves.

Even though this phenomenon is widespread around the world, the drivers of its development and "form" in each country are different, but there can be similarities and patterns associated. This highlights the need for evaluating each country such that trends and the direction of evolution are put in evidence and can be compared with the other countries'.

Nowadays, cross-country databases like the OECD Programme for International Student Assessment (PISA) are available, but its use has not been intense in cross-country analysis, especially with a quantitative approach.

Therefore, understanding the extent of shadow education and how countries differ regarding this phenomenon has an undeniable relevance. The main goals of this thesis are to map the evolution of the phenomenon across three cycles of PISA understanding how preponderant variables evolved in culturally different countries, and benchmarking schools in the same countries to understand in what ways do countries differ on their use of shadow education and the results obtained. Thus, in the end of this dissertation, two questions should be answered: how do we characterize and map the evolution of shadow schooling globally? What are the main differences and similarities in the evolution between countries? This is going to be answered using data from PISA 2009, 2012 and 2015 starting with a descriptive analysis and ending in a benchmarking analysis to compare different performances of schools from various countries. A descriptive analysis can be found soon in Silva & Silva (forthcoming) showing a summary of what we will see here, but in a slightly different perspective.

The first chapter, "Literature Review", highlights relevant literature on the topic and clearly exposes what has been done in research and the main results. The following chapter, "The Data Envelopment Analysis Technique", explains the research method that is going to be applied, and "Empirical Analysis: Data and Variables" explains the data collection process and explores the data obtained, putting forward a more descriptive analysis of the evolution. After that, there is another empirical analysis chapter, but this time applying a Data Envelopment Analysis model to assess the performance of schools on an international level named “Empirical Analysis: Performance Scores”. The final chapter, “Conclusions and Future Research”, summarizes the main conclusions obtained and their importance when it comes to educational policy. It also states the main limitations of the study, as well as some recommendations regarding future research on the topic.

Chapter 1
Literature Review

Shadow schooling relates to paid tutoring on academic subjects given outside the school environment to students from primary to university level (Bray, 2020). This concept can be broad and, therefore, it needs some clarification when it is used.

Shadow education is spreading globally and has recently been gaining more attention from researchers in the field of Education and Social Sciences, especially since the 2000s (Manzon & Areepattamannil, 2014), even though it still remains in the "shadow" for many regulators and policymakers throughout the world. It is important to deeply understand this phenomenon and its implications on the role of the national education systems (Bray, 2020), and on the socioeconomic inequalities that might emerge with its establishment in each country (Berberoğlu & Tansel, 2014; Javadi & Kazemirad, 2020; Liao & Huang, 2018; Liu & Bray, 2017; Pallegedara & Mottaleb, 2018; Safarzyńska, 2013; Šťastný, 2016).

Private tutoring has been a reality for a long time in East Asia (Ireson, 2004; Manzon & Areepattamannil, 2014), mainly in countries such as Japan, South Korea, China, and, also, Hong Kong, being considerably embedded in their culture by now.

Recently, it has been growing significantly in Europe where it can be divided in four regions that are distinguishable according to the reasons that triggered its development (Bray, 2017). These regions are Northern Europe, Southern

Europe, Western Europe, and Eastern Europe. Around the same time, countries around the world, irrespective of whether they were developed or developing economies, like Canada (Davies, 2004), Ethiopia (Melese & Abebe, 2017), Sri Lanka (Pallegedara, 2012), Turkey (Berberoğlu & Tansel, 2014) and others, started to see this phenomenon grow, each at its own pace, with different drivers and objectives. Thus, it is preponderant to characterize, map and compare this evolution around the world where, in a general manner, it seems that there is a growth tendency.

1. Geography and scope

Looking at the literature, there are not many cross-country studies on shadow education. Researchers have been focusing more on understanding shadow education in depth in a specific country, as nearly 66% of the 29 empirical articles reviewed for this research have a single-country scope. In line with what was previously mentioned, most single-country articles were on Asian countries, as we can see in Figure 1.

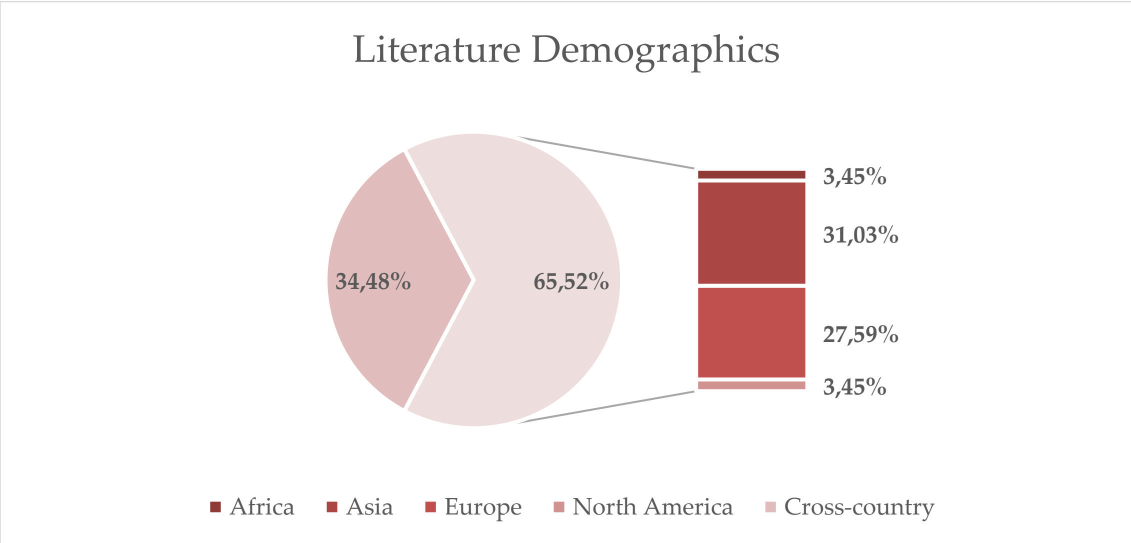

Figure 1: Literature Demographics.

Some studies follow quantitative approaches, which are the most common, representing almost 69% of the articles reviewed (e.g. Baker et al. (2001) and Song et al. (2013), that will be better analyzed later), while others followed qualitative approaches, relying on gathering documentation from several national studies to understand the phenomenon (e.g. Bray (2020), Manzon & Areepattamannil (2014) and Ventura & Jang (2010)).

There are studies that were developed within the scope of cross-country comparisons and with the final objectives of understanding the prevalence of the phenomenon worldwide, as well as similarities and differences between countries. In many cases, researchers were also interested in analyzing the drivers of demand and expenditure on private tutoring. Also, shadow education grew considerably due to the common belief that it leads to an improvement in students' achievement. Nevertheless, its effectiveness is one of the most questioned topics in the literature, as well as how to measure it (Ireson, 2004; Kobakidze & Suter, 2020).

Liao & Huang (2018) developed a study to access the effectiveness of private tutoring in science in Mainland China. According to the authors, private tutoring was not significant in explaining the increase in students' scientific literacy scores. Sharing the same view, Berberoğlu & Tansel (2014) concluded that private tutoring is not significant to explain achievement in science in Turkey. However, when it comes to mathematics and language, they acknowledged a slight, but statistically significant, impact of private tutoring in students' performance. On the contrary, Byun (2014) found that, in general, for the Republic of Korea shadow education was not significant in explaining students' achievement in mathematics. And Baker et al. (2001) found the same but applied worldwide. As for Ryu & Kang (2013), they considered that there is lack of evidence favoring a strong impact of private tutoring expenditure on the overall achievement of students in South Korea.

These disparities can also be attributed to differences in the method used in the research, different samples, countries as well as different variables considered to measure the effectiveness and the extent of private tutoring.

As for Portugal, all research until the present day, to the authors' knowledge, used primary collected data on schools from certain regions. Most of these studies have a small sample varying from 400 students to 1 000 maximum (Azevedo & Neto-Mendes, 2009; Bento & Ribeiro, 2013; Costa et al., 2007). There is one questionnaire distributed by the Portuguese Ministry of Education cited in Ventura et al. (2008) that has a considerable sample of 30 886 students, corresponding to the candidates for the national university access, where 57,9% of the respondents said that they enrolled in private tutoring at some point during their school years.

However, none of the studies tried to benchmark Portugal against other European countries and their respective schools in order to understand differences and similarities in the demand for private tutoring between them.

Note that the main goal of the thesis is not to study the Portuguese case but the international scenery. Nevertheless, it is intended to understand the phenomenon in Portugal, in a general manner, through the conclusions we obtain for other similar countries in Europe (Bray, 2020), and through a benchmarking exercise of Portugal against other countries.

2. Samples and Data

As we can confirm by the articles mentioned above, a considerable variety of studies on shadow schooling are developed nowadays, both at a single-country and cross-country scope.

Inside the single-country studies we can differentiate between studies that use national data collected by the authors, 32%, by official entities, 47%, (in Europe, Germany is the only country where entities develop a national official survey (Kobakidze & Suter, 2020)), and the use of cross-national data available in Programme for International Student Assessment (PISA) or other databases, 16%.

Researchers prefer to collect primary data on this topic or data from national entities so that the data are more detailed and tailored to their goal, as mentioned by Bray (2020). Some examples are Byun (2014) and Ryu & Kang (2013) who used Korea Education Longitudinal Study (KELS) to study the phenomenon in Korea; Šťastný (2016) collected data from two regions of Czech Republic; Pallegedara & Mottaleb (2018) used data from the Household Income and Expenditure Survey in Bangladesh to study private tutoring in this country; In Germany, Guill et al. (2020) used data from a German project to access private tutoring; and, to understand this reality in China, Liu & Bray (2017) used data from the China Family Panel Studies. Note, however, that sometimes primary collected data corresponds to a non-representative small sample.

Other studies, although in smaller number, use cross-national data from international databases such as Liao & Huang (2018) that used data from PISA 2015 to study Mainland China; Safarzyńska (2013) also used PISA, but the 2006 data, to evaluate private tutoring in Poland; and Runte-Geidel (2013) that evaluated the phenomenon in Spain through PISA 2012.

When it comes to cross-country studies, as it involves many countries, there is a need for standardized data to assure comparability. Therefore, existent cross-national databases are the most common choice because these data are hard to collect. It is important to mention that since comparative surveys were created, such as PISA or Trends in International Mathematics and Science Study

(TIMSS), the number of studies on European countries increased considerably (Kobakidze & Suter, 2020).

2.1 Level of analysis

Most quantitative research articles regarding the participation in private tutoring focus on a student-level analysis, representing 85% of the quantitative research articles. However, other levels such as school-level or even country-level are needed to understand this phenomenon, so they should not be ignored (Manzon & Areepattamannil, 2014; Song et al., 2013). In Table 1 there are some examples of different levels of analysis from the literature reviewed.

Table 1: Literature by level of analysis.

Level of Analysis	Articles
Country-level	Baker et al. (2001)
School-level	Addi-Raccah & Dana (2015) and Costa et al. (2007)
Student-level	Azevedo & Neto-Mendes (2009); Bento & Ribeiro (2013); Berberoğlu & Tansel (2014); Byun (2014); Byun et al. (2018); Davies (2004); Guill et al. (2020); Kim & Park (2010); Liao & Huang (2018); Liu & Bray (2017); Pallegedara & Mottaleb (2018); Ryu & Kang (2013); Safarzyńska (2013); Song et al. (2013); Šťastný (2016)

In a country-level study the unit of assessment is the country. It is important as a means to compare trends and differences in the demand for private tutoring among countries. The limitation concerning a study using this level of analysis is the impossibility of accounting for the differences within each country, not allowing for such level of detail. A country-level study can only use variables at a country-level or variables with characteristics of the schools or the students but aggregated at the country-level.

We can also have studies where the level of analysis is the school, which allows the comparison between schools within the countries and between countries. This level of analysis is not yet the most detailed as it does not account for variations in shadow schooling that are due to differences between students. For school-level studies, variables regarding characteristics of the school and the country can be used. Student related variables can only be used if aggregated at the school-level.

The most common type of study, as already mentioned above, is at a student-level which allows for greater detail. Variables indicating student, school and country characteristics can be used in this level of analysis.

Regarding the national case, Costa et al. (2007) developed a study at a school-level, aggregating various student related variables at a school-level to compare four schools in the Portuguese seaside. In addition, Azevedo & Neto-Mendes (2009) and Bento & Ribeiro (2013) fit inside the student-level studies in Table 1.

2.2 Relevant variables in explaining the shadow schooling phenomenon

According to Lee et al. (2009), there are three level factors that can affect the demand for shadow education: the macro-level factors are related to the country system; meso-level factors have to do with the characteristics of the school system; and micro-level factors concern the individual characteristics where we can also include family characteristics.

2.2.1 Student related variables

Figure 2 shows the variables that are usually used on the characteristics of the students.

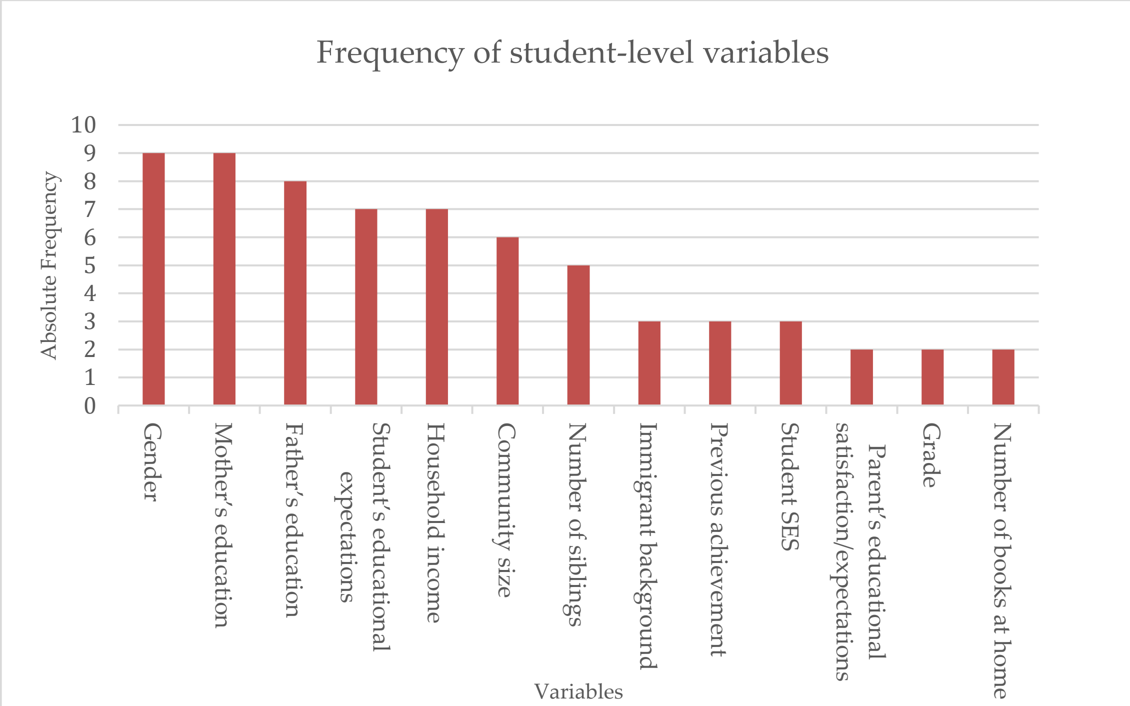

Figure 2: Frequency of student-level variables in the literature.

Normally, the student-level variables that are mostly relevant are: gender (in some articles, female students are more likely to enroll in private tutoring, especially in mathematics); students' educational expectations (this has a positive influence on the demand for shadow education, even though in some studies it is not statistically different from 0); immigrant background (it has very inconsistent results. In some studies, it is not relevant, in others, it positively influences demand as students may present more difficulties, especially in languages, and in the others, it has a negative impact, which may be due to the lower purchasing power of some immigrant families); previous achievement (it, typically, negatively impacts the demand for shadow education); and urbanicity/community size (it is presented as having a positive association with shadow education).

There are also family related factors that consistently influence the demand for private tutoring such: as the household income (in general, it has a positive and significant influence on the demand for private tutoring); and parents' level of education (it has a positive significant influence. If separated, mother's education usually has more impact than father's education, but it highly depends on the cultural context).

Student socioeconomic status (SES) is usually relevant to explain the demand for private tutoring, and normally leads to the emergence of educational inequalities because students with high SES can afford shadow education and students with lower SES may not have that possibility (Byun et al., 2018).

Once again, the results for the same variables may differ due to different samples, different countries, and different measurement of the variables.

2.2.2 School related variables

There are several studies that include school-level variables in their models. We display in Figure 3 the most frequently assessed variables in the reviewed literature.

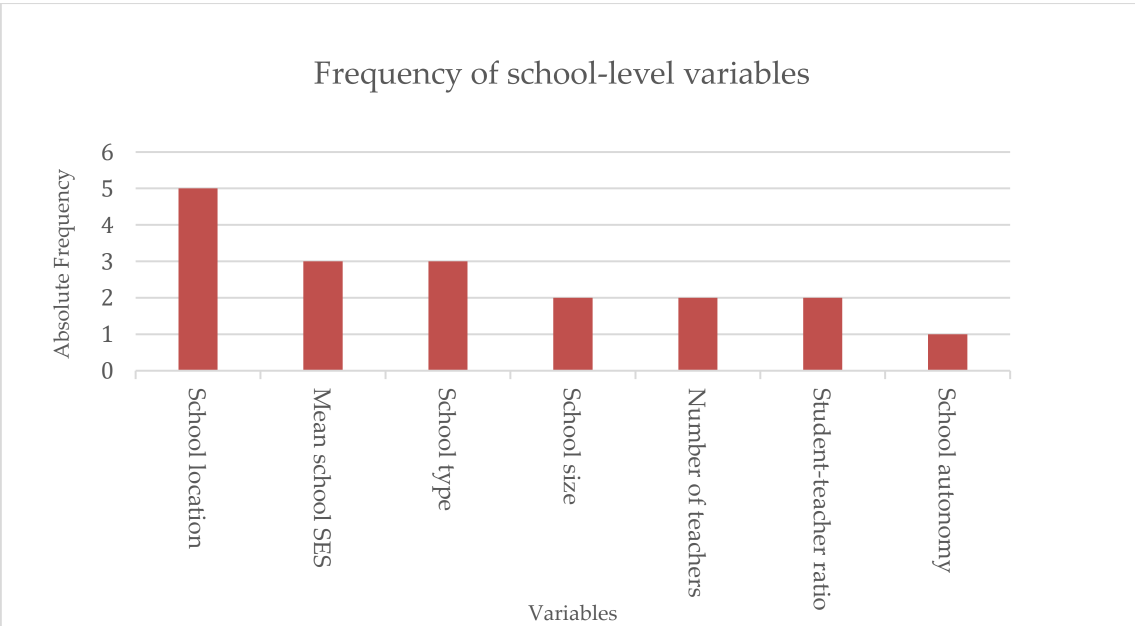

Figure 3: Frequency of school-level variables in the literature.

From all the variables considered, school size was used by Liao & Huang (2018) in their study – that uses both student- and school-level variables - and proved to be negatively associated with participation in private tutoring in China. Liao & Huang (2018) also found that the school's autonomy is negatively associated with private tutoring, meaning that if the school is more autonomous on its decisions regarding teachers, curriculum and resources, then private tutoring will be less prevalent. Using the school location, Addi-Raccah & Dana (2015), Song et al. (2013) and Šťastný (2016) agree on its impact, concluding that schools in cities with more population have a higher demand for private tutoring than in the suburbs/periphery. It is interesting to note that Liao & Huang (2018) did not find statistical evidence on the importance of school location to explain participation in private tutoring.

Regarding the mean school SES, for Liao & Huang (2018) there is no significant association with participation in private tutoring in Mainland China. However, Song et al. (2013) found a statistically significant positive association between mean school SES (measured by the inverse of the poverty rate in a certain school) and private tutoring in countries where there is high incidence of private tutoring and high education quality, specifically Japan. For Addi-Raccah & Dana (2015) this association exists and it is also positive in Israel (the school SES in this article was the combination of the percentage of immigrants, distance from central cities, and the aggregation of two student characteristics at a school-level - parents' education and family income).

2.2.3 Country related variables

Country-level variables are almost not used by researchers. The article by Baker et al. (2001) indicated that public expenditure as a percentage of GNP (Gross National Product), and gross enrolment ratio at elementary and secondary levels were the two most significant variables. As for Byun et al. (2018), lower GDP per capita and the region of the world the country is in are relevant variables to explain the likelihood of participating in fee-paying out-of-school activities. High gross postsecondary enrolment rates were important in easing the negative relationship between academic achievement and participation in shadow education.

Song et al. (2013) also use country-level variables in their study to compare the countries. The authors categorize the countries according to two variables: the high/low demand for private tutoring and the high/low quality of education in the country. Demand for private tutoring was measured by the percentage of students participating in mathematics' private tutoring in each country, where a threshold of 49% was used to distinguish between high and low values (49% is the mean participation rate), and quality of education was measured by the mean mathematics achievement score for each country, where the threshold used was a score of 500.

2.3 Methods

To determine which factors affect participation in private tutoring, Safarzyńska (2013), for example, opted for a logit regression based on the BRR with the Fray factor 0.5 method. Guill et al. (2020) decided on a multilevel logistic regression. Davies (2004) and Šťastný (2016) chose a binary logistic regression. Even though Šťastný (2016) uses both student-level and school-level variables, the author does not recognize the nested structure as Guill et al. (2020) do. When researchers want to know what factors account for the decision of private tutoring and then, separately, what factors determine the expenditure on private tutoring, or elasticity of expenditure, they usually opt for the Hurdle Model as Liu & Bray (2017) and Pallegedara & Mottaleb (2018) did.

Liao & Huang (2018) and Byun (2014) applied very similar strategies to examine private tutoring. Both articles aimed at understanding, firstly, the drivers of the participation in private tutoring and then, its effectiveness on students' performance. For this purpose, they divided their method in three steps. First, Liao & Huang (2018) applied a multilevel logistic regression (which is a hierarchical linear model in line with Song et al. (2013)) to find the drivers, whereas Byun (2014) chose a simple logistic regression. The difference is that the former included student- and school-level variables, but the latter only used student-level variables. Then, both pre-processed the samples: Liao & Huang (2018) through Coarsened Exact Matching (CEM) and Byun (2014) through an alternative which is Propensity Score Matching (PSM). Lastly, to access the effectiveness, the former applied again a hierarchical linear model, and the latter an ordinary least squares regression.

The interest of our study is cross-country comparative research, where we may have studies at country-level, school-level and student-level. At the country-level, the study of Baker et al. (2001) applied Ordinary Least Squares regressions to test the hypotheses. The authors used as independent variables a set of indicators on the use of shadow education for each country. As dependent variables, they selected the existence of high stakes testing, public expenditure as a percentage of GNP, gross enrolment ratio and mean student response rate.

At the student-level, the article by Song et al. (2013) used a two-level hierarchical generalized linear model so that the nested structure between student-level and school-level factors could be recognized. The authors selected student's participation in extra lessons or tutoring in mathematics as the dependent variable. The independent variables can be classified in two main categories: student factors and school factors. The student related factors considered were gender, motivational factors, family background and academic achievement. Whereas related to school they selected the location, mean SES, school's effort to respond to different needs of students through curriculum adjustment, class organization by ability grouping, provision of enrichment/remedial classes, frequent homework, and classroom assessment.

In the article developed by Byun et al. (2018), at the student-level as well, there were different methods for the different goals. Firstly, similarly to Baker et al. (2001), a logistic regression was used to assess the cross-national variations in the association between shadow education and academic performance. Secondly, to evaluate the factors that affect students' participation in shadow education, the chosen method allowed the recognition of the nested structure between individual- and country-level factors, being a two-level logit analysis. The dependent variable was a dummy variable that represented whether the student participated or not in fee-paying out-of-school activities. Whereas the independent variables at the country-level were the GDP per capita, gross postsecondary enrolment rates and geographical regions; and at the individual-level were the family SES, academic achievement, gender, the grade, urbanicity and type of program students are enrolled in.

3. Findings from cross-country studies

In 2001, Baker et al. (2001) used TIMSS cross-national data to examine private tutoring in several aspects. The authors concluded that 39,6% of the international sample of seventh and eighth graders regularly participate in out-of-school activities to increase their performance in mathematics, but there was a significant variation between countries. Regarding the students that indeed participate in shadow schooling activities, in 75% of the sampled countries, as it was the case of Portugal, the purpose was remedial (in opposition to an enrichment purpose which is based on enrolling in private lessons to secure competitive advantage), which occurs when students have a bad performance and attend private tutoring to improve it. To evaluate the purpose of the additional instruction, the authors looked at the logit coefficient of the effects of the scores in math on the participation in additional instruction, controlling for certain factors such as gender and SES. If the coefficient was positive, meaning that more high math ability students participate in shadow education than low ability ones, then the shadow education is essentially for enrichment purposes. The opposite with a negative coefficient.

In the study of Baker et al. (2001), four hypotheses were tested concerning the reasons for the existence of cross-national variations in shadow education. Only the second one, which stated that "In nations with more pronounced qualities of mass schooling, such as high expenditures on public education or large enrolment rates at both elementary and secondary levels, shadow education will be more prevalent as a remedial strategy" (Baker et al., 2001, p. 4), was statistically significant, but in the opposite direction of what was expected. Lower public educational expenditure and smaller enrolment rates resulted in a higher use of shadow education with a remedial purpose. High-stake examinations were found not relevant to explain the prevalence of shadow education, but this can be due to the fact that the sample consists of younger students and, usually, these examinations predominate in latter stages of education, for example end of high school exams.

Later, another cross-country comparative study using TIMSS 2003 by Song et al. (2013) emerged with a similar goal of understanding differences between countries involving the factors that motivate the demand for private tutoring. Song et al. (2013) divided the 46 sampled countries into four groups according to the demand for private tutoring and the quality of education. Then, referring only to groups with high prevalence of private tutoring, they picked two countries where education quality was considered high (Korea and Taiwan) and other two where it was low (Philippines and Romania). The results show that in countries with high incidence of private tutoring and educational quality, the demand is essentially explained by student characteristics such as high academic aspirations, self-confidence and father's education, which shows that it might be leaning towards enrichment purposes (students have high aspirations but also high achievement). A high mean school SES is also a reason behind the demand in this group. Whereas, in the group with low quality of education, shadow education is more demanded by students with high aspirations but low performance (remedial purposes), especially if the school assigns a significant quantity of homework and has more assessments.

In 2018, Byun et al. (2018) developed a study aiming to improve and update the work developed by Baker et al. (2001). The authors used, this time, data from PISA 2012 which had a higher quantity and wider range of countries. They found that, in 2012, about 33% of the 15-year-old students from 64 countries participated in fee-paying out-of-school classes, and, as concluded by Baker et al. (2001), the cross-national variation on the use of these classes is considerable. Regarding the nature of the classes, Byun et al. (2018) also found that they were mainly remedial, as lower achievers are more likely to enroll.

However, Baker et al. (2001) only used characteristics of the countries when assessing out-of-school activities, and Byun et al. (2018) considered that, based on empirical research, individual factors played a major role on the differences between the uses of shadow education. Therefore, in the updated article, they decided to account for individual and country variables. The results at the country-level were very much the same: countries with lower Gross Domestic Product (GDP) per capita and less educational development have a higher incidence of shadow education. Therefore, these conclusions still verify nowadays and are even more evident with new statistical approaches and bigger samples (Byun et al., 2018).

It is interesting that in Baker et al. (2001) Romania was one of the few countries where private tutoring was essentially for enrichment purposes, and, in Song et al. (2013) and Byun et al. (2018) its nature was classified as remedial.

Note that, to the authors' knowledge, there is no cross-country comparative study with a school level of analysis, as the three references we found after applying all the "filters" have a student or country level of analysis. In addition, only Byun et al. (2018) use data from PISA as we are going to use in our research.

4. Limitations and caution in comparative studies using PISA or TIMSS

PISA and TIMSS had a considerable impact around the world because they allowed the development of substantial work in the field of comparative education. By increasing the consciousness of national governments, they led to great changes in educational systems. However, this type of data also has some methodological and measurement problems (Bray et al., 2020; Bray & Kobakhidze, 2014; Rutkowski & Rutkowski, 2016).

The first problem is the definition of the concept of shadow education. Researchers should mention the definition of shadow education they considered in their articles and, even after that, they can be subjected to problems about the broadness of the concept. In PISA 2000, for example, there was a question that required the students to know the distinction between "training to improve your study skills" and "private tutoring" which may have resulted in an overlapping of the answers (Bray & Kobakhidze, 2014).

Furthermore, translation problems regarding PISA have also been highlighted in the past years. National PISA teams try to adapt the concepts to those traditionally used in their country. Still, there are problematic translations which hinder the possibility of cross-national comparison. In Bray et al. (2020) there is an example about a question in PISA 2015 which has the option "Live instruction by a person" and, in China, the translation meant mass tutoring, whereas in Bulgaria indicated individual tutoring.

Attempting to consider the critics made by researchers concerning the questions, PISA's responsible team has been trying to improve the questions throughout the years. Nevertheless, this has a cost which is the comparison among iterations because the questions are not necessarily comparable overtime.

The above stated also applies to TIMSS. Two of the articles evidencing some problems mentioned by Bray & Kobakhidze (2014) were developed by Baker et al. (2001) and Song et al. (2013). Whereas the article of Bray et al. (2020) pointed out some problems on the work of Liao & Huang (2018). Therefore, the main existent studies on cross-national comparisons should be interpreted with caution as they reveal some issues that may affect the reliability of the results, even though they remain valuable references regarding this topic and should be taken into consideration.

5. Conclusion

After considering the existent literature on this topic, we can identify several gaps, such as: the lack of general comprehension of the phenomenon at different levels of analysis, especially, school- and country-level; the lack of data adequality at a national-level in many countries where shadow schooling is not as developed as, for example, in Korea or China; few studies make a detailed international comparison between countries; and researchers still prefer primary national data to cross-national data on the topic.

In this study we aim to address these gaps by performing a cross-country comparative analysis as we have already seen, but considering countries that simultaneously participated in the 2009, 2012 and 2015 cycles of PISA in order to compare the evolution overtime and evaluate differences and similarities between countries relating to the additional instruction phenomenon. This analysis will be at a school-level so that the phenomenon can be understood through a different perspective. Note that, given data availability in PISA, we will use additional instruction in or outside the schools as the variable of interest. This differs from considering only private tutoring classes outside the school system. As a result, our research will not be able to completely assess the extension of shadow schooling across countries but will contribute to that literature in the sense that much of the additional instruction is indeed private tutoring.

Chapter 2
The Data Envelopment Analysis Technique

Charnes, Cooper and Rhodes (1978) were the first authors to develop a seminal work on Data Envelopment Analysis (DEA), which is a non-parametric, linear programming technique. DEA allows the measurement of the relative efficiencies of decision-making units through a ratio composed by the weighted sum of outputs to the weighted sum of inputs, where the weights are the ones that allow each unit to be assessed in its best possible light, i.e., the combination of weights that allows the maximization of the efficiency of the production process, which is represented in Equation 1.

Equation 1: DEA model.

$$Max\ h_0 = \frac{\sum_r u_r y_{rj_0}}{\sum_i v_i x_{ij_0}}$$

$$subject\ to$$

$$\frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \leq 1\ for\ each\ unit\ j.$$

$$u_r, v_i\ \geq\ \varepsilon$$

x_{ij} = quantity of input i for unit j

v_i = weight attached to input i

y_{rj} = quantity of output r for unit j

u_r = weight attached to output r

h_0 = efficiency score

If the DEA model is input-oriented, then the efficiency represents the capacity of a certain decision-making unit to contract its inputs (resources) without compromising the output level (production). Whereas a model that is output-oriented aims at maximizing the output levels while maintaining the level of inputs.

As the field we are currently studying is education, the decision-making units can be schools that use their inputs, such as teachers and computers, to maximize their output which is, normally, their students' academic attainment.

Therefore, DEA is a global performance indicator that provides an efficiency measure which is a distance to a best practice frontier, and this distance to a common frontier allows for the benchmarking of schools in a certain area.

Considering an output-oriented model such as in Figure 4, DMU4 is underperforming on the two outputs because DMU2 consuming the same amount of input (a unitary amount in the figure) produces higher values of both outputs. The radial efficiency of DMU4 to the frontier is given by OA/OA', which is expressed in percentage terms showing how much are the observed outputs of the DMU compared to its target outputs, being this percentage related to the distance of DMU4 to the projected point A'.

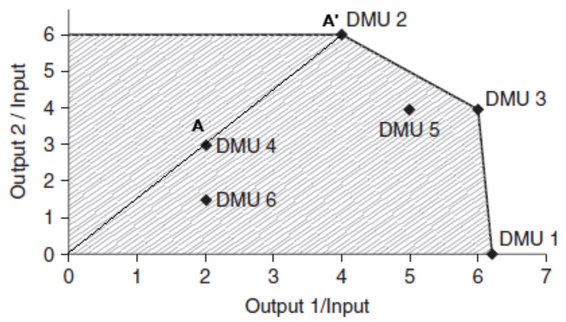

Figure 4: Frontier from an output-oriented model. Adapted from Thanassoulis et al. (2008).

Note that DEA presents a weakness that has to do with the fact that weights can be zero if a certain input or output is not favorable to a decision-making unit. Thus, a unit can be considered efficient just because it neglected a certain input or output when assessing its efficiency. On the other hand, DEA has the advantage of being able to evaluate inputs and outputs expressed in different units of measurement.

1. Meta-Malmquist Index

When we have a sample that can be divided into different homogenous groups according to a certain characteristic, for example time or the type of production unit, we can assess the unit within the homogeneous group and with all units in the sample (Thanassoulis et al., 2008). The meta-Malmquist index is an approach used by Portela et al. (2011) to compare efficiency over time, even though it can be adapted to other circumstances such as the comparison between countries, as we are going to see further.

In Figure 5, we have an illustration of the meta-frontier approach. The meta-frontier is the bold frontier on the figure enveloping all the year-specific frontiers, and it measures the global efficiency of the units on the sample, i.e., ignoring the effect year can have on the efficiency of a certain unit due to technological conditions verified in that year.

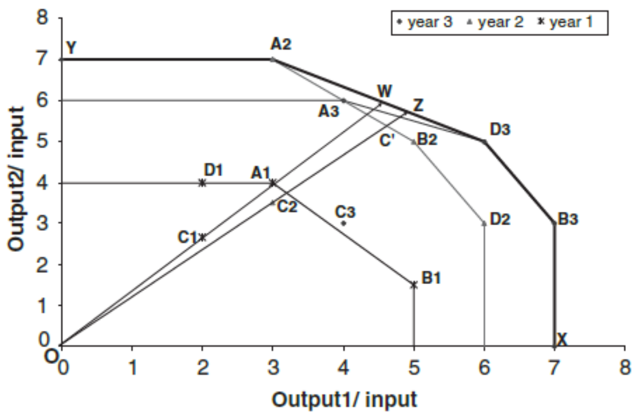

Figure 5: Meta-frontier approach. Source: Portela et al. (2011)

Unit A is an efficient unit in year 1 (OA1/OA1) when we assess it in relation to the year 1 frontier. Comparing unit A with the year 1 frontier, we are assuring that we are comparing it with units under the same technological conditions, controlling for that effect. However, we can compute a meta-efficiency for unit A, which is its radial efficiency irrespective of the year, dividing OA1 by OW. This meta-efficiency is formed by two components that, multiplied, give the global measure efficiency (OA1/OW). These are the within-year efficiency (OA1/OA1) and a program efficiency (OA1/OW), named technological gap, that captures the inefficiency a unit has that is not attributable to the unit itself, but rather to a technological shift that happened between years and made it impossible for unit A to be as productive in year 1 as it was in year 2 (point A2).

2. Empirical research using DEA

DEA models have been used for many years now to study different questions regarding education, especially to study the effectiveness and efficiency of schools. These studies are performed more often with pupil-level or school-level data because it allows for more detailed information.

The first applications of DEA to education using aggregated data at the school level were Bessent & Bessent (1980) (55 elementary schools in a certain district), followed by Charnes et al. (1981).

Bessent & Bessent (1980) considered as outputs students' actual attainments in reading and math (scores measured by the California Achievement Test in 1977); the inputs were the students' prior attainments in reading and math (scores measured by the California Achievement Test in 1976), variables related to the socioeconomic background of the students at the school and the school resources (regarding expenditure in instruction, number of staff, organizational climate and the type of educational process followed - group or individually oriented).

Charnes, Cooper and Rhodes (1981) selected data from the Project Follow Through (a social experiment in US public schools) and considered a DEA model where the output was also the attainment in reading and math, but added a measure of self-esteem. As inputs, they selected the educational level of the mother, the highest occupation of a family member, number of parental visits to the school and number of teachers. The aim of this study was to distinguish managerial from program efficiency, isolating the ineffectiveness on the implementation of the Project Follow Through (managerial part), so that the program effectiveness itself could be identified.

From that point on, many studies were developed, namely using data from PISA, such as the one from Agasisti (2013) that performed a two-stage analysis. Firstly, the author computed efficiency scores for 651 schools in Italy through the data from PISA 2006. There were three DEA output-oriented models where the outputs considered were the mean PISA scores in math or science, and the inputs were the inverse of the student-teacher ratio, the proportion of computers with access to the internet (as a proxy for the school's resources), and the highest educational level or the highest occupational status of the parents (as a proxy for the students' background). A DEA bootstrapping procedure was used to assure the reliability of the results. Secondly, the author performed a Tobit regression with the efficiency scores and a set of external factors that may affect those efficiency scores (such as the type of school, school and class size, percentage of girls, location of the school and parental pressure).

The results of this study show that *licei*, technical schools, and private schools in general, perform better than vocational schools; schools outside the city also perform better; the higher the percentage of girls and parental pressure the lower the performance of the school; and, finally, the more schools there are in the region (competition), the higher the schools' achievement, so more autonomy should be given to schools.

Mancebón et al. (2012) also developed a study aggregating the student information at the school-level with the data from PISA 2006. The authors performed a DEA output-oriented model assuming variable returns to scale. The aim of the study was to conclude on the efficiency of publicly subsidized school versus public schools by decomposing the inefficiencies in managerial and program, concluding that, once the context of the students and school resources are removed, public schools are more efficient.

Note that, in this study, as in opposition to Agasisti (2013), the autonomy of the school proved to be irrelevant to explain the achievement.

Hence, the DEA has been widely used in various topics on the education field. However, to the author’s knowledge, there are no empirical analysis using DEA with the aim of assessing the shadow schooling phenomenon specifically, which is one of the goals of this study.

Chapter 3
Empirical Analysis: Data and Variables

1. Data

To answer the questions and, bearing in mind that the aim is to perform a cross-country comparative analysis, we used the OECD Programme for International Student Assessment (PISA). PISA started in 2000 and it is administrated every three years, being the last one 2018. The main goal of PISA tests is to evaluate whether 15-year-old students have the literacy needed to fully participate in the society after compulsory education (OECD, 2017b). Their literacy is evaluated on mathematics, reading and science. In each cycle, PISA has a focus in one of the subjects which is tested in detail. In 2009 and 2018, it was reading, in 2012, mathematics, and, in 2015, science.

Besides evaluating the cognitive level of students, some mandatory questionnaires are administrated to students to understand their socioeconomic background, their lives, routines/habits regarding learning in and outside school, motivations, and interest in learning. Also, there are mandatory questionnaires for the schools' principals to assess the conditions, resources, environment, and other aspects of schools. Furthermore, there are optional questionnaires to evaluate other domains, in which the Educational Career Questionnaire is the one that matters to our study, because, in certain years, it included questions about additional instruction.

Over the years there has been an increase in the number of participating countries and economies, the ones belonging to OECD and other partners, as we can see in Table 2.

Table 2: Levels of participation in PISA from 2009 to 2018.

Year	Number of economies	Number of schools	Number of students
2009	65	18 641	470 000
2012	65	18 139	510 000
2015	72	17 908	540 000
2018	79	21 903	710 000

The target of PISA, as mentioned before, are 15-year-old students. To guarantee that the sample is representative of the full target population of 15-year-old, OECD uses a two-stage stratified sample design[1]. The first stage is sampling schools eligible from a national list where the probabilities of selection are related to a measure of the size of the schools. Then, the selected schools are put into groups according to their characteristics (stratification variables). The second stage consists of selecting the students within the selected schools, where, from the list of students with 15 years-old, 42 or 35 students are selected with equal probability according to certain requirements (OECD, 2017a).

Following the limitations referred in the last chapter regarding the use of data from PISA, the responsible expert team has been trying to improve the questions so that they are clearer, more informative and avoid overlapping, resulting in comparison problems over the years.

2. Variables

The variables collected are from the Student questionnaire, School questionnaire and, in some cases, the Education Career questionnaire.

[1] Except for Russia.

2.1 Student variables

Two of the variables we consider important to evaluate shadow education are the participation in shadow education and its intensity. These variables can be selected from the Student or Education Career questionnaires. We must be careful when choosing the questions because, as we want to compare this phenomenon over the years, the questions must be asked in the same way or, at least, in a comparable way in the years we are considering.

Comparing the questionnaires from 2009 to 2018, we decided to evaluate the years 2009 (OECD, 2020b), 2012 (OECD, 2020a) and 2015 (OECD, 2020c) where we considered that the questions were comparable (unlike in 2018) and fulfilled our purposes for both variables (see Table 3).

Table 3: Questions about additional instruction in PISA[2].

Year	Question	Scale
2009 / 2012	**"How many hours do you typically spend per week attending <out-of-school-time lessons> in the following subjects (at school, at home or somewhere else)?"** (OECD, 2008b, 2011a)	Ordinal variable with 5 categories: 1 – I do not attend <out-of-school-time lessons> in this subject; 2 - Less than 2 hours a week; 3 - 2 or more but less than 4 hours a week; 4 - 4 or more but less than 6 hours a week; 5 - 6 or more hours a week.
2015	"The following questions ask about any additional instruction in school subjects and other domains that you attend in this school year. This instruction might take place at school or somewhere else, but it is not part of your mandatory school schedule. (…) **In this school year, approximately how many hours per week do you attend additional instruction in the following domains in addiction to mandatory school lessons?"** (OECD, 2014a)	Continuous variable from 0 to 20 hours.

From the questions in Table 3, we constructed two variables for each subject (science, math and reading) for each school: the percentage of participation of students in additional instruction, and the average time of additional

[2] Original question code in Appendix I – PISA questionnaire references.

instruction for those that participated. In this case, for the years 2009/2012, we converted the ordinal scale into a value corresponding to the half of the interval. As a complement to the last two variables, we created a variable regarding the average time of additional instruction in a school, including those students that do not attend additional instruction (i.e., 0 hours).

Furthermore, we also considered relevant the grades students attained in the three tests developed by PISA to study the shadow education phenomenon. In 2009, the test of each student resulted in five plausible values drawn upon distributions of test scores. Therefore, to achieve a mean score for each student, a simple average was calculated. The same was computed for 2012 and 2015 with ten plausible values. After that, we aggregated the mean scores of all students at the school-level obtaining a mean mathematics score, mean science score and mean reading score for each school.

Regarding the socioeconomic background of the students in each school, we are going to use the individual Index of Economic, Social and Cultural Status (ESCS). The ESCS is an index variable created by PISA and derived from the highest parental occupation, highest parental education and how many books students have at home. In 2009, the Principal Component Analysis only considered OECD countries, and, since 2012, all the participating countries were taken into account to estimate the ESCS scores (OECD, 2017a).

Besides that, the components of the ESCS index also changed over the different cycles, leading to an impossibility of comparison over the years. To overcome this barrier, in 2015, OECD recomputed the ESCS from 2009 and 2012 using the same methodology as in 2015 so that comparison in time was possible (OECD, 2017a). Therefore, we used the new ESCS values of 2009 and 2012.

For socioeconomic background, the mean is not the only important variable to take into consideration because we can have the same mean for two schools, one of them being much more heterogenous than the other. So, a measure of dispersion/heterogeneity is also preponderant, which can be the percentage of advantaged students and disadvantaged students in each school.

To classify a student as advantaged we considered that he must be on the top quarter of the ESCS index student distribution and to be considered disadvantaged he must be on the bottom quarter (OECD, 2016, p. 205). As we are considering three PISA cycles, we decided to compute the mean value of the first quartile and the third quartile for the three years. Students bellow the first quartile, which is approximately -1,07, were disadvantaged and above the third quartile, which is approximately 0,56, were advantaged.

2.2 School variables

To evaluate and characterize the phenomenon at the school level, it is preponderant to consider the characteristics of each school presented in Table 4. Here, it is also essential to assure that all the variables (selected from the School questionnaire) are equal or, at least, comparable in all the years.

Table 4: Variables from the School questionnaire.

Variable	Question	Scale
School location	**"Which of the following definitions best describes the community in which your school is located?"**	Categorical variable with 5 categories: 1- a village, hamlet or rural area (fewer than 3 000 people); 2- a small town (3 000 to about 15 000 people); 3- a town (15 000 to about 100 000 people); 4- a city (100 000 to about 1 000 000 people); 5- a large city (with over 1 000 000 people).
School type	**"Is your school a public or a private school?"**	Categorical variable with 2 categories: 1- a public school; 2- a private school.
School size	**"As at <February 1, 20XX>, what was**	Continuous variable with any possible

	the total school enrolment (number of students)?"	natural value (number of boys and number of girls).
Number of teachers	**"How many of the following teachers are on the staff of your school?" Teachers in TOTAL**	Continuous variable with any possible natural value (full-time and part-time).
Fully certified teachers	**"How many of the following teachers are on the staff of your school?" Teachers fully certified by <the appropriate authority>**	Continuous variable with any possible natural value (full-time and part-time).
Computers per pupil	**"Approximately, how many computers are available for these students for educational purposes?" / "At your school, what is the total number of students in the <national modal grade for 15-year-olds>?"**	Continuous variables with any possible natural value.
Computers per pupil with access to the internet	**"Approximately, how many of these computers are connected to the Internet/World Wide Web?" / "At your school, what is the total number of students in the <national modal grade for 15-year-olds>?"**	Continuous variables with any possible natural value.

Source: (OECD, 2008a, 2011b, 2014b)[3]

There are variables that we will not use directly but transform them into new variables. From the number of teachers (see Table 4), which gives us the number of full-time and part-time teachers separately, we can take the total "number of teachers full-time equivalent". To obtain it, we add the number of full-time teachers to the part-time teachers considered at half time. A part-time teacher is any teacher that is employed less than 90% of the time for the full school year (OECD, 2008a, 2011b, 2014b). Consequently, all part-time teachers are counted as being employed 50% of the time. If a school has 10 full-time and 10 part-time teachers, it has 15 teachers full-time equivalent.

[3] Original question code in Appendix I – PISA questionnaire references.

The number of teachers is very different from their qualifications, so we also consider the variable "percentage of fully certified teachers full-time equivalent" that shows how many teachers from the teaching staff in a certain school is fully certified by the appropriate authority.

From the total number of students (see "School size" in Table 4) and total number of teachers full-time equivalent we can obtain the student-teacher ratio.

3. Dataset used for the analysis

In 2015, the Education Career questionnaire was optional and, therefore, many countries did not respond to it. Consequently, those countries did not provide information regarding the participation of their students in additional instruction and were eliminated from the dataset. Portugal did not have data for this questionnaire in 2015, but we did not eliminate it.

Furthermore, it is important to guarantee that all the years have the same countries so that a yearly evolution can analyzed. As a result, every country that was not common to all datasets was not considered in our analysis.

Additionally, there are variables that are considered preponderant to explain the shadow education phenomenon stated in the literature, such as the scores and the socioeconomic background. Bearing that in mind, we decided on eliminating schools with missing values on their mean socioeconomic background. There were no schools with missing values on the scores.

Data were collected at the student-level from PISA and then, aggregated at the school-level, which is our level of analysis. Schools which had more than 40% missing values[4] in our variables of interest (related to additional learning)

[4] We analyzed other alternatives like the 50% and 30% and chose the one that implied lower losses in data but assured some representativity of the schools.

were excluded from the analysis. After cleaning the data, we have 22 countries and 13536 schools.

In Table 5 we show the descriptive statistics of our dataset.

Table 5: Descriptive Statistics.

	Mean			Standard deviation		
	2009	**2012**	**2015**	**2009**	**2012**	**2015**
Math score	499,26	488,00	472,95	58,38	65,63	69,59
Reading score	497,48	489,18	470,33	56,47	66,03	73,35
Science score	505,71	496,70	474,31	56,47	63,29	70,48
% of students attending additional instruction in math	0,32	0,28	0,62	0,18	0,14	0,19
% of students attending additional instruction in reading	0,18	0,19	0,52	0,16	0,14	0,21
% of students attending additional instruction in science	0,19	0,20	0,51	0,16	0,13	0,21
Mean hours of additional instruction in math (total intensity)	0,99	1,12	3,37	0,74	0,76	1,86
Mean hours of additional instruction in reading (total intensity)	0,60	0,77	2,93	0,60	0,67	1,86
Mean hours of additional instruction in science (total intensity)	0,63	0,75	2,86	0,60	0,63	1,86
Mean hours of additional instruction in math	2,49	2,54	4,30	0,91	0,95	1,81
Mean hours of additional instruction in reading	2,71	2,59	4,24	1,19	1,14	1,86
Mean hours of additional instruction in science	2,60	2,46	4,16	1,12	1,05	1,85
Mean ESCS	-0,12	-0,13	-0,31	0,63	0,68	0,73
% of disadvantaged students	0,19	0,19	0,23	0,20	0,21	0,26
% of advantaged students	0,28	0,28	0,24	0,21	0,22	0,22
School size	703,04	704,00	645,67	464,56	521,91	644,58
Number of teachers FTE	59,74	58,22	49,21	35,42	36,41	33,70
% of fully certified teachers FTE	0,94	0,94	0,92	0,16	0,17	0,21
Computers per pupil	0,65	0,77	0,88	0,60	0,79	1,74
Computers with access to the internet per pupil	0,62	0,76	0,83	0,58	0,80	1,69
Student-teacher ratio	12,23	12,19	12,61	4,89	6,50	12,15
Number of schools	5 048	5 626	2 862	5 048	5 626	2 862

4. Data Exploration

4.1 Additional instruction

We started our analysis with our variable of interest: additional instruction.

In Figure 6 we plot the participation rate and intensity, as well as the socioeconomic background aggregated at a country level for mathematics.

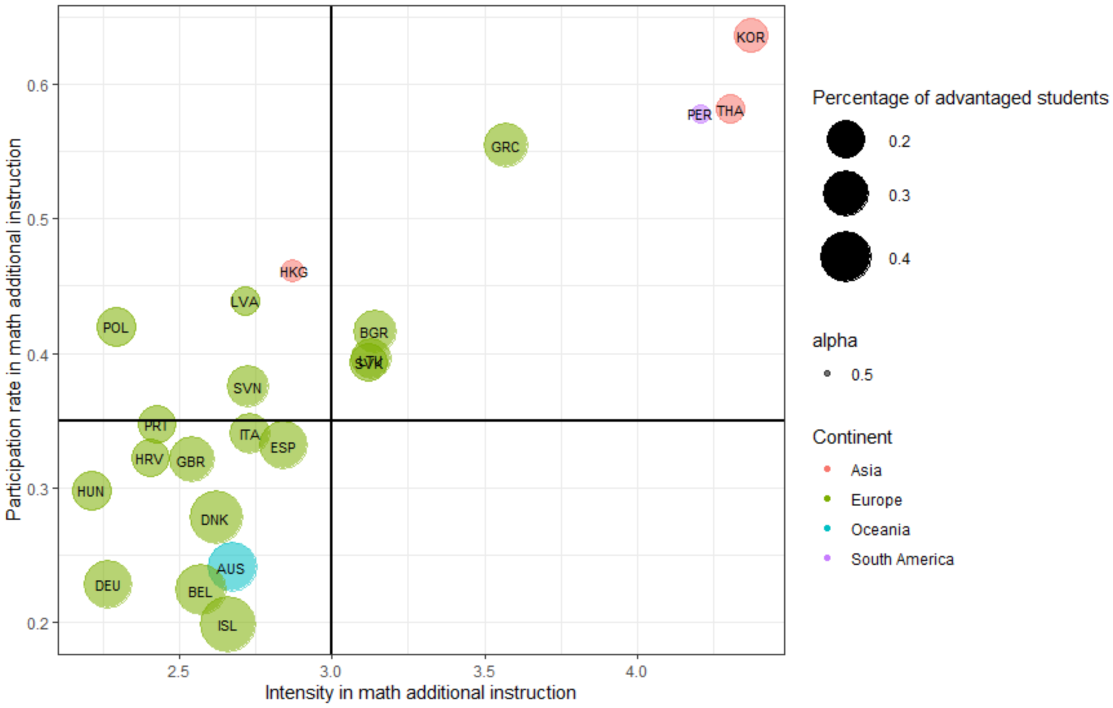

Figure 6: Matrix comparing the participation rate and intensity of math additional instruction per country.

As we can see in Figure 6, the countries with the highest percentage of advantaged students, seen in the size of the bubbles (Island, Belgium, Denmark and Australia), are on the quadrant where intensity and participation in math additional instruction are low (3rd quadrant), so the phenomenon is residual.

In math, the countries with the lowest percentage of advantaged students (Peru, Hong Kong, Latvia, Thailand and Korea) are on the quadrants with high participation rates in additional instruction (1st and 2nd quadrant). Peru, Thailand and Korea also have high intensity (1st quadrant), meaning that in those countries additional instruction takes much time from their students beside the normal school time. Hong Kong and Latvia have a very generalized additional instruction phenomenon in their country but taking less time from their students (quadrant where the intensity is below the reference). This can make us understand that maybe countries with schools with a higher percentage of disadvantaged students (considerably negatively correlated to the percentage of advantaged students) are more intense in additional instruction, and, therefore, this extra support is probably given by schools.

Portugal is located on the quadrant where the phenomenon can be considered residual (both low intensity and participation rates in math additional instruction), but the participation rate is very close to the reference.

In math, there is no country on the quadrant where the intensity is high, and participation is low. This is because there is a high positive correlation between the variables, so the quadrants high-low and low-high are less populated.

In Figure 7, we redo the above plot, this time for reading.

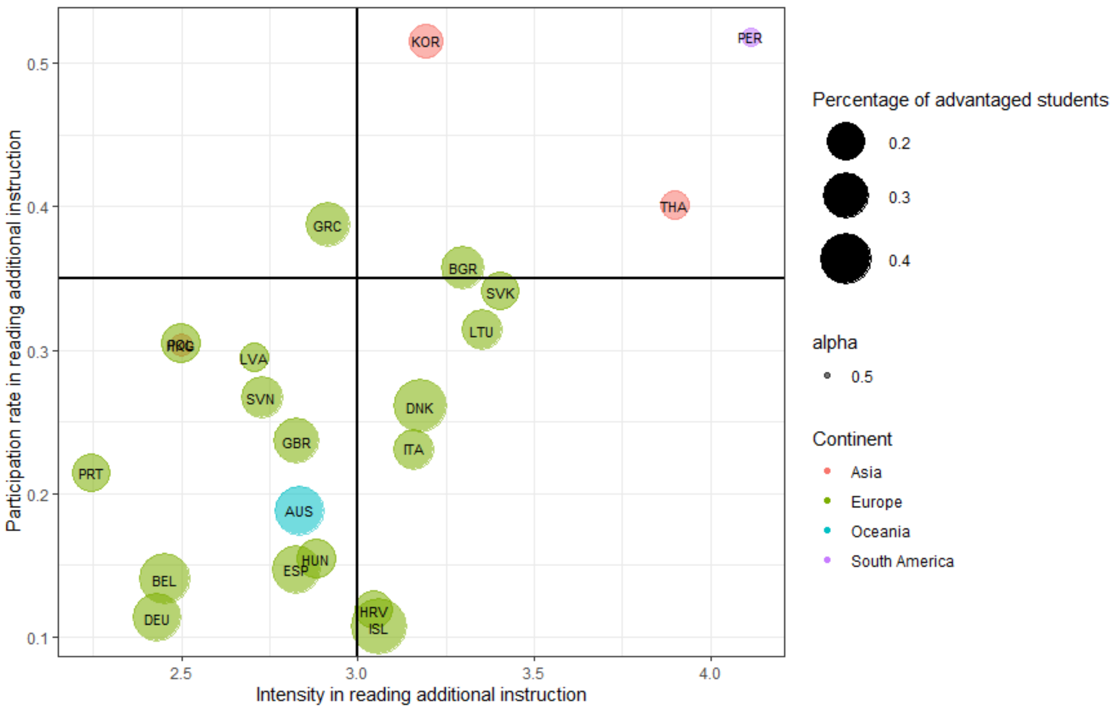

Figure 7: Matrix comparing participation rate and intensity of reading additional instruction per country.

In reading, in general, we can see in Figure 7 more diversity than in math when it comes to possible scenarios. The countries mentioned before as having the highest percentages of advantaged students are more spread around the matrix. Despite that, they still predominate in the area below the reference participation rate (3rd and 4th quadrants), but Iceland is now on the 4th quadrant, meaning that the phenomenon is more targeted in reading because few students have additional instruction, but those who have, have a lot of time consumed by it. Regarding the countries with the lowest percentage of advantaged students, they present roughly the same pattern as in math.

Portugal is in the 3rd quadrant as in math but presenting lower participation rate and intensity.

In Figure 8, we redo the plot, this time for science.

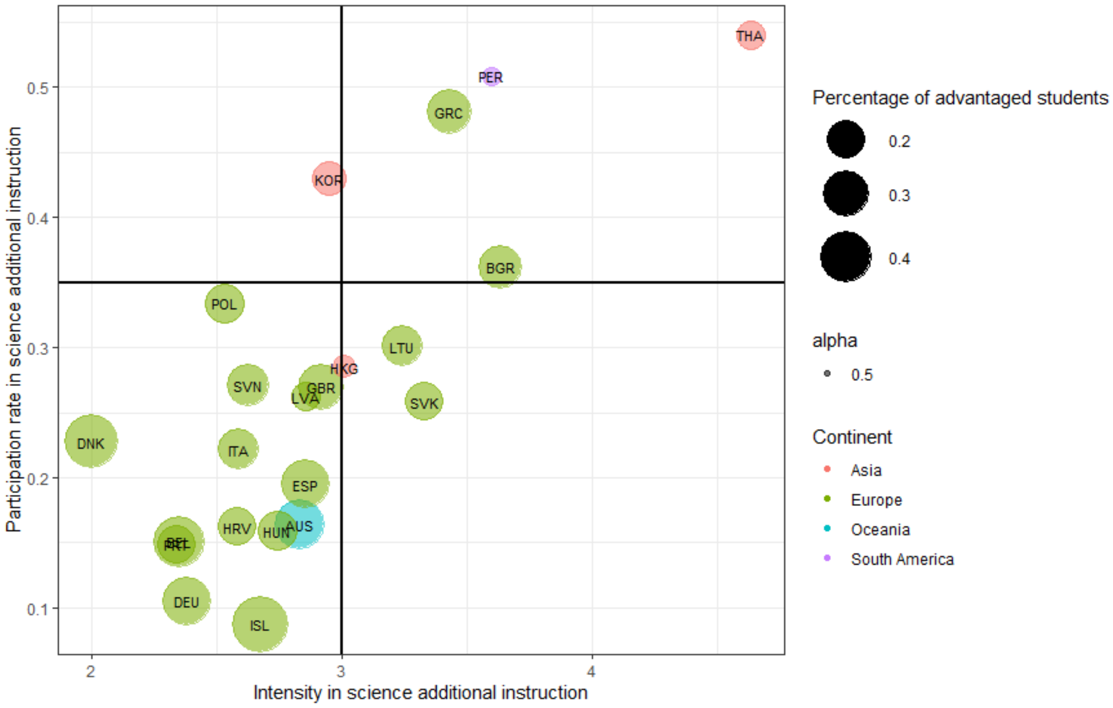

Figure 8: Matrix comparing participation rate and intensity of science additional instruction per country.

In science, through Figure 8, we can see that all the countries mentioned before as having the highest percentages of advantaged students are, once more, on the 3rd quadrant, where the phenomenon is residual.

On the other hand, in the countries with the lowest percentage of advantaged students, there is a slight change of pattern. We can observe that they all still present higher participation rates in science than the reference, except Latvia and Hong Kong.

Portugal is still on the 3rd quadrant where the phenomenon is residual, with lower participation rate than in math and reading, but similar intensity to math.

In math and science (see Figure 6 and Figure 8), we can see a concentration of European countries in the quadrant where the phenomenon is residual, and we can also notice that they are the countries that present the highest percentage of advantaged students. In reading (see Figure 7), there is a slightly higher dispersion. Whereas in Asia, we have three countries mainly located where the phenomenon is generalized (superior quadrants), except Hong Kong, which means the phenomenon is generalized in Asia and more intense than in Europe. Also, Asian countries present a lower percentage of advantaged students.

Concerning Oceania and South America, we do not have enough countries represented, only one per continent. However, in Australia the phenomenon is residual and in Peru it is considerably generalized and intense for all subjects.

Through these figures we can conclude that there is a strong positive correlation between the rate of participation in additional instruction and the intensity of that participation. Therefore, the variable regarding total intensity of participation in additional instruction is a good proxy for these two. Bearing this in mind, from now on we will only be using the mean hours of additional instruction attendance, also considering people that do not participate in additional instruction, which mixes both the intensity and the percentage seen before (we call this variable total intensity of additional instruction).

4.1.1 Country ranking

In Table 6, we present a ranking for the additional instruction variable to understand which countries have the highest mean hours of additional instruction in each subject (including the students that do not attend additional instruction).

Table 6: Country ranking according to the total intensity of additional instruction.

Place	Math	Mean	Reading	Mean	Science	Mean
1º	**KOR**	4,50	**PER**	2,77	**THA**	3,49
2º	**THA**	3,46	**THA**	2,53	**PER**	2,43
3º	**PER**	3,11	**KOR**	2,13	**GRC**	2,28
4º	**GRC**	2,62	**BGR**	1,83	**BGR**	1,99
5º	**BGR**	1,96	**SVK**	1,77	**KOR**	1,67
6º	SVK	1,85	GRC	1,67	LTU	1,44
7º	LTU	1,72	LTU	1,54	SVK	1,38
8º	HKG	1,70	LVA	1,22	HKG	1,20
9º	LVA	1,69	DNK	1,14	POL	1,19
10º	SVN	1,49	HKG	1,09	LVA	1,19
11º	ITA	1,31	ITA	1,07	GBR	1,10
12º	POL	1,31	SVN	1,07	SVN	1,07
13º	ESP	1,26	POL	1,06	ITA	0,83
14º	**PRT**	1,13	GBR	0,96	AUS	0,82
15º	GBR	1,13	AUS	0,91	ESP	0,78
16º	HRV	1,10	**PRT**	0,71	HUN	0,69
17º	AUS	1,04	HUN	0,70	HRV	0,65
18º	**DNK**	1,03	**ESP**	0,58	**DNK**	0,64
19º	**HUN**	0,97	**HRV**	0,58	**BEL**	0,51
20º	**DEU**	0,82	**BEL**	0,53	**PRT**	0,50
21º	**BEL**	0,80	**DEU**	0,48	**DEU**	0,40
22º	**ISL**	0,73	**ISL**	0,47	**ISL**	0,36

We can observe in Table 6 that the countries that have schools with the highest mean hours of additional instruction are similar for all subjects. Therefore, Bulgaria, Greece, Korea, Peru and Thailand are among the countries with the highest intensities. Whereas in countries with the lowest values for the mean hours of additional instruction, we can conclude that Belgium, Deutschland and Island are always among them.

Even though Portugal is only on the bottom five for additional instruction in science, it is always on the bottom half of the total intensity ranking. Math is the subject in which Portuguese schools have the highest mean hours attended. Note that Portugal can be penalized in this ranking because it does not have data for 2015, which was the year when the values declared were the highest.

4.1.2 National evolution of additional instruction

As a way of trying to understand the phenomenon more in-depth, we developed Figure 9, Figure 10 and Figure 11 that have boxplots with the yearly evolution of additional instruction per country, distinguishing between types of school.

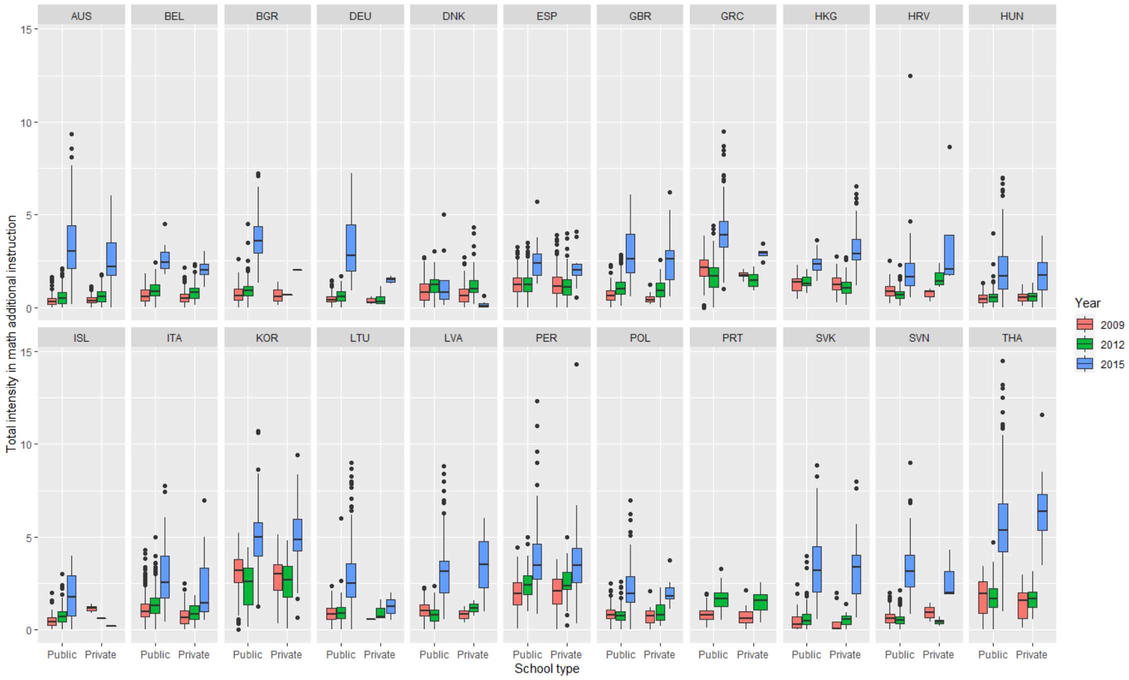

Figure 9: Boxplots for the total intensity of math additional instruction.

In general, there was a growth on the mean hours of participation in math additional instruction throughout the years (see Figure 9), both for private and public schools. The exceptions were Greece, Hong Kong, Korea and Poland, where, from 2009 and 2012, the median of the mean hours of additional math instruction decreased. Besides that, in public schools in Latvia, Croatia and Thailand there was a decrease between 2009 and 2012, as well as in private schools in Spain.

It is important to highlight that Denmark was the only country where the phenomenon decreased in 2015.

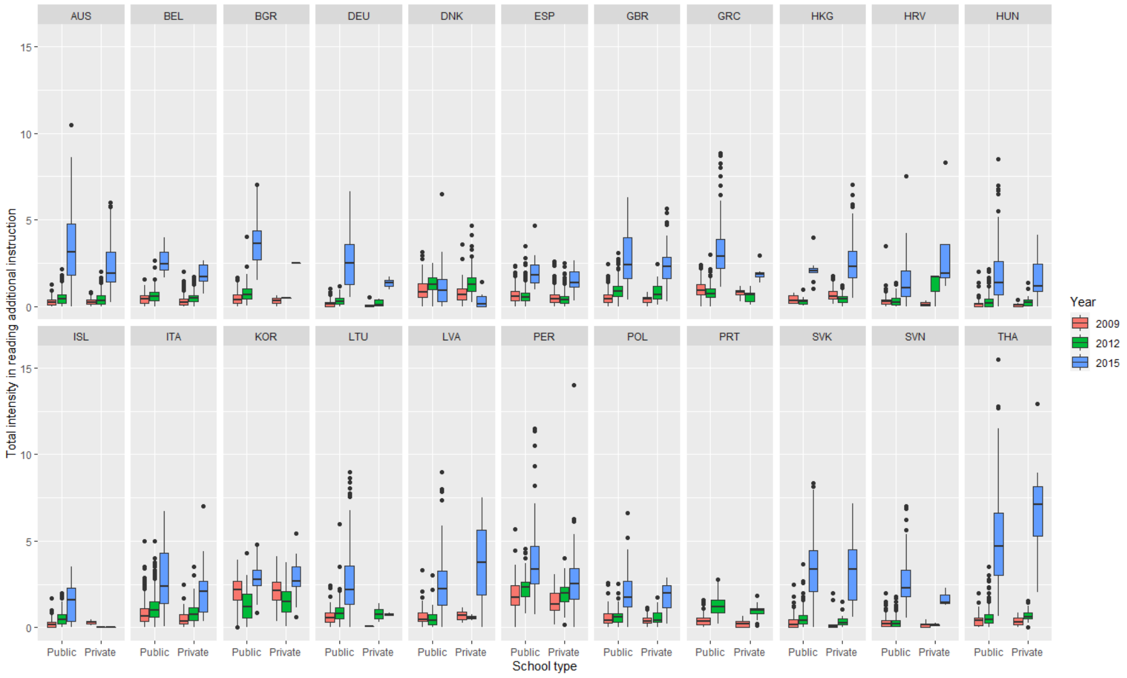

Figure 10: Boxplots for the total intensity of reading additional instruction.

When it comes to reading, we can see in Figure 10 that the increasing trend still verifies, in a general manner, irrespective of the type of school. Greece, Hong Kong, Korea, Latvia and Spain present a decrease from 2009 to 2012, but, in reading, Poland has been increasing from 2009 to 2015. Public schools in Croatia verified a decrease in the mean hours of reading additional instruction from 2009 to 2012, whereas Thailand in reading increased in this period. Denmark maintains its decreasing trend in additional instruction in reading too.

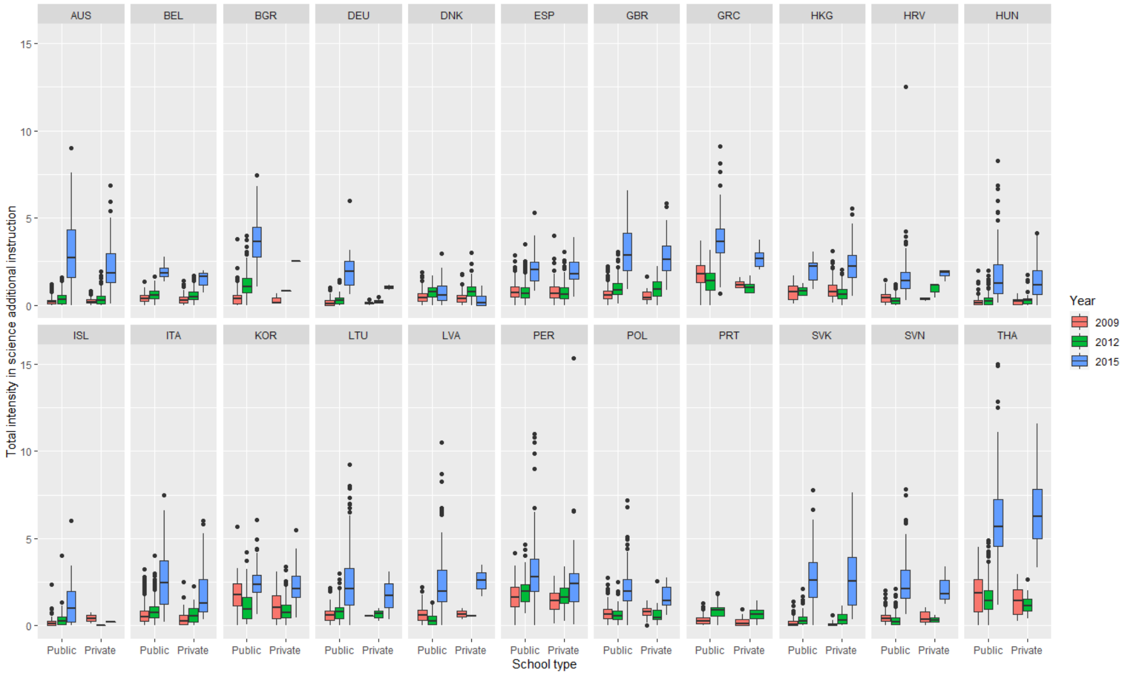

Figure 11: Boxplots for the total intensity of science additional instruction.

It is perceptible through Figure 11 that the evolution verified in math is very similar to that verified in science. The only differences are in Spain, where the intensity of additional instruction decreased from 2009 to 2012, both in private and public schools, and in Slovenian public schools where, unlike in math and reading, there was a decrease in intensity in 2012 compared to 2009.

Something very interesting to note is that the heterogeneity between schools in the countries increased in a general manner in 2015 for all the subjects.

There are countries where there are not enough schools to be evaluated which may have occurred due to the lack of data in relevant variables. This is the example of private schools in Island and Bulgaria.

In general, we can conclude that the total intensity of additional instruction has been increasing across the countries throughout the cycles and the intensity is higher in countries like Peru, Korea and Thailand where schools are, also, more heterogenous, and lower in countries such as Belgium, Deutschland, Denmark, and others depending on the year.

4.2 Relationship between variables of interest

In order to analyze the relationship between variables, as we have a considerable number of them, we are going to choose the most important ones in terms of correlations with the others. We firstly verified if there were any differences in the direction of correlations between years and, also, per type of school, which did not happen. Therefore, we performed an overall correlation presented in Figure 12.

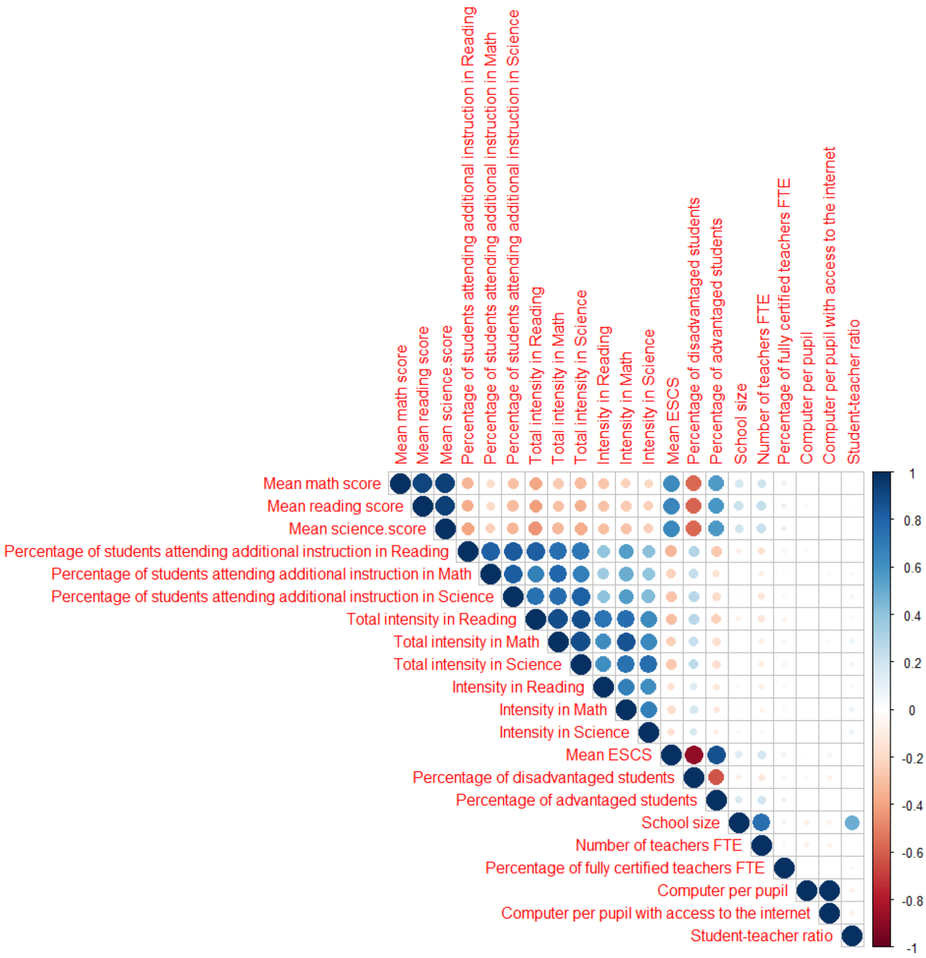

Figure 12: Overall correlation plot.

By looking at the correlation plot in Figure 12, we can see that total intensity is the variable regarding additional instruction that presents the strongest correlations with the scores and, also, the socioeconomic background. This, once again, proves that total intensity is the most adequate variable to represent the additional instruction phenomenon.

To represent the socioeconomic background, we considered not only the mean ESCS but also de percentage of advantaged and disadvantaged students. All the variables have similar correlation with scores. For simplicity, the variable that will be used is the percentage of advantaged students.

In general, we can conclude by Figure 12 that there is a strong negative correlation between additional instruction and the PISA scores in all the subjects. Additional instruction in reading is the variable that presents the strongest correlation with all the subject scores, and not only with reading. Therefore, we can observe that, probably, the nature of additional instruction is mainly remedial because schools where the phenomenon is more intense tend to perform worse in PISA.

Regarding the school variables, we can easily see three possible influential variables: the school size; the number of teachers full-time equivalent, which is strongly correlated to the former; and the number of computers per pupil. Variables such as the percentage of fully certified teachers are, in theory, important to explain the resources of the school in terms of teachers' qualifications. However, our dataset shows that there is no relationship between this variable and all the others. The student-teacher ratio is also a variable that does not appear to be relevant for this purpose.

Therefore, we will proceed with the number of full-time equivalent teachers, because it presents stronger correlations with the other variables than the school size, and the number of computers per pupil as a proxy for the schools' resources.

4.2.1 Relationship between additional instruction, socioeconomic background and the PISA scores

Correlation plots are very useful to understand at what level variables impact each other. In the previous section we saw that, overall, there was a negative correlation between the percentage of advantaged students and the mean hours of additional instruction. This seems to suggest that socioeconomic deprivation implies, in general, more hours of additional instruction, which leads us to think that additional instruction is offered at school and not outside the school – since tutoring classes are usually expensive and only more advantaged students have access to it. It is interesting to understand whereas this negative correlation is maintained in each country or not. Therefore, we performed correlation plots per country and found three different types of countries: the first type is represented in Figure 13 with the case of Denmark; the second type is represented in Figure 14 with the case of Hong Kong; and the third type is represented in Figure 15 with the case of South Korea.

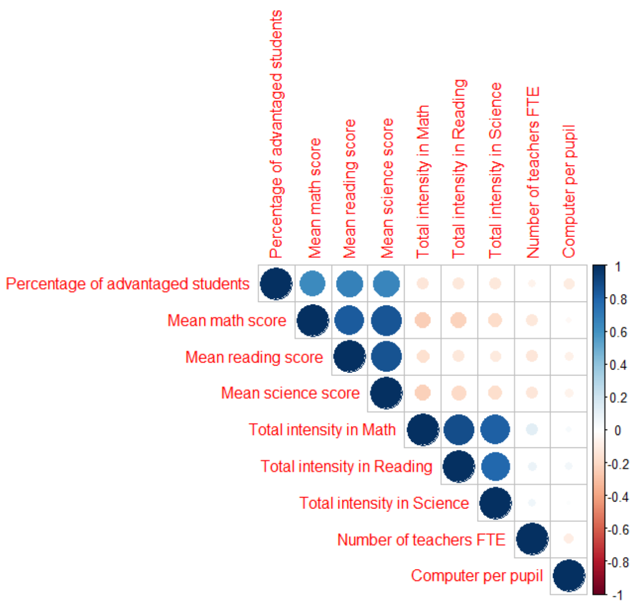

Figure 13: Correlation plot for Denmark.

Despite the strength of the correlation, there are countries where, for all subjects, the total intensity of additional learning is negatively correlated with the percentage of advantaged students. Some examples are Bulgaria, Denmark, Deutschland, Great Britain, Hungary, Island, Italy, Latvia, Peru and Slovakia.

We show in Figure 13 the correlation plot for Denmark, where we can see the negative correlation between the total intensity of additional instruction and the percentage of advantaged students. This goes in line with the "Economic model" theory exposed in Byun et al. (2018) that states that "students in poorer societies will be more likely to use shadow education than their counterparts in wealthier societies" (Byun et al., 2018, p. 9) due to the lower quality of education in those countries.

We may conclude that, in such countries, students that attend additional instruction do it mainly in schools, because in schools with a low percentage of advantaged students the intensity of additional instruction tends to be higher, and those students do not have the capacity to pay for expensive private tutoring classes.

Note that for these countries, where the correlation with the percentage of advantaged students is negative, the correlation with the scores is also negative. This puts in evidence the remedial nature of additional instruction.

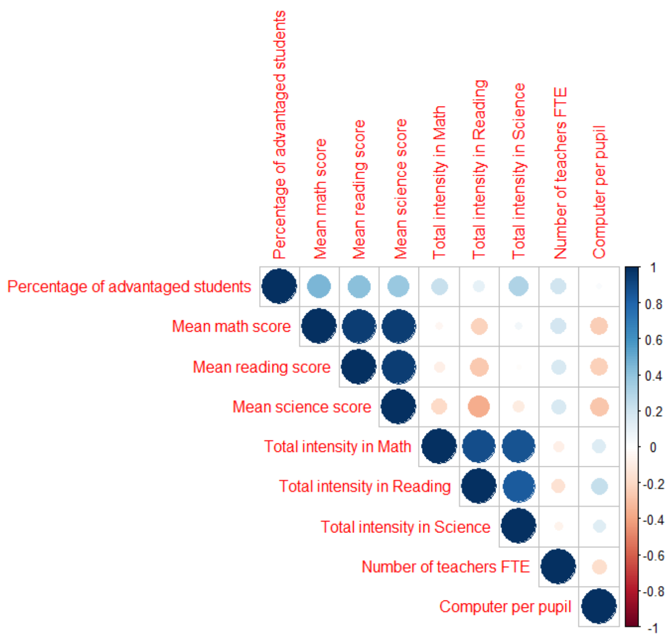

Figure 14: Correlation plot for Hong Kong.

There are other countries where the incidence and intensity of additional learning, irrespective of the subject, are positively correlated with the percentage of advantaged students as we can see in Figure 14. Such countries are Hong Kong, Thailand, and, in Europe, Greece is a country that approximates this group pattern, except in 2015.

We show the correlation plot for Hong Kong in Figure 14, where we can see the positive correlation between the total intensity of additional instruction and the percentage of advantaged students, especially in math and science.

Hence, we may conclude that as the intensity of additional instruction tends to be higher in schools with a higher percentage of advantaged students, probably, this additional instruction is not necessarily given at school but has more private tutoring character. For those countries, the correlation with the PISA scores is negative, so the additional instruction also tends to be remedial.

Korea is represented in Figure 15 where we can see that the type of additional instruction is more success related, as the correlation between the PISA scores and additional instruction is positive. In addition, the relationship between additional instruction and the socioeconomic background is positive.

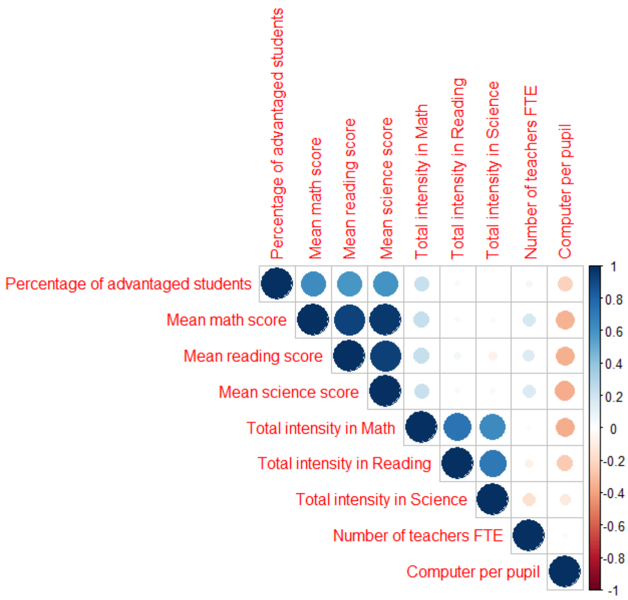

Figure 15: Correlation plot for Korea.

Note that this goes in line with the conclusions obtained by Baker et al. (2001) and Byun et al. (2018) that considered Korea as having an additional instruction leaning towards enrichment, whereas Hong Kong and Thailand, despite having an enrichment component, have a strong remedial character.

Then, there are countries where the correlation between additional instruction and the percentage of advantaged students depends on the subject and countries where, over the years, the pattern changes, such as Belgium, Lithuania and others. Note that in some countries the relations are very weak and there is not much connection between the shadow education phenomenon, the socioeconomic background of the schools and their PISA scores.

In Portugal, represented in Figure 16, through the data of 2009 and 2012, we can conclude that it does not have a clear pattern when it comes to additional instruction. However, it is more similar to the group where the correlation between additional instruction and the percentage of advantaged students is negative. The correlation between additional instruction and the PISA scores is essentially negative, denoting a remedial character of additional instruction in Portugal, especially in reading.

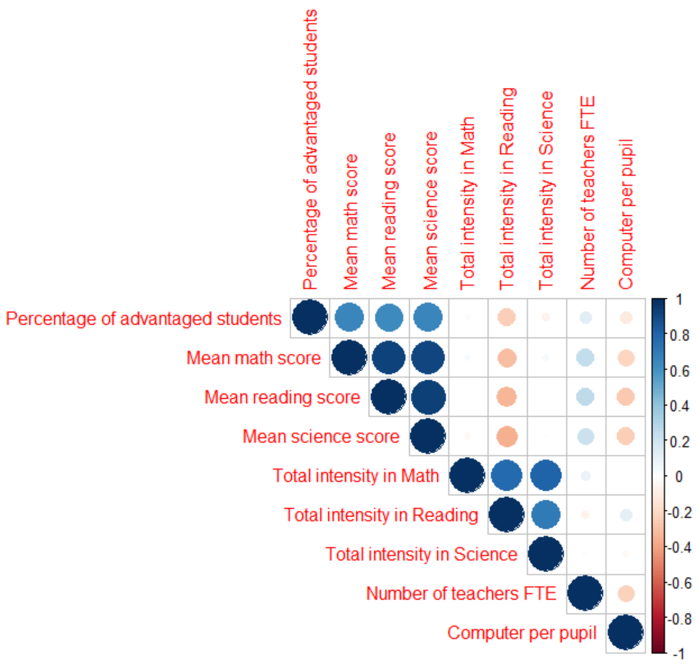

Figure 16: Correlation plot for Portugal.

Chapter 4
Empirical Analysis: Performance Scores

Data Envelopment Analysis (detailed in Chapter 2) was used to compare the schools across various countries on their ability to produce PISA scores. Our data is at the school-level and we are going to apply an output-oriented model considering variable returns to scale.

The output variables are the mean PISA scores in math, science and reading. Whereas the input variables are the total intensity of additional instruction in each subject in schools, their percentage of advantaged students (representing the socioeconomic background of the schools), the number of teachers full-time equivalent and the number of computers per pupil (representing the schools' resources).

The model is output-oriented because PISA scores are more manageable than the inputs, for example the percentage of advantaged students. Besides that, in our empirical application, we consider variable returns to scale since we cannot consider that when inputs increase, a proportional increase in the scores will happen because they are bounded.

Consequently, our frontier is going to be populated by the schools that achieve the best results in PISA with the lowest values of: total intensity of additional instruction; percentage of advantaged students; and school's resources (number of teachers full-time equivalent and computers per pupil). Schools that are not on the frontier present scores below 100%, that correspond to the percentage that observed outputs are of the target outputs.

Our study distinguishes from others using the DEA model due to the addition of an input, which is additional instruction, our variable of interest in this study. Note that we decided to avoid missing values in a DEA model, resulting in a reduced sample of 11 529 schools.

Running the DEA output-oriented model results in 200 schools on the frontier, which can achieve the highest results in PISA with the lowest resources. Firstly, it is important to verify whether these schools may be outliers, that is, sets of units that do not represent attainable performances. In order to check for outliers, we ran a super-efficiency model and realized that the maximum super-efficiency obtained was 135%, which means that no school appears to be a significant outlier (as the super-efficiency is not very distant from 100%).

In Table 7, we present the frequency table for the various scores obtained in the whole sample (when performance is assessed in relation to the meta-frontier). Most schools have a performance score between 77% and 83%, and the mean score is 79%. The school with the worst performance has a score of 46%.

Table 7: Distribution of the performance scores.

Efficiency Range	Absolute frequency	Relative frequency
Eff = 1	200	1,74%
0,91 < Eff < 1	1235	10,71%
0,83 < Eff ≤ 0,91	2801	24,30%
0,77 < Eff ≤ 0,83	3328	28,87%
0,67 < Eff ≤ 0,77	3181	27,59%
0,5 < Eff ≤ 0,67	775	6,72%
0,2 < Eff ≤ 0,5	9	0,078%

In terms of characteristics of the schools at the frontier, we can highlight that: most of them are public schools (even tough almost 31% are private schools, which is more than the percentage of schools represented on the total sample

(22%)); there are only sixteen schools at the frontier that represent observed values of 2015, and efficient schools with observed values from 2009 predominate (almost 53% of schools on the frontier belong to 2009, and the other years are underrepresented); when it comes to the countries, Australia, Belgium, Deutschland, Iceland, Italy, Hong Kong, Hungary, Korea, Peru, Poland and Slovenia predominate as frontier countries. Countries like Belgium, Deutschland, Hong Kong, Iceland, Korea, Poland and Slovenia are overrepresented on the frontier in comparison to their representativity in the sample. Note that Lithuania and Portugal do not have any school on the frontier.

1. Top quartile schools and bottom quartile schools

In order to understand the distribution of the schools according to the performance scores, we show in Table 8 the characteristics that differentiate schools that perform well (top quartile - where the efficient units are included) from schools that present a poor performance (bottom quartile).

Table 8: Performance of the efficient, top quartile and bottom quartile units.

	Efficient units	Top quartile units	Bottom quartile units
Observations	200	2882	2882
Mean performance score	1	0,92	0,69
Mean math score	563,80	552,6	423,1
Mean reading score	550,60	552,1	416,6
Mean science score	565,10	557,8	426,2
Total intensity of math additional instruction	0,83	1,12	2,20
Total intensity of reading additional instruction	0,44	0,62	1,99
Total intensity of science additional instruction	0,53	0,66	1,85

Percentage of advantaged students	0,33	0,36	0,16
Number of teachers FTE	33,35	53,25	56,24
Computers per pupil	0,68	0,63	0,83

For all the subjects the PISA scores of schools in the top performance quartile are above those of schools in the bottom quartile (Table 8) and the heterogeneity within the groups is very similar as we can see in Figure 17.

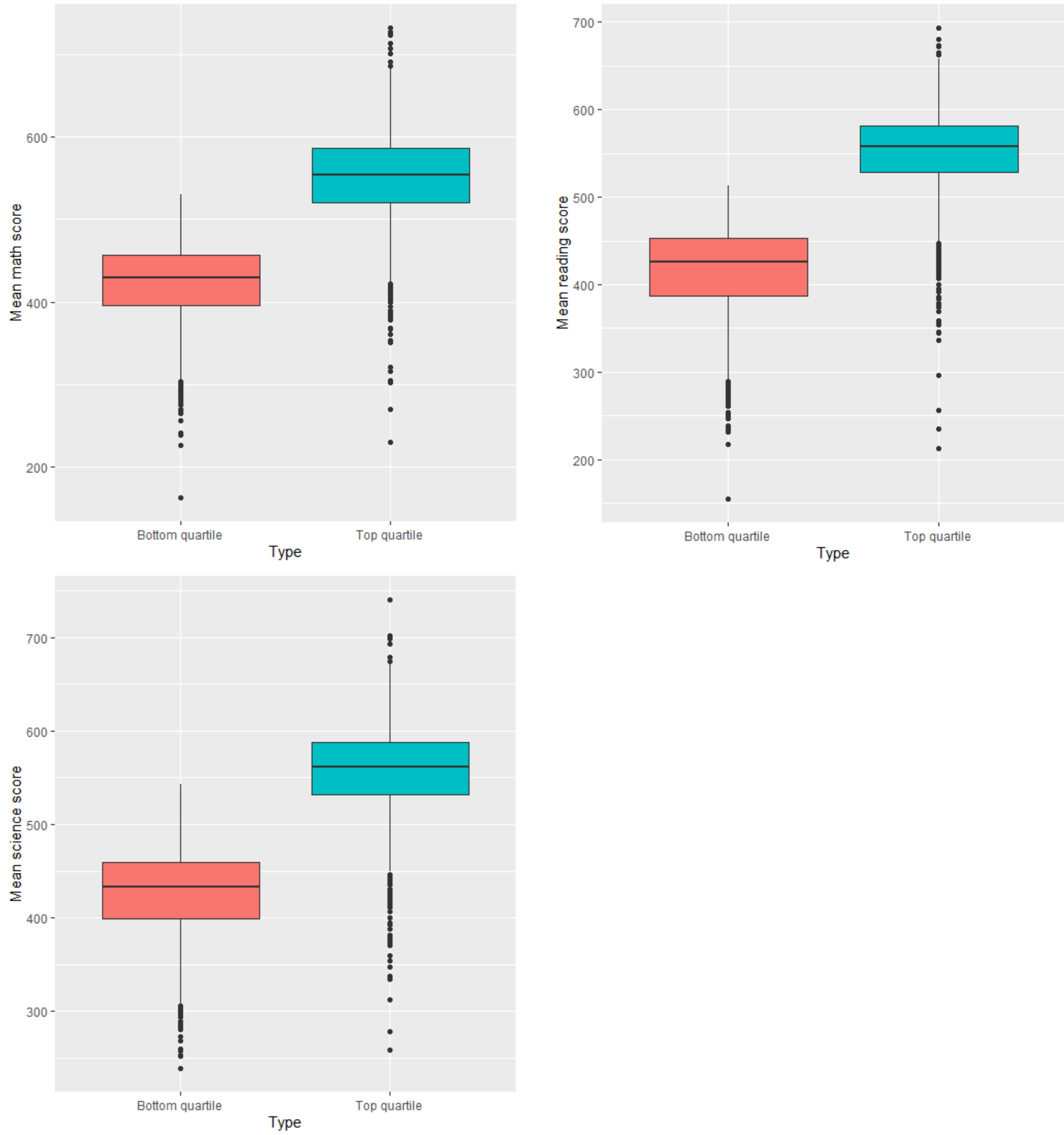

Figure 17: Boxplots of the distribution of outputs.

When it comes to the inputs related to additional instruction, our variable of interest, the intensity is similar between reading and science, and always higher for math. The intensity of additional instruction is, on average, higher for schools of the bottom quartile of performance, as it can be seen in Table 8. Also, we can see in Figure 18 that the dispersion between the bottom quartile schools is higher than what is verified on the top quartile schools.

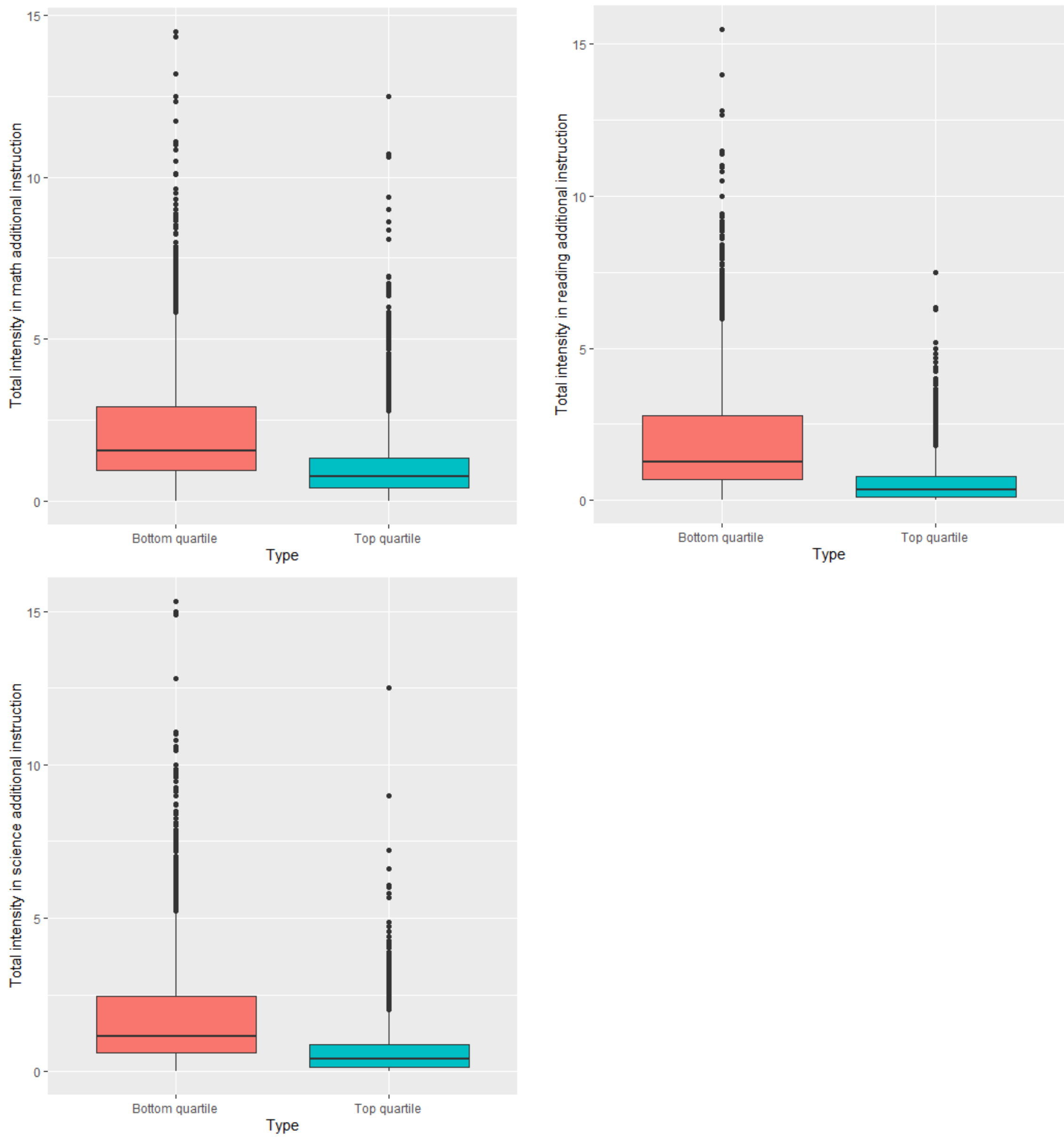

Figure 18: Boxplots of the distribution of additional instruction.

When it comes to the socioeconomic background of the students, top quartile schools have, on average, a higher percentage of students from advantaged socioeconomic backgrounds, even though they attend less additional instruction (see Table 8). This once again reenforces the idea that additional instruction is given essentially by the schools to students with lower performance (complemented with the information in the output section). However, the schools from the top quartile have more heterogenous backgrounds than the schools on the bottom, as Figure 19 puts in evidence.

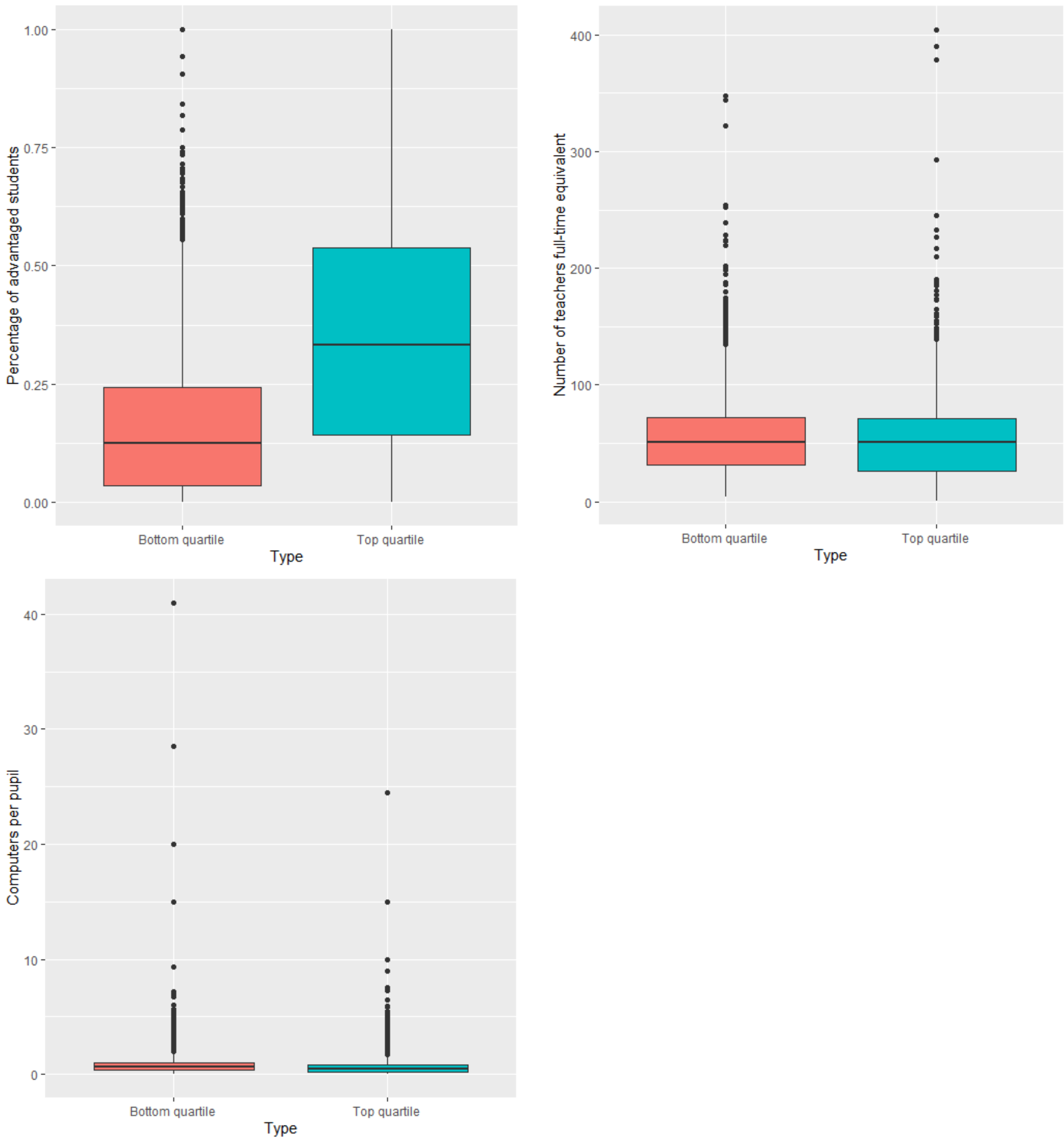

Figure 19: Boxplots of the distribution of other inputs.

As for schools' resources, we have the number of teachers full-time equivalent whose distribution is very similar between the two groups (see Figure 19). Regarding the number of computers per pupil, the top quartile has a slightly lower number of computers per pupil than the bottom (see Table 8).

We can conclude that the main factors contributing to the higher performance scores obtained by the top quartile schools are the scores in PISA and the low intensity of additional instruction when compared to the schools on the bottom quartile, which is confirmed by Figure 20. This is so because bottom and top schools are not very different when it comes to the other inputs.

Figure 20 shows the correlations between the variables included in the DEA analysis and the final scores, in an attempt to understand which of these variables contributed more to the performance scores.

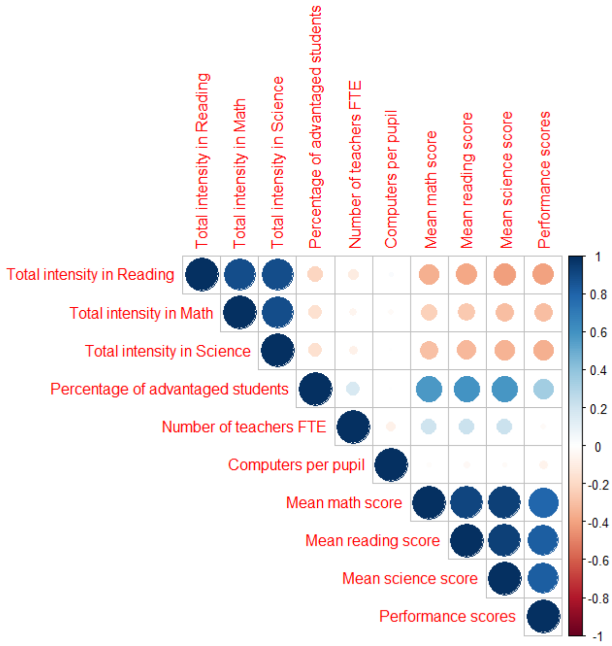

Figure 20: Correlation plot of the outputs, inputs and performance scores.

There is a non-expected positive correlation between the percentage of advantaged students and the performance scores in Figure 20 that should be object of further research. Indeed, the expectation was for a negative correlation of inputs and a positive correlation of outputs with performance scores.

2. School performance overtime

If we consider a meta-frontier, i.e., all the years aggregated in only one frontier, we can see through the first column of Table 9 that, from 2009 to 2015, the performance scores have been decreasing. To understand this evolution in depth a meta-Malmquist Index was computed. The approach used was the circular meta-Malmquist index applied by Portela et al. (2011), and the results are in Table 9.

Table 9: Meta-Malmquist index results.

Year	**Average meta-efficiency (%)**	**Average within-year efficiency (%)**	**Average technological frontier gap (%)**
2009	81,82	83,80	97,64
2012	79,87	81,00	98,62
2015	76,42	81,96	93,24
Period	**Malmquist index**	**Efficiency change**	**Technological change**
2009/2012	0,9762	0,9666	1,0100
2012/2015	0,9568	1,0119	0,9454

Even though we are evaluating a short period of time, we can conclude that there is no trend when it comes to the within-year efficiency or the technological gap. However, the average meta-efficiency seems to be decreasing (see Table 9), which means that, from 2009 to 2015, schools were moving away from the meta-frontier, especially since 2012.

From 2009 to 2012, the main reason for the decrease of the productivity was the decrease on the efficiency/performance of the schools. Whereas from 2012 to 2015, the main driver of the decrease was, supposedly, a negative technological change, that could not be compensated by an almost stable efficiency, as we can see in Table 9.

Trying to evaluate more deeply this negative trend, we looked at the evolution of the outputs and inputs overtime and concluded that this decrease on the schools' performance has to do, mainly, with the reduction on the mean PISA scores in the three subjects, but also with the increase on the intensity of additional instruction in those subjects.

The frontier seems to be regressing, but there are no reasons to suspect that the mean PISA scores obtained by the students in 2009 could not be attained in 2015 (there is a "superior limit", so the possibility of increases on the scores is smaller each year, but they can remain "stable"). As a result, in the analysis that follows, we will ignore the technological regress of the frontier in 2015 because there was no observable technological change implying that students could learn less in that year than the preceding years. For example, the advent of online teaching with the pandemic could have been considered such a technological change that prevented the comparability in attainments between years. But no such thing happened in 2015 to the authors' knowledge.

3. Performance scores across countries

As our sample consists of many countries, we decided on evaluating whether there are considerable differences between countries when it comes to their performance. Through Figure 21, that presents the distribution of the performance scores per country, we can see that the differences between some the countries are substantial.

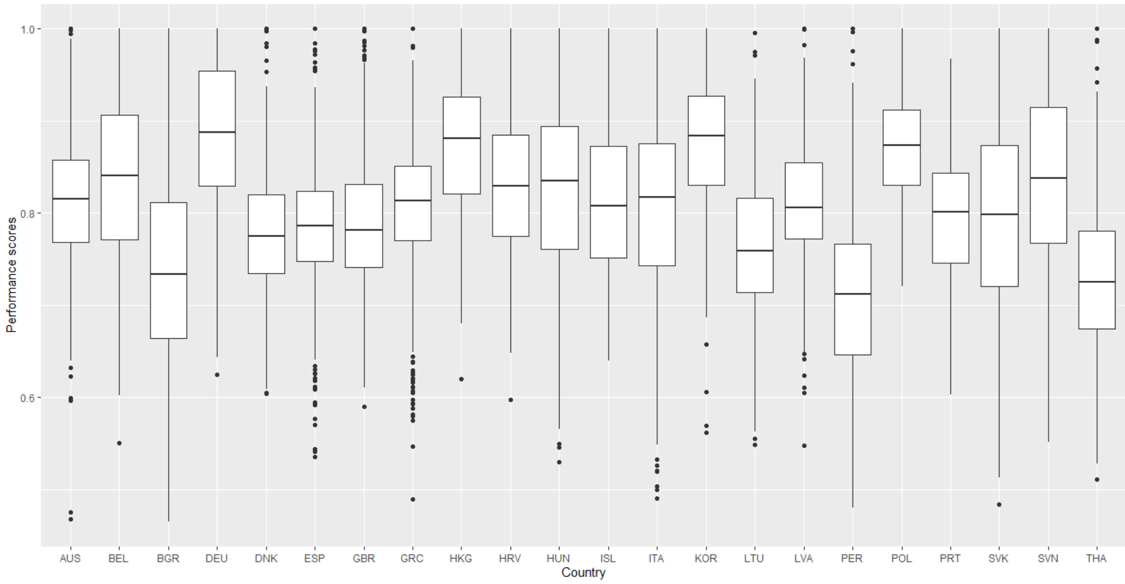

Figure 21: Boxplots of the performance scores per country.

Countries such as Bulgaria, Peru and Thailand present the lowest efficiency scores on average, which means that, for the same level of inputs, schools in these countries attain lower PISA scores than other schools on the sample. Looking at Table 10, where we have the mean values per country, these countries are, simultaneously, countries where the total intensity of additional instruction is considerable and the mean scores in all the subjects are the lowest, being these the main apparent reasons for their low performance scores.

Peru is the only country where the mean scores in all the subjects are below 400, and 32% of all the schools that have the mean scores below 400 are from Peru, which means there is an over-representation because schools from Peru are only 3,5% of the schools on our sample. When it comes to Thailand, their grades are slightly above those from Peru, but 15% of the schools with the mean scores below 400 are from Thailand and schools from Thailand are only 4% of the total. Together with Bulgaria, these three countries have a representation of approximately 53% on the schools with all PISA scores below 400.

Note that Peru and Thailand have more than 75% of schools with an efficiency lower than 80% (see Figure 21).

Table 10: Mean output and input values per country.

	Efficiency score	Mean math score	Mean reading score	Mean science score	Additional instruction in math	Additional instruction in reading	Additional instruction in science	Percentage of advantaged students	Number of teachers FTE	Computer per pupil
AUS	0,810	503,45	509,69	520,03	0,97	0,83	0,74	0,37	72,29	1,37
BEL	0,829	531,86	523,12	524,03	0,79	0,51	0,51	0,39	79,99	0,72
BGR	0,733	454,73	456,35	464,37	1,78	1,66	1,83	0,29	52,52	0,51
DEU	0,879	539,19	527,73	549,47	0,81	0,48	0,40	0,35	47,33	0,51
DNK	0,778	493,22	488,63	488,78	0,99	1,10	0,62	0,44	40,62	0,99
ESP	0,783	496,56	493,75	501,59	1,24	0,56	0,77	0,36	59,20	0,70
GBR	0,787	494,45	500,35	515,37	1,07	0,90	1,03	0,32	68,07	0,97
GRC	0,795	461,76	481,18	468,06	2,64	1,68	2,28	0,30	29,52	0,23
HKG	0,868	559,58	540,20	548,38	1,72	1,10	1,21	0,11	62,93	0,71
HRV	0,824	475,65	494,10	496,27	1,07	0,56	0,66	0,21	48,66	0,33
HUN	0,814	483,74	488,44	495,92	0,92	0,63	0,64	0,23	47,66	0,66
ISL	0,814	496,37	493,53	487,42	0,75	0,50	0,38	0,47	28,52	1,29
ITA	0,802	493,69	495,63	500,50	1,29	1,03	0,80	0,24	69,74	0,58
KOR	0,870	544,04	532,39	532,67	3,48	2,08	1,64	0,17	64,53	0,36
LTU	0,759	472,55	465,09	478,46	1,79	1,61	1,51	0,24	46,35	0,91
LVA	0,809	485,15	486,41	495,54	1,71	1,24	1,20	0,13	44,43	1,51
PER	0,707	379,61	390,08	388,66	3,26	2,89	2,55	0,10	35,10	0,48
POL	0,873	513,62	515,61	518,41	1,31	1,05	1,18	0,21	31,56	0,44
PRT	0,791	488,67	491,43	493,79	1,11	0,70	0,48	0,21	103,7	0,55
SVK	0,789	487,74	469,04	476,49	1,90	1,81	1,44	0,21	31,10	0,83
SVN	0,833	493,42	478,59	502,74	1,37	0,96	0,97	0,25	32,30	0,70
THA	0,724	435,91	432,92	443,52	3,21	2,27	3,24	0,13	78,81	0,53

On the other hand, Deutschland, Hong Kong, Korea and Poland are the countries that present the highest performance scores, having more than 75% of the schools with performance scores above 80% (see Figure 21). This happens even though Hong Kong, Korea and Poland present an intensity of additional instruction above the mean (see Table 10), something that, in practice, lowers their scores. If we take a closer look at the inputs and outputs for these three countries, we can see that, in the case of Hong Kong, its performance score is mainly due to the high scores the country achieves (the highest on the sample except in science) and, also, the low percentage of advantaged students in the schools.

In Korea, the intensity of additional instruction is higher than in Hong Kong and the scores are slightly lower, but the number of computers per pupil is lower. In Poland, with lower inputs they can still produce high PISA scores.

Deutschland has the highest performance of all countries because, despite the considerable percentage of advantaged students, they use few inputs to produce considerable high PISA scores.

These performance scores can be decomposed so that we can understand the main drivers of the inefficiencies on the countries and, for that, we used once again the circular meta-Malmquist index applied by Portela et al. (2011) with assessments within each country to understand country effects or country gaps (the distance between the country frontier and the meta-frontier). The results are presented in Figure 22.

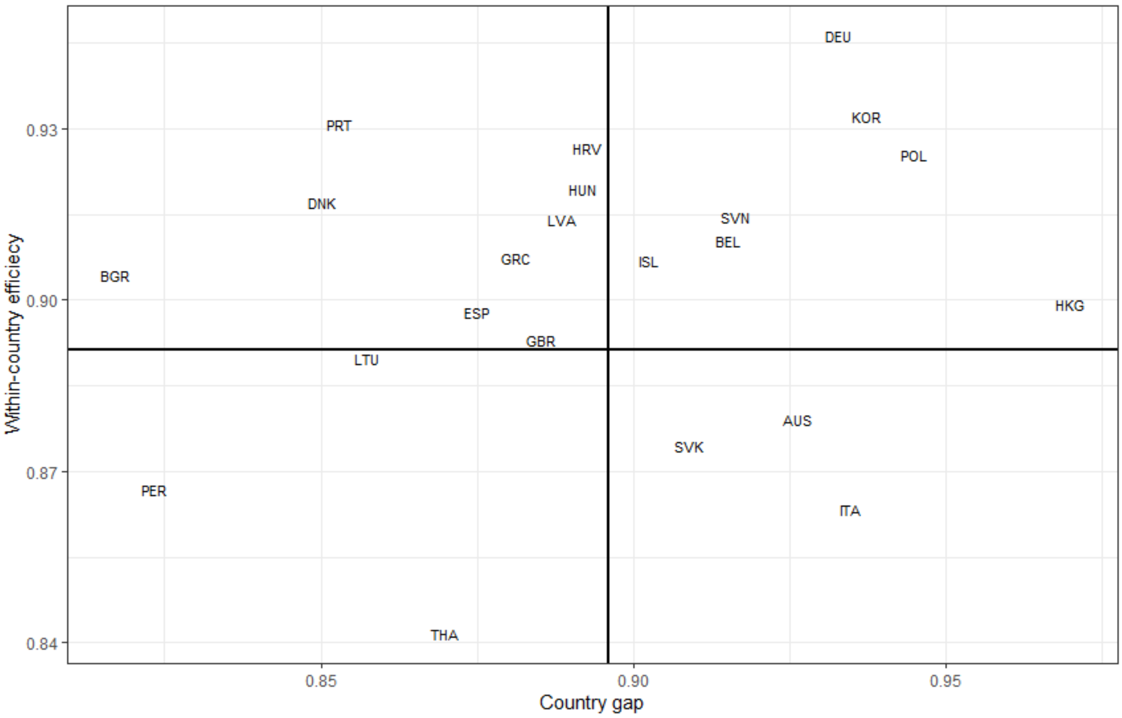

Figure 22: Matrix of the countries according to their within-country efficiency and country gap.

The average meta-efficiency can be decomposed on the average within-country efficiency and the average country frontier gap, as referred in Chapter

2. The within-country efficiency can be seen as an indicator of the homogeneity of schools in a given country. The smaller it is, the further away schools are from the country frontier. However, there is also another type of inefficiency that has not to do with the schools themselves, but rather with the country that was not able to guarantee that schools could achieve the same PISA scores as schools in other countries did with the same level of inputs, i.e., the educational system does not seem to be leveraging the same results as in other countries.

Australia, Italy, Lithuania, Peru, Slovakia and Thailand present the lowest within-country efficiencies which show that the schools within those countries are very heterogenous when it comes to their performance, having several schools performing poorly when compared to others. Whereas in Croatia, Deutschland, Korea and Portugal, for example, the schools are more concentrated around the frontier, presenting in principle better and more balanced performances when compared between each other (see Figure 22).

The country gap represents differences between countries motivated by their education system, and not an inefficiency on the production process of the schools themselves. On the sample, the countries that show the lowest country frontier gap (and therefore the largest average distance between the country frontier and the meta-frontier) are Bulgaria, Peru, Denmark and Portugal as it can be seen in Figure 22. Their schools present a lower performance just because they belong to these countries, whereas schools in Hong Kong, Korea and Poland can achieve considerably better results with the same inputs.

Note that in Korea we have a situation that is mirroring the situation of Portugal. Both are the countries where the schools are more homogenous (together with Deutschland). This means that Portuguese schools, when compared between themselves, are as good as Korean schools when being compared with each other. However, when comparing between countries, the situation is absolutely the opposite (see the country gap axis in Figure 22). Even though Portuguese schools are very homogenous within the country, when comparing with other countries' performances, they are not that good. Whereas schools in Korea are all good performers within their country and, when compared with schools from other countries, are also within the best.

This means that despite being good performers in Portugal, there is a certain factor, that has to do with the national educational system, that does not allow them to achieve such high PISA scores as schools in Korea, or other countries, with the same level of inputs.

4. Performance scores per type of school

The type of school can be a factor affecting the performance of schools and, therefore, we investigated the extent to which private and public schools present different performance scores through Figure 23.

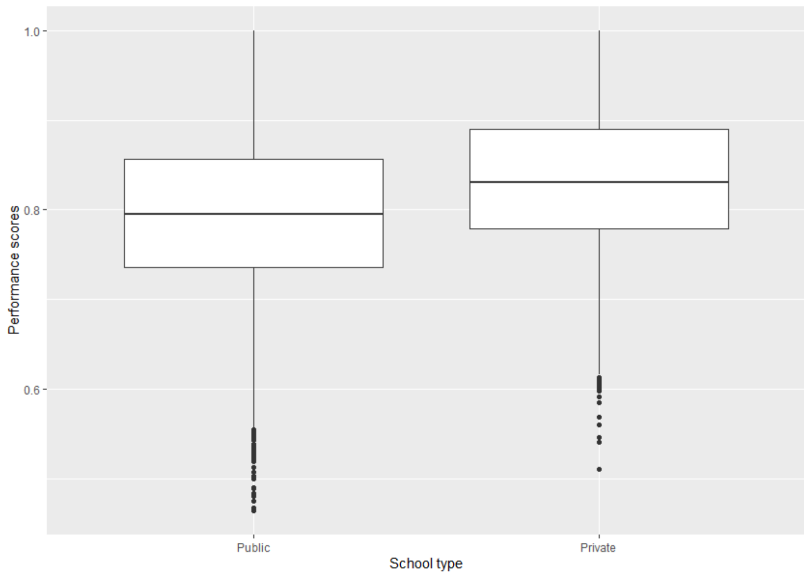

Figure 23: Boxplots of the efficiency scores per type of school.

It is perceptible that, evaluating how the efficiencies are distributed between private and public schools, the formers are, on average, slightly more efficient (see Figure 23). To make sure this difference is statistically significant we performed a T-test which had a p-value very close to 0, and, therefore, confirmed our suspicion.

With the same inputs, private schools can achieve considerably higher PISA scores. There is then a difference between the types of schools that may be attributable to different teaching approaches.

Thus, the type of school should be, in theory, considered as a factor. Yet, note that this would imply a deep level of disaggregation that is not possible in this sample. This is due to some countries, such as Iceland and Latvia, only having 4 to 5 private schools in the sample, so it would not allow a reliable frontier (i.e., a frontier representing the private schools of these countries).

Taking this into account we are not considering the type of school as a factor, as it would not add much value to our study for the reason above mentioned.

5. Differences and similarities between countries

As we have seen before, the performance scores of the countries are mainly dependent on the students' achievement in PISA, and the additional instruction on the three subjects, even though other inputs also play a role.

Therefore, we selected those variables and the efficiency scores aggregated at the country-level to try to define different educational groups in Figure 24, together with the information obtained in Figure 13, Figure 14 and Figure 15. We selected only the matrix with math because the situation is similar for the other subjects, except in science where Korea presents a considerable different pattern (locates on the quadrant with low intensity and high performance).

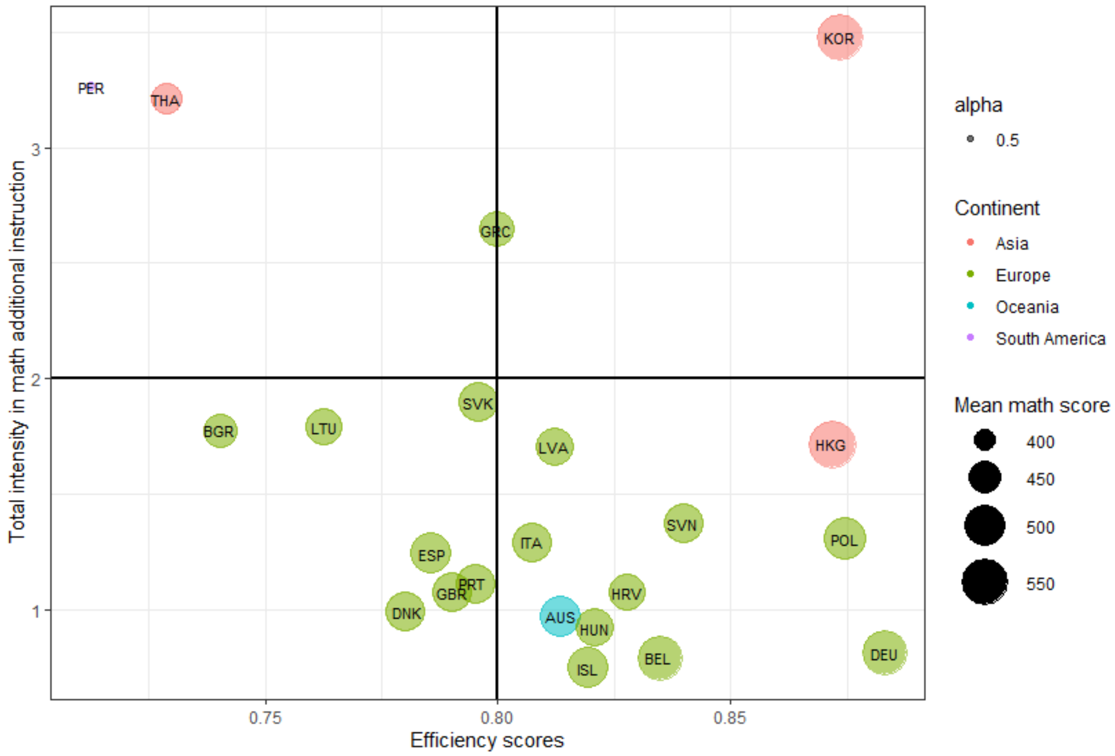

Figure 24: Matrix of the countries according to additional instruction and efficiency scores.

On the quadrant with low performance and high additional instruction intensity in Figure 24, we have countries that cannot reflect their effort in additional instruction into the PISA scores, being both the reason for such low performance. These countries have an additional instruction that may be essentially remedial in nature as the correlation between the PISA scores and additional instruction is negative (see Figure 14).

On the other side, we have countries with high performance and high intensity of additional instruction that, in this case, is only Korea (see Figure 24). Korea is a country where the shadow education phenomenon is intense, however, that investment can be converted into considerably high PISA scores resulting in a strong performance. It is the only country on the sample that shows a positive correlation between the PISA scores and additional instruction, evidencing an enrichment type of additional instruction (see Figure 15).

Note that Greece is an in-between situation in Figure 24. This can be due to the fact that additional instruction in Greece is intense and essentially remedial in nature, but it cannot be compared with Peru and Thailand because they achieve superior scores in PISA and perform considerably better, showing that additional instruction may be leveraging their grades.

In most countries additional instruction is not intense, taking less time from their students, as it is the case of Hong Kong, or is even residual in incidence, such as in Deutschland. However, some of the countries in this situation present strong performances and others poor performances when it comes to achieving high PISA scores with the same level of inputs.

The most populated quadrant is characterized by high performance with low intensity of additional instruction (see Figure 24), where it seems that low intensity of additional instruction is one of the main factors contributing to the good performance of the countries. The mean scores in PISA seem to be within the average in these countries. Note that the additional instruction is essentially remedial, as all countries in this quadrant present a negative correlation between the mean PISA scores and intensity of additional instruction (see Figure 13).

The other quadrant in Figure 24 is characterized by low performances with low additional instruction intensity, which means that their poor performance is not related to additional instruction but to other factors. The nature of the additional instruction that exists is also remedial (see Figure 13).

In order to understand the characteristics of the countries in each quadrant in terms of the level of inputs and outputs, one country was selected per quadrant, except the low performance with low intensity quadrant where both Spain and Portugal were selected. Two radar charts with normalized values (values normalized by the mean global value of the variable) were developed, one for the outputs - Figure 25 - and another for the inputs that affect the performance scores the most - Figure 26 -, and we can conclude that the higher the efficiency score, the higher the PISA scores.

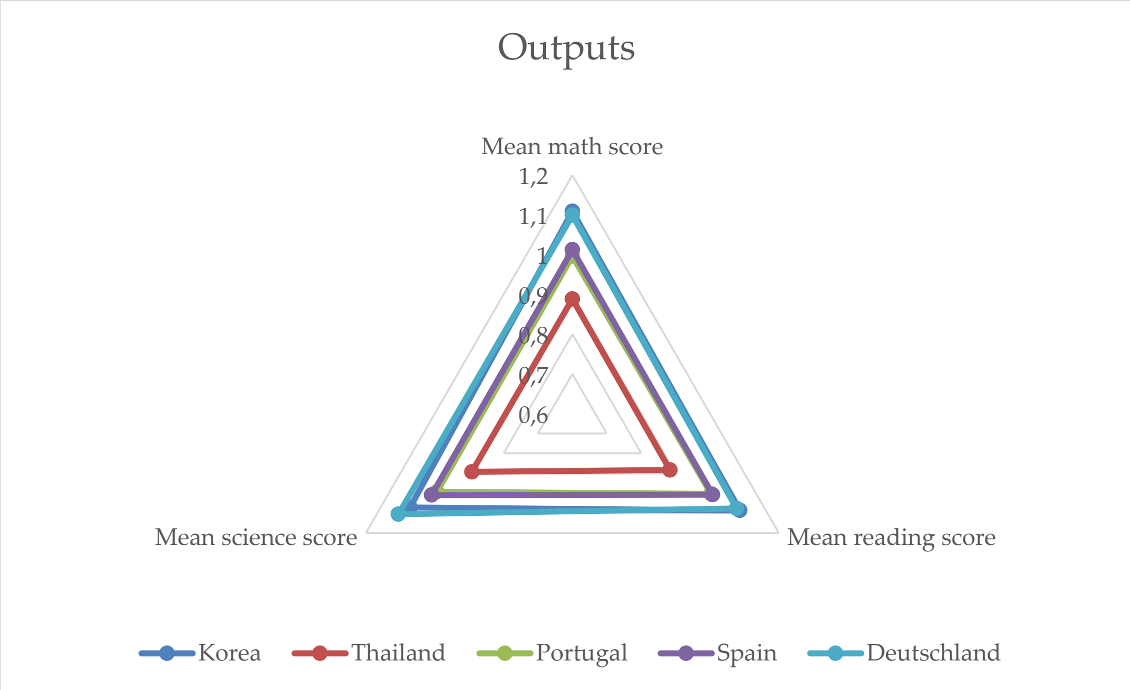

Figure 25: Radar chart of the normalized outputs.

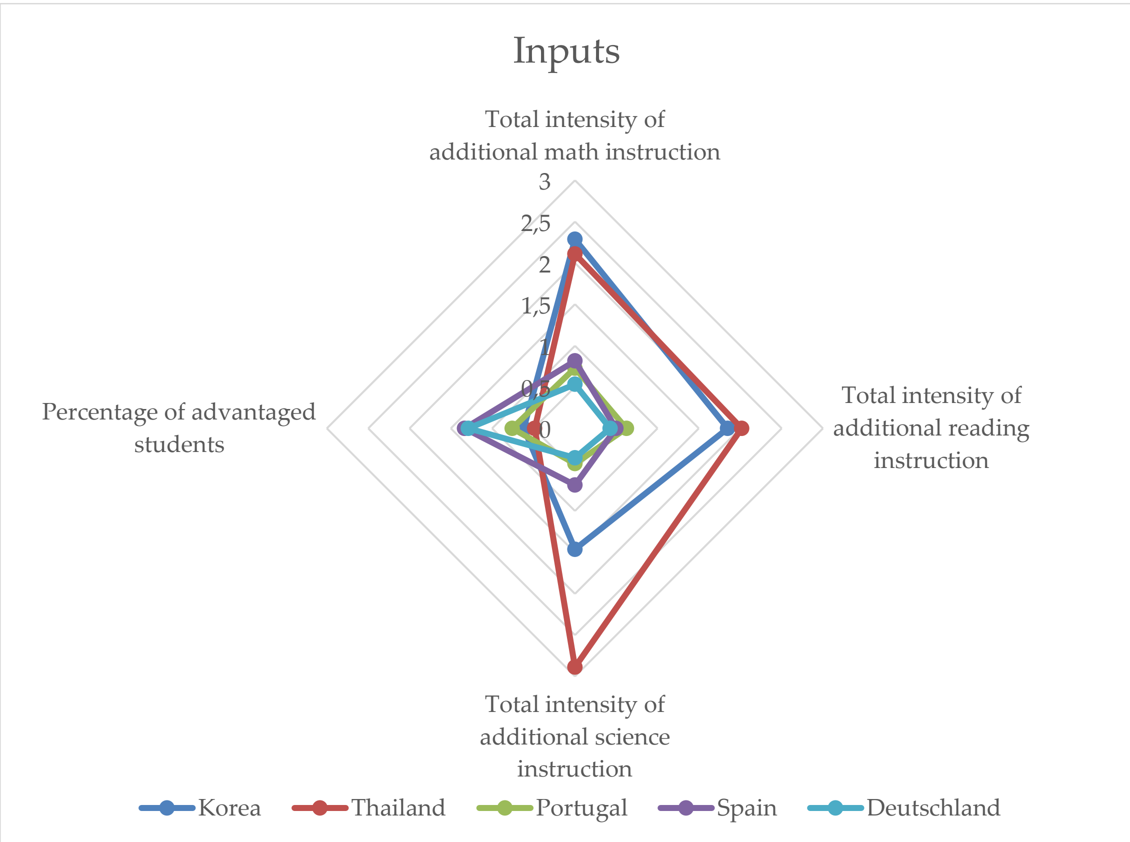

Figure 26: Radar chart of the normalized inputs.

Also, we can see that in the quadrant where there are low performances with low intensity, which is the case of Spain and Portugal, the main factor contributing to the low performance may be the considerable percentage of advantaged students (or one of the other inputs not represented in the graph), put in evidence in Figure 26.

Chapter 5
Conclusions and Future Research

This research provided implications for understanding the shadow schooling phenomenon, more specifically additional instruction, through a comparative study between many participants in PISA over the years, which is something not very explored in research until now.

1. Main conclusions

It is important to note, firstly, that additional instruction has been increasing from 2009 to 2015, whereas the scores obtained by the students in PISA have been decreasing, something that was not expected by any technological change.

As already put in evidence by the literature on the field, the additional instruction phenomenon is generalized at schools in East Asian countries and still residual in Europe, apart from Greece that clearly stands out when it comes to the incidence and intensity of additional instruction.

Even though it depends on the subjects, we were able to find countries where: both incidence and intensity were high, namely, East Asian countries and Greece; where the phenomenon was residual such as in Deutschland, Denmark, Iceland and Portugal included (despite having an incidence almost above the mean); and countries where the phenomenon is generalized but takes less time from the students, which is the case of Hong Kong.

One finding that has been controversial in literature, but found by many researchers, relates to the relationship between the incidence and intensity of additional instruction and the grades obtained by students. In this research, a negative correlation between the total intensity of additional instruction and the PISA scores was found, which means that schools, and, therefore, countries with high intensity of additional instruction tend to perform worse in PISA. It was expected that more additional instruction would be associated with better scores, except if the support was given only to the worst students and, then, be more intense in countries where the performances were worse (which does not seem to be the case as there are countries with a high incidence of additional instruction and good performances in PISA, simultaneously). A reason we may find for this has to do with the fact that additional instruction is normally used to reinforce school content. PISA, however, does not evaluate that content but rather the comprehension and reasoning abilities, critical thinking and other skills. Therefore, it may happen that time spent in reinforcement of school content penalizes time that could be used to promote other skills related to knowing how to use what we learn at school in real life.

This negative correlation also puts in evidence a probable remedial nature of additional instruction because additional instruction is more intense where the performance in PISA is worse. Korea is an exception presenting a positive correlation between these variables, meaning that in Korea, in general, PISA scores are higher in schools with higher intensity of additional instruction.

Despite DEA models being widely used to study the education sector, there are not studies, to the authors' knowledge, using DEA to study additional instruction and understand its impact on the education sector and on the efficiency of schools. One of the variables that impacts the efficiency of schools the most is, indeed, additional instruction. This input together with the PISA scores are the variables that essentially define the performance score of the schools, i.e., their capacity to maximize the PISA scores for the same level of additional instruction, socioeconomic background and resources.

Note that we concluded that the performance scores are not exclusively given to the performance of schools, but there is also a factor that influences their performance that is not directly in their control, which is the country they belong to. Schools in countries such as Hong Kong, Korea and Poland have education systems that potentiate the attainment of higher scores making them schools within the best, whereas schools in Bulgaria, Peru, Denmark and Portugal are penalized because they belong to a country that cannot leverage their results as the former countries can for the same level of inputs.

The benchmarking exercise allowed for the detection of different educational systems, specifically regarding the additional instruction variable, and resulted in information that should be taken into consideration by the sampled countries on future decisions regarding educational policy.

Countries such as Peru and Thailand are characterized by having a considerable intensity of additional instruction but very poor performance when compared to other countries' schools with the same level of inputs. This puts in evidence the fact that, in this case, additional instruction is not contributing positively to the PISA scores and, therefore, reduces the performance of the schools. In such countries, the role of additional instruction should be revisited by official entities to avoid the use of resources that do not produce, necessarily, better results.

Other countries, as opposing to Korea, present low intensity in additional instruction and high performance scores, which is the case of Deutschland. There are also countries that present low intensity but also low performance, which means their poor performance is not due to additional instruction but other factors. This is the case of Portugal where the driver of this inefficiency should be assessed.

2. Limitations and Further Research

In this study, there was an evident difficulty related to the format of the PISA questionnaires. Firstly, due to the limitations referred in Chapter 1 section 4, questions in PISA keep changing across cycles, attempting to improve their formulation, which makes it complicated to find years when questions are compatible and a unique variable regarding additional instruction can be used to map the evolution over time. This may be one reason why there are not many studies that follow a quantitative approach using more than one year of PISA.

Moreover, another limitation found regarding PISA was that the question we were able to collect from PISA asks about additional instruction, and not private tutoring in specific, which means that it can take place in the form of a private lesson, but also includes the extra support given to students at school. Therefore, our research was not able to fully assess the extension of shadow schooling, even though it contributes to the literature on this topic.

The sample used in the future should include more observations regarding private schools so that the role of the type of school could be considered without compromising the results, as the T-test proved the type of school to be important in explaining differences between performance scores.

References

Addi-Raccah, A., & Dana, O. (2015). Private tutoring intensity in schools: a comparison between high and low socio-economic schools. *International Studies in Sociology of Education*, *25*(3), 183–203. https://doi.org/10.1080/09620214.2015.1069719

Agasisti, T. (2013). The efficiency of Italian secondary schools and the potential role of competition: a data envelopment analysis using OECD-PISA2006 data. *Education Economics*, *21*(5), 520–544. https://doi.org/10.1080/09645292.2010.511840

Azevedo, S., & Neto-Mendes, A. (2009). As explicações (aulas particulares) no ensino secundário e superior em Portugal: apresentação de um estudo. *Revista Eletrônica de Educação*, *3*(2), 84–104. https://www.reveduc.ufscar.br

Baker, D. P., Akiba, M., LeTendre, G. K., & Wiseman, A. W. (2001). Worldwide Shadow Education: Outside-School Learning, Institutional Quality of Schooling, and Cross-National Mathematics Achievement. *Educational Evaluation and Policy Analysis*, *23*(1), 1–17. https://doi.org/10.3102/01623737023001001

Bento, A., & Ribeiro, M. (2013). The Phenomenon of Private Tutoring: Implications for Public Education. *Global Education Review*, *1*(5), 70–74.

Berberoğlu, G., & Tansel, A. (2014). Does Private Tutoring Increase Students' Academic Performance? Evidence from Turkey. *International Review of Education*, *60*(5), 683–701. https://doi.org/10.1007/s11159-014-9436-y

Bessent, A. M., & Bessent, E. W. (1980). Determining the Comparative Efficiency of Schools through Data Envelopment Analysis. *Educational Administration Quarterly*, *16*(2), 57–75. https://doi.org/10.1177/0013161X8001600207

Bray, M. (2017). Schooling and Its Supplements: Changing Global Patterns and Implications for Comparative Education. *Comparative Education Review*, *61*(3), 469–491. https://doi.org/10.1086/692709

Bray, M. (2020). Shadow Education in Europe: Growing Prevalence, Underlying Forces, and Policy Implications. *ECNU Review of Education*, 1–34. https://doi.org/10.1177/2096531119890142

Bray, M., & Kobakhidze, M. N. (2014). Measurement Issues in Research on Shadow Education: Challenges and Pitfalls Encountered in TIMSS and PISA. *Comparative Education Review*, *58*(4), 590–620. https://doi.org/10.1086/677907

Bray, M., Kobakhidze, M. N., & Suter, L. E. (2020). The Challenges of Measuring Outside-School-Time Educational Activities: Experiences and Lessons from the Programme for International Student Assessment (PISA). *Comparative Education Review*, *64*(1), 87–106. https://doi.org/10.1086/706776

Byun, S. (2014). Shadow Education and Academic Success in Republic of Korea. In H. Park & K. Kim (Eds.), *Korean Education in Changing Economic and Demographic Contexts* (Vol. 23, pp. 39–58). Springer, Singapore. https://doi.org/10.1007/978-981-4451-27-7_3

Byun, S., Chung, H. J., & Baker, D. P. (2018). Global Patterns of the Use of Shadow Education: Student, Family, and National Influences. In *Research in the Sociology of Education* (Vol. 20, pp. 71–105). https://doi.org/10.1108/S1479-353920180000020004

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating Program and

Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through. *Management Science*, *27*(6), 668–697. https://doi.org/10.1287/mnsc.27.6.668

Costa, J. A., Neto-Mendes, A., Ventura, A., & Azevedo, S. (2007). O fenômeno das explicações: aspectos da realidade portuguesa e do contexto global. *Ensaio: Avaliação e Políticas Públicas Em Educação*, *15*(57), 475–488. https://doi.org/10.1590/S0104-40362007000400002

Davies, S. (2004). School Choice by Default? Understanding the Demand for Private Tutoring in Canada. *American Journal of Education*, *110*(3), 233–255. https://doi.org/10.1086/383073

Guill, K., Lüdtke, O., & Schwanenberg, J. (2020). A two-level study of predictors of private tutoring attendance at the beginning of secondary schooling in Germany : The role of individual learning support in the classroom. *British Educational Research Journal*, *46*(2), 437–457. https://doi.org/10.1002/berj.3586

Ireson, J. (2004). Private Tutoring: how prevalent and effective is it? *London Review of Education*, *2*(2), 109–122. https://doi.org/10.1080/1474846042000229458

Javadi, Y., & Kazemirad, F. (2020). Worldwide Shadow Education Epidemic and Its Move toward Shadow Curriculum. *Journal of Language Teaching and Research*, *11*(2), 212–220. https://doi.org/10.17507/jltr.1102.09

Kim, J. H., & Park, D. (2010). The determinants of demand for private tutoring in South Korea. *Asia Pacific Education Review*, *11*(3), 411–421. https://doi.org/10.1007/s12564-009-9067-3

Kobakidze, M. N., & Suter, L. E. (2020). The global diversity of shadow education. *European Journal of Education*, *55*(3), 316–321. https://doi.org/10.1111/ejed.12411

Lee, C.-J., Park, H.-J., & Lee, H. (2009). Shadow Education Systems. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of Education Policy Research* (1st ed., pp. 901–919). Routledge. https://doi.org/10.4324/9780203880968.ch70

Liao, X., & Huang, X. (2018). Who Is More Likely to Participate in Private Tutoring and Does It Work ?: Evidence from PISA (2015). *ECNU Review of Education*, *1*(3), 69–95. https://doi.org/10.30926/ecnuroe2018010304

Liu, J., & Bray, M. (2017). Determinants of demand for private supplementary tutoring in China: findings from a national survey. *Education Economics*, *25*(2), 205–218. https://doi.org/10.1080/09645292.2016.1182623

Mancebón, M. J., Calero, J., Choi, Á., & Ximénez-De-Embún, D. P. (2012). The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006. *Journal of the Operational Research Society*, *63*(11), 1516–1533. https://doi.org/10.1057/jors.2011.156

Manzon, M., & Areepattamannil, S. (2014). Shadow educations: mapping the global discourse. *Asia Pacific Journal of Education*, *34*(4), 389–402. https://doi.org/10.1080/02188791.2014.969194

Melese, W., & Abebe, M. (2017). Demand and Supply of Supplementary Private Tutoring in Upper Primary Schools of Ethiopia. *International Online Journal of Educational Sciences*, *9*(3), 629–640. https://doi.org/10.15345/iojes.2017.03.004

OECD. (2008a). *School Questionnaire for Pisa 2009*. OECD Publishing. https://www.oecd.org/pisa/

OECD. (2008b). *Student Questionnaire for PISA 2009*. OECD Publishing. https://www.oecd.org/pisa/

OECD. (2011a). *PISA Questionnaire Student A*. OECD Publishing. https://www.oecd.org/pisa/

OECD. (2011b). *PISA School Questionnaire*. OECD Publishing. https://www.oecd.org/pisa/

OECD. (2014a). *Educational Career Questionnaire for Pisa 2015 (International Option)*. OECD Publishing. https://www.oecd.org/pisa/

OECD. (2014b). *School Questionnaire for Pisa 2015*. OECD Publishing. https://www.oecd.org/pisa/

OECD. (2016). Overview: Excellence and Equity in Education. In *PISA 2015 Results: Vol. I* (pp. 33–47). OECD Publishing. https://doi.org/10.1787/9789264266490-5-en

OECD. (2017a). PISA 2015 Technical Report. In *OECD Publishing*. https://www.oecd.org/pisa/data/2015-technical-report/

OECD. (2017b). What is PISA? In *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving* (pp. 11–18). OECD Publishing. https://doi.org/10.1787/9789264281820-2-en

OECD. (2020a). *Data base - PISA 2012*. https://www.oecd.org/pisa/data/pisa2012database-downloadabledata.htm

OECD. (2020b). *Data base PISA 2009*. https://www.oecd.org/pisa/data/pisa2009database-downloadabledata.htm

OECD. (2020c). *PISA 2015 Database*. https://www.oecd.org/pisa/data/2015database/

Pallegedara, A. (2012). Demand for private tutoring in a free education country. The case of Sri Lanka. *International Journal of Education Economics and Development*, *3*(4), 375. https://doi.org/10.1504/IJEED.2012.052321

Pallegedara, A., & Mottaleb, K. A. (2018). Patterns and determinants of private tutoring: The case of Bangladesh households. *International Journal of Educational Development*, *59*, 43–50. https://doi.org/10.1016/j.ijedudev.2017.10.004

Portela, M. C. A. S., Thanassoulis, E., Horncastle, A., & Maugg, T. (2011). Productivity change in the water industry in England and Wales: application of the meta-Malmquist index. *Journal of the Operational Research Society*, *62*(12), 2173–2188. https://doi.org/10.1057/jors.2011.17

Runte-Geidel, A. (2013). La incidencia de las clases particulares en España a través de los datos de PISA. *Revista Española de Educación Comparada*, *21*, 249–282. https://doi.org/10.5944/reec.21.2013.7622

Rutkowski, L., & Rutkowski, D. (2016). A Call for a More Measured Approach to Reporting and Interpreting PISA Results. *Educational Researcher*, *45*(4), 252–257. https://doi.org/10.3102/0013189X16649961

Ryu, D., & Kang, C. (2013). Do Private Tutoring Expenditures Raise Academic Performance ? Evidence from Middle School Students in South Korea. *Asian Economic Journal*, *27*(1), 59–83. https://doi.org/10.1111/asej.12002

Safarzyńska, K. (2013). Socio-economic Determinants of Demand for Private Tutoring. *European Sociological Review*, *29*(2), 139–154. https://doi.org/10.1093/esr/jcr045

Silva, R., & Silva, M. C. (forthcoming). Apoio escolar – Breve resumo de 3 anos de PISA. *Cadernos de Economia*.

Song, K.-O., Park, H.-J., & Sang, K.-A. (2013). A cross-national analysis of the student- and school-level factors affecting the demand for private tutoring. *Asia Pacific Education Review*, *14*(2), 125–139. https://doi.org/10.1007/s12564-012-9236-7

Šťastný, V. (2016). Private Supplementary Tutoring in the Czech Republic. *European Education*, *48*(1), 1–22. https://doi.org/10.1080/10564934.2016.1147006

Thanassoulis, E., Portela, M. C. S., & Despić, O. (2008). Data Envelopment Analysis: The Mathematical Programming Approach to Efficiency Analysis. In *The Measurement of Productive Efficiency and Productivity Change*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195183528.003.0003

Ventura, A., Costa, J. A., Neto-Mendes, A., & Azevedo, S. (2008). Dimension and characteristics of the use of private tutoring in the 12th Grade: From the local to the national level. *Xplika: Research on the Private Tutoring Market. Aveiro: University of Aveiro [in Portuguese]*.

Ventura, A., & Jang, S. (2010). Private tutoring through the internet: globalization and offshoring. *Asia Pacific Education Review*, *11*(1), 59–68. https://doi.org/10.1007/s12564-009-9065-5

Appendix I – PISA questionnaire references

The questions we used from the PISA questionnaires in 2009 can be found in:

- question related to additional instruction: STQ32;
- school location: SC04Q01;
- school type: SC02Q01;
- school size: SC06Q01 and SC06Q02;
- number of teachers: SC09Q11 and SC09Q12;
- fully certified teachers: SC09Q21 and SC09Q22;
- computers per pupil: SC10Q01 and SC10Q02;
- computers per pupil with access to the internet: SC10Q01 and SC10Q03.

The questions we used from the PISA questionnaires in 2012 can be found in:

- question related to additional instruction: STQ42;
- school location: SC03Q01;
- school type: SC01Q01;
- school size: SC07Q01 and SC07Q02;
- number of teachers: SC09Q11 and SC09Q12;
- fully certified teachers: SC09Q21 and SC09Q22;
- computers per pupil: SC11Q01 and SC11Q02;
- computers per pupil with access to the internet: SC11Q01 and SC11Q03.

The questions we used from the PISA questionnaires in 2015 can be found in:

- question related to additional instruction: EC001;
- school location: SC001Q01TA;
- school type: SC013Q01TA;
- school size: SC002Q01TA and SC002Q02TA;
- number of teachers: SC018Q01TA01 and SC018Q01TA02;
- fully certified teachers: SC018Q02TA01 and SC018Q02TA02;

- computers per pupil: SC004Q01TA and SC004Q02TA;
- computers per pupil with access to the internet: SC004Q01TA and SC004Q03TA.