UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DOTTORATO DI RICERCA IN MODELLI E METODI PER L'ECONOMIA E IL MANAGEMENT
(ANALYTICS FOR ECONOMICS AND MANAGEMENT - AEM)

SECS-S/06 Metodi Matematici dell'Economia e delle Scienze Attuariali e Finanziarie

ciclo XXXV

THE ROLE OF ICT IN EDUCATION: AN EFFICIENCY ANALYSIS

Candidato: MUHAMMAD MUJIYA ULKHAQ

NOME DEL RELATORE:
Prof.ssa. ROSSANA RICCARDI

NOME DEL CO-RELATORE:
Prof.ssa. GIORGIA OGGIONI

Anno accademico 2021/2022

UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE DOCTOR OF PHILOSOPHY
IN ANALYTICS FOR ECONOMICS AND MANAGEMENT

SECS-S/06 Mathematical Methods of Economics and Actuarial and Financial Sciences

cycle XXXV

THE ROLE OF ICT IN EDUCATION: AN EFFICIENCY ANALYSIS

Candidate: MUHAMMAD MUJIYA ULKHAQ

Name of Supervisor:
Prof. ROSSANA RICCARDI

Name of Co-supervisor:
Prof. GIORGIA OGGIONI

Academic year 2021/2022

ABSTRACT (ITALIAN)

Nell’ambito dell’educazione, l’utilizzo delle tecnologie dell’informazione e della comunicazione (TCI) si è notevolmente intensificato negli ultimi decenni grazie agli investimenti effettuati. Il concetto di TCI è molto ampio. In questo lavoro di tesi, TCI non si riferisce solo alle infrastrutture fisiche (ad esempio radio, telefono, video, televisione, computer), ma include anche l’uso e l’intensità di utilizzo (ad esempio l’impiego giornaliero, settimanale, ecc.), la qualità e l’ubicazione dell’infrastruttura (ad esempio, a scuola oppure a casa), il motivo del suo utilizzo (ad esempio, per svago o per motivi di studio) e la spesa relativa alle TIC. Questa dissertazione discute il ruolo delle TIC nell’istruzione concentrandosi sull’analisi dell’efficienza. La tesi comprende quattro lavori ripartiti in diversi capitoli. Il Capitolo II propone una sistematica literature review sull’argomento. Il Capitolo III esegue un’analisi transnazionale dell’efficienza dell’istruzione a livello scolastico in sei Paesi del sud-est asiatico, ossia in Brunei Darussalam, in Malesia, in Indonesia, nelle Filippine, a Singapore ed in Tailandia. L’analisi viene effettuata mediate l’approccio della stochastic frontier analysis (SFA) che consente di considerare l'eteroschedasticità. Da questo studio risulta che Singapore è comparativamente il Paese con la migliore performance. Nell’analisi condotta, le variabili TIC, ovvero (1) il rapporto tra computer a scuola e (2) il numero totale di studenti ed il rapporto tra computer connessi a Internet, sono assunte essere determinanti dell’inefficienza ed entrano come input nella funzione di produzione (istruzione). Dall’analisi condotta, emerge che il primo rapporto non influenza in modo significativo gli esiti scolastici mentre il secondo ha un significativo impatto. Come determinanti dell’inefficienza, il primo rapporto influisce sull’inefficienza della scuola in nelle aree di matematica e scienze, mentre il secondo non ha alcuna influenza. Il Capitolo IV utilizza l'approccio DEA (non-parametric data envelopment analysis) del modello di super-efficienza che consente alle scuole efficienti di avere punteggi di efficienza superiori a uno (nell’approccio DEA tradizionale, il punteggio di efficienza è limitato da zero a uno). Per studiare i fattori che potenzialmente influenzano l’efficienza, questo studio include anche una seconda analisi basata sull’approccio bootstrapped quantile regression. I risultati suggeriscono una serie di implicazioni politiche per le scuole del sud-est asiatico, indicando diverse linee d’azione per le scuole sia con livelli di efficienza più alti sia per quelle con efficienza minore. Il Capitolo V estende l'analisi condotta nel Capitolo III sia dal punto di vista metodologico che empirico. L’analisi, basata sull’approccio SFA, non include solo le infrastrutture TCI nel modello, ma aggiunge anche l’uso delle TCI (compreso l’indice del tempo trascorso dagli studenti nell’uso delle TCI a scuola, fuori dalla scuola per scopi di intrattenimento e a casa per compiti scolastici). Ciò viene fatto utilizzando il “modello di frontiera stocastica a quattro componenti” in cui le TCI sono modellate sia come input che come determinanti di inefficienza variabile nel tempo. Inoltre, questo modello viene testato utilizzando un set di dati di 24 Paesi OCSE. I risultati mostrano che tutte e tre le variabili che appartengono all’uso delle

TIC influenzano i risultati sul livello di istruzione degli studenti, mentre come determinanti di inefficienza, queste variabili hanno solo un effetto marginale. Questo studio dovrebbe quindi fornire una visione più olistica del ruolo delle TIC nell'efficienza dei processi educativi.

ABSTRACT (ENGLISH)

In education sector, the application of information and communication technology (ICT) has increased substantially over the last decades as many countries have been investing their resources in ICT for educational purposes. The ICT is a broad concept. In this dissertation, ICT does not only refer to physical infrastructure (e.g., radio, telephone, video, television, computer), but it also includes the use and the intensity of use (e.g., every day, one a week, twice a week), the quality and location of the infrastructure (e.g., at school, at home), the reason for using it (e.g., for entertainment or for study purposes), and the expenditure related to the ICT. This dissertation then discusses the role of ICT in education focusing on the efficiency analysis. It comprises four studies starting with a systematic literature review presented in Chapter II, which offers a clear overview of what has and has not been done in the literature towards this particular topic. Chapter III performs cross-country analysis of efficiency of education at school level in six countries in South-East Asia (i.e., Brunei Darussalam, Malaysia, Indonesia, the Philippines, Singapore, and Thailand). The stochastic frontier analysis (SFA) allowing for heteroscedasticity is used. The result reveals that Singapore has the (relatively) best performance among other countries. The ICT infrastructure variables, i.e., the ratio of computers at school to the total number of students and the ratio of computers connected to the internet, are modeled as inputs in the (education) production function and determinants of inefficiency. The first ratio is found to be not significant influencing education outcomes while the second one does influence. As determinants of inefficiency, the first ratio affects school's inefficiency in terms of mathematics and science, while the second one has no influence. Relying the finding of Chapter III that there are many higher efficiency level schools, Chapter IV uses the non-parametric data envelopment analysis (DEA) approach of the super-efficiency model which has the ability to differentiate among the higher efficiency level schools. This model allows the efficient schools to have efficiency scores of more than one (in the traditional DEA approach, the efficiency score is bounded from zero to one). To investigate factors that potentially influence efficiency, this study performs the "second-stage" analysis by using bootstrapped quantile regression. The results suggest a number of policy implications for South-East Asian schools, indicating different courses of action for schools with higher and lower efficiency levels. Chapter V extends the analysis conducted in Chapter III both from methodological and empirical point of views. The analysis, based on the SFA approach, not only includes the ICT infrastructure in the model, but it also adds the ICT use (including the index of time spent by students in using ICT at school, outside school for entertainment purposes, and at home for school-related tasks). This is done by using the "four-component stochastic frontier model" where ICT is modeled both as inputs and determinants of time-varying inefficiency. In addition, this model is tested using a dataset of 24 OECD countries. Results show that all three variables belong to ICT use influence education outcomes, while as the determinants of time-varying inefficiency, these variables have only marginal effect on inefficiency. This study is

then expected to provide a more holistic view of the role of ICT in the efficiency of education measurement as the previous studies only addressed the ICT infrastructure.

This is dedicated to:
Trisna Nagris Pratiwi (my lovely wife),
Nazala Arjunka Ulkhaq (my adorable son),
Naziha Araela Ulkhaq (my pretty daughter).

ACKNOWLEDGMENT

I would like to express my sincere gratitude to everyone who was involved in completing this dissertation. Firstly, I'd like to express my appreciation to my supervisors, Rossana Riccardi and Georgia Oggioni, who was providing me valuable guidance and feedback through my Ph.D. period Secondly, to Kristof De Witte, whose insight and knowledge into the subject matter steered me through this research. Finally, my biggest thanks to my family for all the unconditional support you have shown me through this research, the culmination of three years of this Ph.D. journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

[This page is intentionally left blank]

CHAPTER I. INTRODUCTION

The role of information and communication technology (ICT) is apparent in this twenty-first century and has become the most distinctive feature in this modern life. It has triggered a huge global impact on the world; thus, the progress and prosperity of countries are commonly associated with the extent of its progress and achievements (Al-araibi et al., 2019). Because of this, most countries have commenced advancing various institutions to keep pace with this technological revolution.

ICT is a broad subject, and the concepts are evolving. It refers to all forms of technologies (e.g., radio, telephone, video, television, computer, and associated services) for creating, storing, sharing or transmitting, exchanging, managing, and analyzing information (Tinio, 2003). However, in relation to education, the previous definition should be redefined and properly measured (Biagi and Loi, 2013) since it only refers to the *physical infrastructure*. It should include the use and the intensity of use (e.g., every day, one a week, twice a week) (Spiezia, 2010); it also does not account for the location of the infrastructure (e.g., at school, at home); the quality of the infrastructure (e.g., Aparicio et al., 2019; Deutsch et al., 2013); the engagement with students (Chiang, 2021a, b); what students do when they use ICT (e.g., for entertainment or for school-related purposes); expenditure related to ICT (e.g., Johnes, 2006, 2008, 2013); or even digital competences (Biagi and Loi, 2012). Education has a unique role to play in providing students with the skills needed in a society in which ICT-related skills and competences are increasingly indispensable. It is, therefore, important to study how education systems are dealing with the integration of ICT in education.

In the education sector, the application of ICT has increased substantially over the last few years (Comi et al., 2017; Falck et al., 2018). Many countries have been investing their resources in ICT infrastructure for educational purposes. Kozma (2008) pinpointed four important reasons for ICT investment in education, i.e., (i) to support economic growth, (ii) to promote social development, (iii) to advance education reform, and (iv) to support the management of education and its accountability. Even though in more recent times the diffusion of ICT in education has lost its status as a policy priority, the investments have not ceased. As OECD (2010) observed, “Education systems keep investing in technology in the belief that, sooner or later, schools and teachers will adopt it and benefit from it”. Therefore, the question is: “has this investment paid off in terms of higher efficiency?”

The term efficiency commonly refers to the ratio of output to input (Cooper et al., 2006) and it is widely used as a measure of performance evaluation. In the field of education, the expression is used to describe the ability of an education institution or education system (school, university, or country) to produce a given level of output with the number of available inputs. The outputs are commonly related to the achievements obtained by students (e.g., test score) or by researchers (e.g., number publications, citations, research grants, patents). On the other hand, there are three kinds of variables reflecting the inputs: (i) characteristics of school, e.g., size of school, number of teachers, (ii) characteristics of teacher, e.g., level of education, salary, and (iii) characteristics of students, e.g., socio-economic characteristics, prior attainment.[1] In this way, the education institution or education system is assumed to be a "producer" that transforms inputs into outputs through a production process (Hanushek, 1979).

This dissertation discusses the role of ICT in education focusing on the efficiency analysis. It consists of four studies depicted in Chapter II until Chapter V. Chapter II presents a systematic literature review on the role of ICT in the efficiency of education. It presents a clear overview of what has and has not been done in the literature towards this topic, including the level of analysis, data source, method used, and ICT-related variable used in the measurement model. At the end of Chapter II, three critical discussions are provided. The first critical remark is that there is limited literature about international comparisons of the efficiency of education analysis. Despite the intrinsic interest of international comparisons, there are two main reasons behind the limited development of studies looking to compare the results between countries, i.e., (i) the lack of reliable datasets and (ii) the substantial differences in institutional (country-specific) settings (Agasisti and Zoido, 2019). Therefore, to overcome both of these limitations, the next three studies use the OECD PISA data, which is well-regarded as an authoritative source of comparison for educational achievement across the world.

The study in Chapter III performs cross-country analysis of efficiency measurement of education in South-East Asia. This study is conducted at the school level by using the recent OECD PISA data 2018. The stochastic frontier analysis allowing for heteroscedasticity is used. The ICT infrastructure variables (i.e., ratio of computers at school to the total number of students and ratio of computers connected to the internet) are modeled as inputs in the (education) production function and determinants of inefficiency.

If Chapter III measures efficiency in a parametric setting, Chapter IV, on the other hand, uses a non-parametric approach to measure efficiency of schools in South-East Asia. Relying on the finding that there are many efficient schools, this study then expands the analysis by using a super-efficiency model to differentiate among those best performers (i.e., the efficient schools). This model allows the efficient schools to have efficiency scores of more than one (note that in the traditional non-parametric approach, the efficiency score is bounded from zero to one). In the first stage, data envelopment analysis is used to measure school's efficiency. To investigate factors that potentially influence school's efficiency, this study performs the "second-stage" analysis by using a bootstrapped quantile regression. ICT infrastructure is again used to represent ICT in this model; it is modeled as input and determinant of efficiency.

[1] See De Witte and López-Torres (2017) who comprehensively reviewed inputs and outputs used in the field of efficiency in education.

The second remark discussed in Chapter II is about the definition of ICT. Previous studies only dealt with the physical infrastructure of ICT. As mentioned previously, ICT is not only about the physical thing; thus, in Chapter V, it adds more ICT-related variables, called ICT use. These variables include the index of time spent by students in using ICT at school, outside school for entertainment purposes, and at home for school-related tasks. This study also uses the OECD PISA data, yet it comes from different samples, i.e., from 24 OECD countries. The ICT familiarity questionnaire offered in those OECD countries—which is not a mandatory practice in the PISA assessment—allows for adding more variables related to ICT.

The third remark discusses the method used in the efficiency measurement. Although there are extensive studies on this field, efficiency is mainly assessed by frontier methods: non-parametric e.g., data envelopment analysis (DEA), and its parametric counterpart, e.g., stochastic frontier analysis (SFA). It should be noted that previous studies reviewed in Chapter II did not apply the recent method in analyzing efficiency. A recent development in SFA dealing with panel data is the "four-component stochastic frontier model" which composes of producer effects, persistent and time-varying inefficiency, as well as statistical noise (Colombi et al., 2014; Kumbhakar et al., 2014; Tsionas and Kumbhakar, 2014). Using the OECD PISA data of 2009 to 2018 waves from 24 OECD countries, Chapter V allows for panel data analysis. The "four-component stochastic frontier model" is used, where ICT is modeled both as inputs and determinants of time-varying inefficiency. Therefore, Chapter V deals with all remarks discussed in Chapter II.

The remainder of this chapter is structured as follows. Chapter I.1 provides the objectives and contributions of four studies in this dissertation. Chapter I.2 presents the outputs of this dissertation.

I.1 Objectives and Contributions

This section describes the objectives and contributions of four studies included in this dissertation. Chapter II presents a systematic literature review on the role of ICT in the efficiency of education. Using the Scopus database, extracted articles are analyzed using the scientometrics analysis of science mapping to analyze bibliographic networks. It presents a visualization of co-authorship, co-citation, co-occurrence, and bibliographic coupling analysis. This chapter then classifies the extracted articles according to the level of analysis, data source, method used, the role of ICT in the measurement model, as well as its influence and significance. It is expected provide a clear overview of what has and has not been done in the literature towards the role of ICT in efficiency of education measurement.

This literature review contributes to the literature by presenting a systematic literature review on the role of ICT in the efficiency of education. To date, there is no review paper which specifically reviews the role of ICT in the measurement of efficiency in education. Previous literature reviews only discussed the efficiency of education in a *general way* (i.e., De Witte and López-Torres, 2017; Rhaiem, 2017; Worthington, 2001); thus, the effect of ICT, in particular, might not be investigated clearly. In addition, different to other studies which described the ICT in the efficiency of education *qualitatively* (e.g., Ciroma, 2014; Făt and Labăr, 2009; Lim et al., 2020; Sosin et al., 2004), this study provides the role of ICT in education in relation to the theory of economics of education about measuring efficiency so that one can identify the influence of ICT in a *quantitative* way.

Chapter III aims to measure, compare, and analyze the efficiency at school level in six countries in South-East Asia (i.e., Brunei Darussalam, Indonesia, Malaysia, the Philippines, Thailand, and Singapore) in a cross-sectional setting. This study also investigates the influence of ICT infrastructure (i.e., ratio of computers at school to the total number of students and ratio of computers connected to the internet) on education outcomes—controlled for other school's characteristics variables—and on inefficiency. Since this study enables international comparisons; thus, it is expected to benchmark against schools in South-East Asia. It is expected to enlarge the perspective of each institution, helping the process of spurring innovation and new ideas of how to employ resources more efficiently.

Chapter III contributes to the literature as the following. First, since this study uses SFA to measure the efficiency, it then attempts to extend the literature as the use of SFA in measuring efficiency of education particularly in South-East Asia is quite limited. As the second contribution, this study includes ICT as inputs and determinants of inefficiency to investigate the influence of ICT on both education outcomes and inefficiency. Previous studies investigating the education systems in South-East Asia did not incorporate ICT into their model; thus, how ICT influences education outcomes as well as inefficiency has not been yet investigated. Lastly, it can be considered as the first study which assesses the efficiency across country in this region. Previous studies only focused on one specific country in the region, while in this study, a more detailed analysis is provided in cross country level.

Similar to Chapter III, Chapter IV also analyses the efficiency of schools in South-East Asia. However, study in Chapter IV has the ability to differentiate among the efficient schools by the means of super-efficiency model. The study also investigates the determinants of efficiency (i.e., factors that potentially explain efficiency) using the bootstrapped quantile regression. This allows comparison of how some percentiles of the efficiency levels may be more affected by certain determinants than other percentiles. Therefore, it is expected to give a more comprehensive picture of the influence of the determinants on efficiency to any part of the distribution of efficiency.

Study in Chapter IV contributes to the literature as it applies the two-stage super efficiency with the bootstrapped quantile regression. To date, there is no study uses this approach in measuring efficiency in education sector. The use of bootstrapped quantile regression is motivated as follows. First, the use of ordinary least squares (OLS) as a traditional way to investigate the effect of determinants of efficiency in the second stage (see e.g., in Banker and Natarajan, 2008; Iliyasu and Mohamed, 2016; Sultan and Crispim, 2018) is flawed by the fact that usual inference on the obtained estimated of the regression coefficient is not available (Simar and Wilson, 2007). The bootstrap procedure is then proposed to obtain more accurate inference. In addition, the bootstrap can be used to correct for the biases resulting from the correlation between the inputs or outputs of the first stage and the regressors of the second stage. The fact that the efficiency distribution is skewed corroborates the use of quantile regression which relies on the conditional quantiles rather than the conditional means as in the OLS. Due to these benefits of the bootstrapped quantile regression, it is suggested that this procedure would present more insightful information compared to the conventional OLS.

Chapter V primarily aims to investigate the influence of ICT infrastructure and ICT use (including index of time spent by students in using ICT at school, outside school for entertainment purposes, and at home for school-related tasks) on both education outcomes and time-varying inefficiency. As the provision of ICT infrastructure is considered a key element for schools to be able to exploit the many benefits that digital technologies bring to teaching and learning, however, the infrastructure-related policies should be accompanied by complementary measures in other areas, such as the use of this infrastructure. Therefore, by including the ICT use variables, this study is expected to both extend the literature and provide a more holistic view of the role of ICT in the efficiency of education measurement by including three under-studied ICT-related variables, called the ICT use, since previous studies only incorporated the ICT infrastructure into the model of efficiency measurement. This analysis is carried out by using the OECD Pisa data from 2009 to 2018 wave from 24 OECD countries. The comparison of different waves and different countries is able to capture the effects of different policies on ICT use and infrastructure to define the benchmark countries in education efficiency and to define which ICT-variables can have positive or negative effects in determining school's performance. From the methodological point of view, in order to capture both persistent and time-varying effects of the ICT-related variables, this study extends the application of the "four-component stochastic frontier model" in the education sector as there is limited study which applied this model in the education sector.

I.2 Outputs of the Dissertation

The main results of this dissertation have been collected in four different works:

1. Ulkhaq, M. M., Riccardi, R., Oggioni, G., and De Witte, K., "Does ICT enhance the efficiency of education? A systematic literature review" under review in *Review of Educational Research*. This paper mainly refers to the study in Chapter II.
2. Ulkhaq, M. M., Oggioni, G., and Riccardi, R., "How efficient are schools in South-East Asia? An analysis through OECD PISA 2018 data" under review in *Asia Pacific Education Review*. This paper refers to the study in Chapter III. In addition, some parts of this work have been presented in The Joint Conference of 8th Annual Conference on Industrial and System Engineering and 1st

International Conference on Ergonomics, Safety, and Health, Semarang-Surabaya, Indonesia [Online], July 13-15, 2021.

3. Ulkhaq, M. M., Oggioni, G., and Riccardi, R., "Two-stage super-efficiency model for measuring efficiency of education in South-East Asia" under review in *Decisions in Economics and Finance*.
4. Ulkhaq, M. M., Oggioni, G., Riccardi, R., and De Witte, K., "Does the use of ICT improve education outcomes and reduce inefficiency?", advanced review in *Socio-Economic Planning Sciences*. This work also has been presented in four conferences: (i) 31st European Conference on Operations Research, Athens, Greece [Hybrid], July 11-14, 2021; (ii) 22nd Conference of the International Federation of Operational Research Societies, Seoul, Rep. of Korea [Online], August 23-27, 2021; (iii) Association for Mathematics Applied to Social and Economic Sciences Conference XLV, Reggio Calabria, Italy [Online], September 13-18, 2021; and (iv) 8th International Workshop on Efficiency in Education, Health and other Public Services, Pisa, Italy, September 8-9, 2022.

Two other papers have been produced during the Ph.D. period and are collected in the Appendix.

1. Ulkhaq, M. M. (2020), "Clustering countries according to the World Happiness Report," *Statistica & Applicazioni*, vol. XVIII, no. 2, pp. 197-220. This study aims to cluster countries according to the World Happiness Country Report 2020. Nine clustering algorithms are used, compared, and analyzed. Interpretation of the finding is also provided. This study was the output from the course of "Multivariate Statistics" in the first year of the Ph.D. period.
2. Ulkhaq, M.M., Pramono, S.N.W., and Adyatama, A., "Assessing the tendency of judging bias in student competition: A data mining approach," *Journal of Applied Research in Higher Education*, in press (https://doi.org/10.1108/JARHE-02-2022-0053). This study aims to investigate the presence of judging bias in a university student competition. It attempts to expand the literature on judging bias by proposing the term "universitarian bias" as the judge coming from a particular university tends to give a higher score to a participant coming from the same university. The association rule of data mining is used to accomplish the objective of the study. This study was the output of a project when the first author was in Indonesia due to the pandemic condition.

CHAPTER II. ICT IN THE EFFICIENCY OF EDUCATION: A SYSTEMATIC LITERATURE REVIEW

Briefly, this chapter describes a systematic literature review on the role of information and technology communication (ICT) in the efficiency of education. The Scopus database is used to extract the relevant articles. First, in presenting the extracted articles, a scientometrics analyses are used to visualize the bibliometric clusters, namely, co-authorship analysis, co-citation analysis, co-occurrence analysis, and bibliographic coupling analysis. Next, a qualitative approach is used to classify the extracted articles according to the level of analysis, data source used, and methods used in measuring the efficiency. In addition, the role of ICT—whether as inputs, outputs, or determinants of (in)efficiency—is summarized. Finally, the influence and significance of ICT on education outcomes and (in)efficiency are investigated.

This chapter is structured as follows. In Chapter II.1, motivational background and contributions of this chapter are presented. Chapter II.2 describes methodology and mapping strategy. In Chapter II.3, several scientometric analyses are conducted to identify what clusters of influence exist between authors, articles, journals, countries, and keywords. Chapter II.4 shows the classification of the articles, including the levels of analysis, data source, and methods used in the assessment of efficiency. Chapter II.5 offers a discussion about the role of ICT in the efficiency measurement model; whereas ICT's influence and significance are discussed in Chapter II.6. Chapter II.7 provides critical discussion and concluding remarks.

II.1 Motivational Background and Contributions

At the first sight, there are two opposite sets of observations in the literature about the influence of ICT on the efficiency of education (De Witte and Rogge, 2014). On the one hand, some scholars found that ICT could reduce educational costs. Other advantages are improving the delivery of education and the learning process, the presence of greater flexibility and autonomy for the students' learning, as well as supporting more interaction and a reduction in the teachers' workload (Grimes and Warschauer, 2008; Lei and Zhao, 2008; Venable et al., 2011). On the other hand, when ICT is not well integrated in the curriculum, due to pedagogical barriers, it might hinder students from learning (Fu, 2013). There are also some barriers that obstruct the use of ICT in education from the teacher perspective, such as a lack of teacher collaboration and pedagogical support, a lack of in-service training on the use of ICT, insufficient time to master new educational software or to integrate ICT during a class period, limited knowledge and experience of ICT in teaching contexts, as well as several technical problems related to

ICT in the classroom that frequently happened. Moreover, when the teachers use ICT in the classroom, it might negatively distract the students.

Due to this inconclusive explorative finding, it is worth to systematically explore in depth the influence of ICT on the efficiency of education. This study contributes to the literature by presenting a systematic literature review on the role of ICT in the efficiency of education. A recent study by Daraio et al. (2020) which investigated review papers in the field of efficiency and productivity analysis revealed only three articles conducted a review on efficiency of education, i.e., Worthington (2001), Rhaiem (2017), and De Witte and López-Torres (2017). To the best of our knowledge, there is no review paper which specifically examines ICT in the field of efficiency of education as the role of ICT is apparent in this twenty-first century and has turned into the most distinguishing feature in this modern life. Different to other studies which described the ICT in the efficiency of education *qualitatively* (e.g., Ciroma, 2014; Făt and Labăr, 2009; Lim et al., 2020; Sosin et al., 2004), in this study, it is presented a clear overview of what has (and has not) been done in the literature towards the role of ICT in education in relation to the theory of economics of education about measuring efficiency (see e.g., Coelli et al., 2005; Fried et al., 2008); so that one can identify the influence of ICT in a *quantitative* way.

II.2 Methodology

The steps to conducting the review are illustrated in Figure II-1. In Step 1, the research objective has to be defined well. In this case, this study aims to systematically review the academic articles regarding the role of ICT in the assessment of efficiency in education.

In Step 2, the relevant search query must be defined. In this study, the search query is the following: TITLE-ABS-KEY((ICT OR information communication technolog* OR computer OR internet) AND efficien* AND (education OR school OR universit*) AND (DEA OR data envelopment analysis OR FDH OR free disposal hull OR SFA OR stochastic frontier analysis OR frontier OR

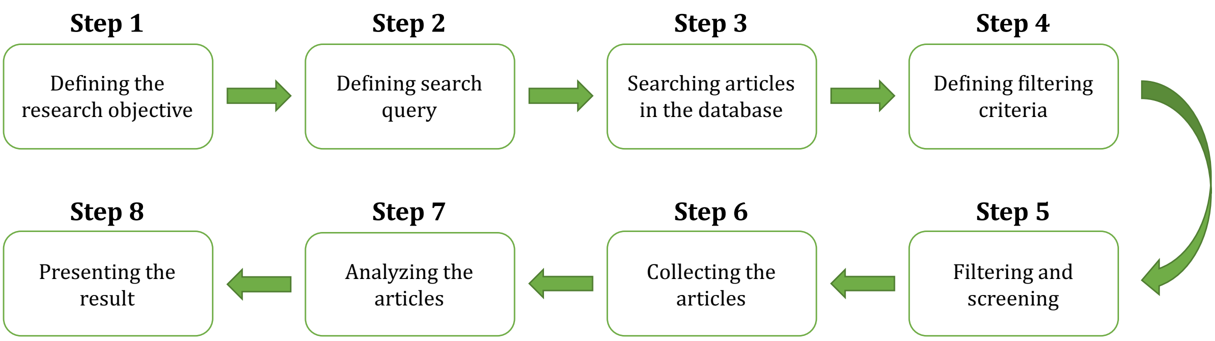


Figure II-1 Steps of the systematic literature review

*parametric)).[2] Thus, articles which contain this search query in the title, abstract, or keywords are extracted. Three most popular frontier methods for assessing efficiency, i.e., data envelopment analysis (DEA), stochastic frontier analysis (SFA), and free disposal hull (FDH) (De Witte and López-Torres, 2017), are included in the search query.[3] The words "frontier" and "*parametric" are also added to cover other quantitative methods in the efficiency measurement. For the sake of quality assurance, the document type is restricted to peer-reviewed research article published in a journal. From a pragmatic point of view, only articles published in English are included. Articles published for the last two decades, during 2002—after the period of the dot-com bubble—to December 2021 are considered. Due to the rapid ICT development in the education, including old literature would be misleading and irrelevant.

In Step 3, to get a widespread coverage of the literature, the biggest scientific database is chosen, i.e., Scopus by Elsevier (https://www.scopus.com/). This database provides access to scientific articles and a wide-ranging of journals from various fields, including science, mathematics, engineering, technology, health and medicine, social sciences, and arts and humanities. The search procedure contains four phases following the PRISMA framework (Moher et al., 2009), i.e., (i) identification, (ii) screening, (iii) eligibility, and (iv) inclusion. Using previous search query, the search yields 693 articles.

The next steps are to define the filtering criteria and conduct filtering and screening processes. The titles and abstracts of the extracted articles are read to verify whether the articles are relevant to this study's objective. In this way, 508 articles are excluded after the first-round inspection. The exclusion is due to the following reasons: first, the excluded articles did not discuss efficiency in the education sector; and/or second, they described efficiency in education qualitatively (in this study, only quantitative study is considered). The second-round inspection is performed by carefully reading full-text of each article to address the eligibility of the articles to be included. Notice that for a practical reason, articles whose full text cannot be accessed are also excluded. Following this procedure, 152 articles do not meet the selection criteria and are not considered for the final extraction. Most of the excluded articles (61.84%) do not incorporate ICT into their models and they solely discussed performance measurement without being accompanied by efficiency analysis (5.92%). Manual forward and backward chaining of the extracted articles are also performed to ensure that the risk of missing articles is minimized. It might be the case that relevant articles do not put the search query in their title, abstract, and keywords. This procedure resulted in the addition of 17 articles that meet the screening and eligibility criteria. Finally, the final screening process results in 41 articles that satisfied the criteria for inclusion in this review. This screening procedure is summarized in Figure II-2.

---

[2] The asterisk sign (*) is used to find a root word plus all the words made by adding letters to the end (or beginning) of it.

[3] It is encouraged to see Fried et al. (2008) who reviewed different frontier methods in measuring efficiency.

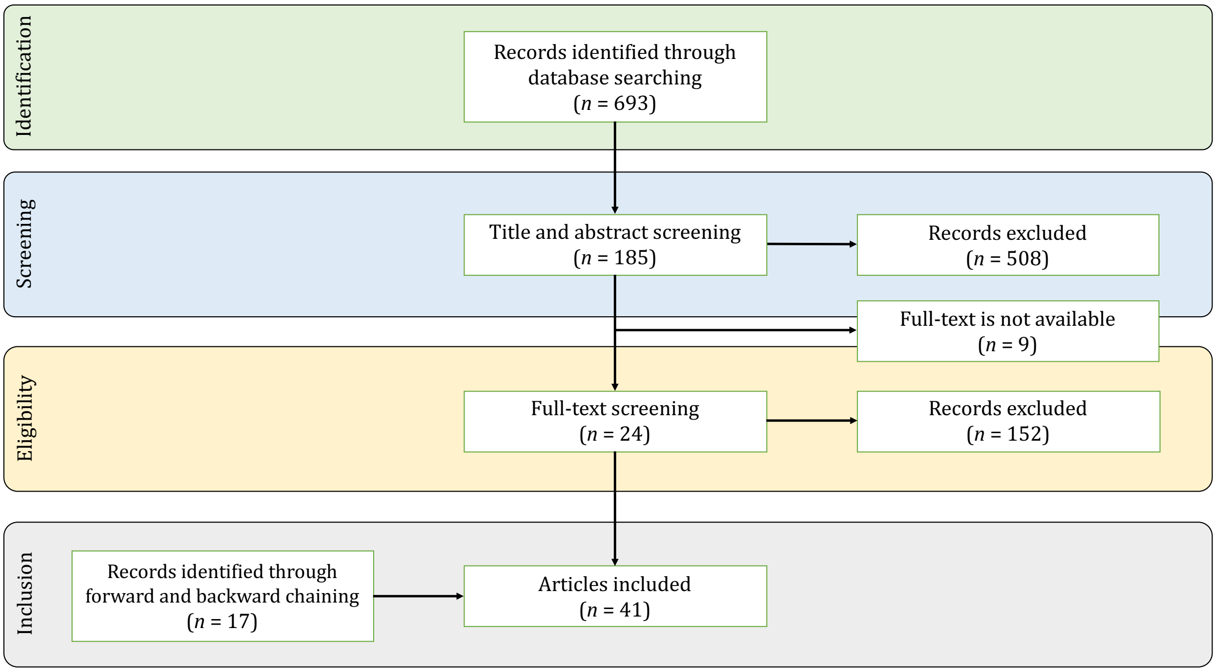
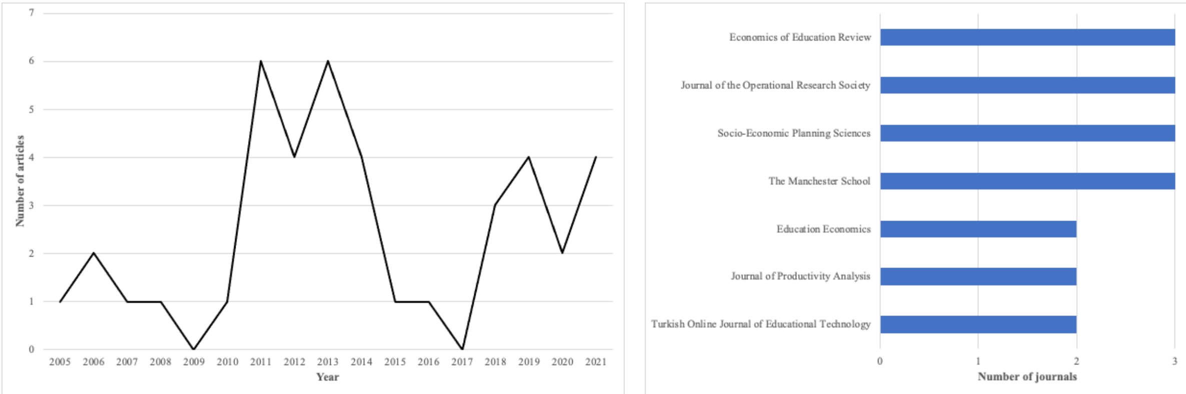


Figure II-2 The PRISMA framework for the screening procedure

In analyzing and presenting the collected articles, the science mapping approach and qualitative approach are used. Several scientometrics analyses are used as science mapping approach, whereas the qualitative approach is employed to classify the extracted articles according to the level of analysis, data source, methods used, the role of ICT-related variables used, as well as the influence and significance of the ICT-related variables used in the efficiency measurement model.

II.3 Scientometric Analysis

Science mapping is a quantitative approach that utilizes visualization techniques as well as statistics to analyze bibliographic networks (e.g., keywords, authors, journals, citations, institutions, and countries) in a specific area (Jin et al., 2019). Scientometric analysis as a tool for performing science mapping is conducted in this review, namely, co-authorship analysis, co-citation analysis, co-occurrence analysis, and bibliographic coupling analysis. This study uses VOSviewer software (van Eck and Waltman, 2010) to conduct the analysis. Fractional counting is used, in which each article has only one unit that it is fractioned according to the number of co-authors (Cancino et al., 2017; Gaviria-Marín et al., 2018; Martínez-López et al., 2020). It is used to normalize the influence of articles with multiple authors. When fractional counting is used, the strength of a co-authorship link between two authors is determined not only by the number of articles co-authored by the authors, but also by the number of authors of each co-authored article.



(a) Number of articles per year

(b) Name of top seven journals

Figure II-3 Number of articles per year and name of top seven journals

The final screening process results in 41 articles. Figure II-3 (a) illustrates the number of articles per year. One might notice random (or perhaps fluctuating) trend; in which, somehow, the trend cannot be estimated (or projected). Including ICT into a model to assess efficiency does depend—as stressed by Portela and Camanho (2007)—on the objective of the study, on the perspective taken (that of the society or of the educational authorities), as well as on the level at which the efficiency analysis is implemented. The collected articles are spread across 30 different journals, whereas top 7 journals—which published more than two articles—are shown in Figure II-3 (b). *Economics of Education Review* (ISSN: 0272-7757), *Journal of the Operational Research Society* (ISSN: 1476-9360), *The Manchester School* (ISSN: 1467-9957) and *Socio-Economic Planning Sciences* (ISSN: 0038-0121) each published three articles in this review.

In the visualized networks using VOSviewer, a network map is formed by applying nodes and lines connecting the nodes. A node symbolizes a particular bibliographic item, such as keywords, article, journal, institution, or country. The node size denotes the counting of the evaluated item, i.e., citation or occurrence. The link denotes the co-citation, co-occurrence, or collaboration relationship. There are three steps to construct the map. In the first step, a similarity matrix is calculated. VOSviewer uses a similarity measure known as the association strength (van Eck and Waltman, 2007; van Eck et al., 2006). Using this association strength, the similarity s_{ij} between two items i and j is calculated as

$$s_{ij} = c_{ij} / (w_i \cdot w_j), \tag{II-1}$$

where c_{ij} denotes the number of co-occurrences of co-cited of items i and j; and w_i and w_j denote either the total number of occurrences (or co-cited) of items i and j or the total number of co-occurrences (or co-cited) of these items. The second step is constructing the map based on the similarity matrix obtained in the previous step. Let n denote the number of items to be mapped. The mapping technique constructs a two-dimensional map in which the items 1, 2, …, n are located in such a way that the distance between any pair of items i and j reflects their similarity s_{ij} as accurately as possible. Items that have a high similarity should be located close to each other, while items that have a low similarity should be located far from each other. The idea of the mapping technique is to minimize a weighted sum of the squared Euclidean distances between all pairs of items. The higher the similarity between two items, the higher the weight of their squared distance in the summation. To avoid trivial maps in which all items have the same location, the constraint is imposed that the average distance between two items must be equal to 1. In a mathematical notation, the objective function to be minimized is given by

$$V(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_n) = \sum_{i<j} s_{ij} \left\| k_i - k_j \right\|^2, \qquad \text{(II-2)}$$

where the vector $\mathbf{k}_i = (k_{i1}, k_{i2})$ denotes the location of item i in a two-dimensional map; and ||•|| denotes the Euclidean norm. Minimization of the objective function is performed subject to the constraint

$$\frac{2}{n(n-1)} \sum_{i<j} \left\| k_i - k_j \right\| = 1. \qquad \text{(II-3)}$$

The optimization problem discussed in the second step does not have a unique globally optimal solution. It is of course important to produce a consistent result, i.e., the same co-occurrence matrix should therefore always yield the same map (ignoring differences caused by local optima). To accomplish this, in the third step, it is necessary to transform the solution obtained for the optimization problem discussed in the second step. There are three transformations applied, namely, translation (the solution is translated in such a way that it becomes centered at the origin), rotation (the solution is rotated in such a way that the variance on the horizontal dimension is maximized), and reflection (if the median of $k_{11}, \ldots, k_{n1}$ is larger than 0, the solution is reflected in the vertical axis; on the other hand, if the median of $k_{12}, \ldots, k_{n2}$ is larger than 0, the solution is reflected in the horizontal axis). Those three transformations are sufficient to ensure consistent results.

II.3.1 Co-authorship analysis

Co-authorship has been operationalized as a proxy for research collaboration (Melin and Persson, 1996). Research collaboration is an interesting issue due to formalized shift in the policy-for-science paradigm from funding individual investigators to funding groups. This is because presumably, the more experts are working together on a particular problem, the better the chances for effectiveness, innovativeness, and/or productivity (Wuchty et al., 2007). Accordingly, recently, many public research investments are made in organized research units with different types of expertise from different economic sectors and/or from different disciplines (Block and Keller 2009). Figure II-4 illustrates the co-authorship network. In this network, nodes represent authors, which are connected when they share

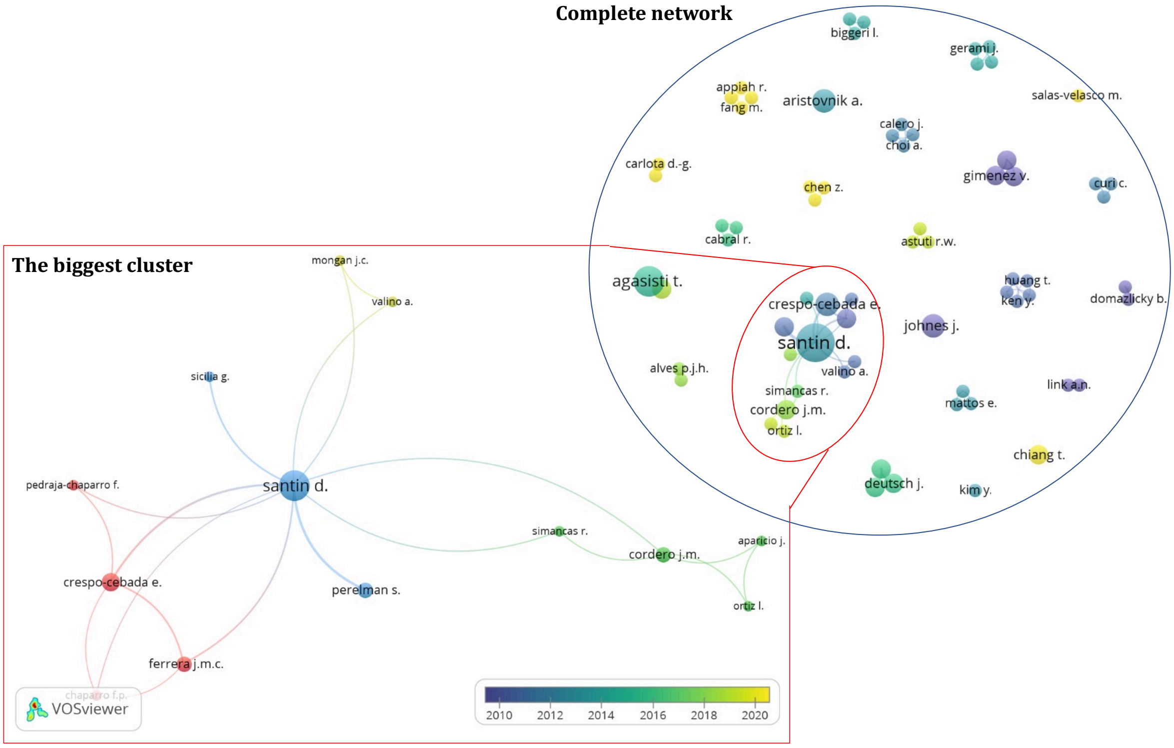


Figure II-4 Co-authorship network

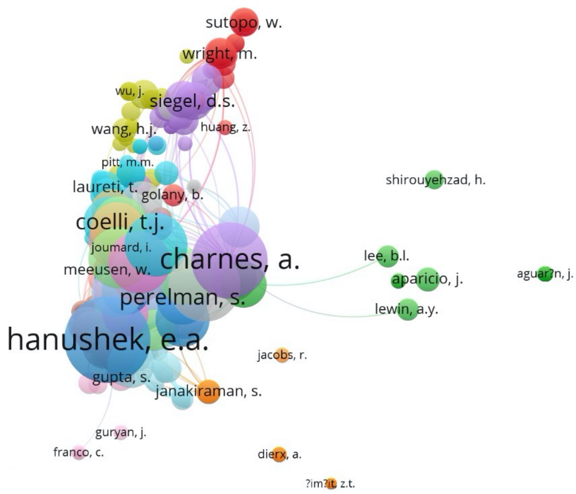


Figure II-5 Author co-citation network

the authorship of an article. In Figure II-4, there is an expanded view of the biggest cluster consisting of 13 nodes. In the expanded view, Daniel Santín acts as an anchor. He wrote 8 articles—the most among others—with Sicilia (Santín and Sicilia, 2018); with Cordero and Simancas (Cordero et al., 2017); with Crespo-Cebada and Pedraja-Chaparro (Crespo-Cebada et al., 2014), with Perelman (Perelman and Santín, 2011a, b), with Mongan and Valiño (Mongan et al., 2011), with Ferrera, Crespo-Cebada, and Chaparro (Ferrera et al., 2011); and with Ferrera and Crespo-Cebada (Ferrera et al., 2010).

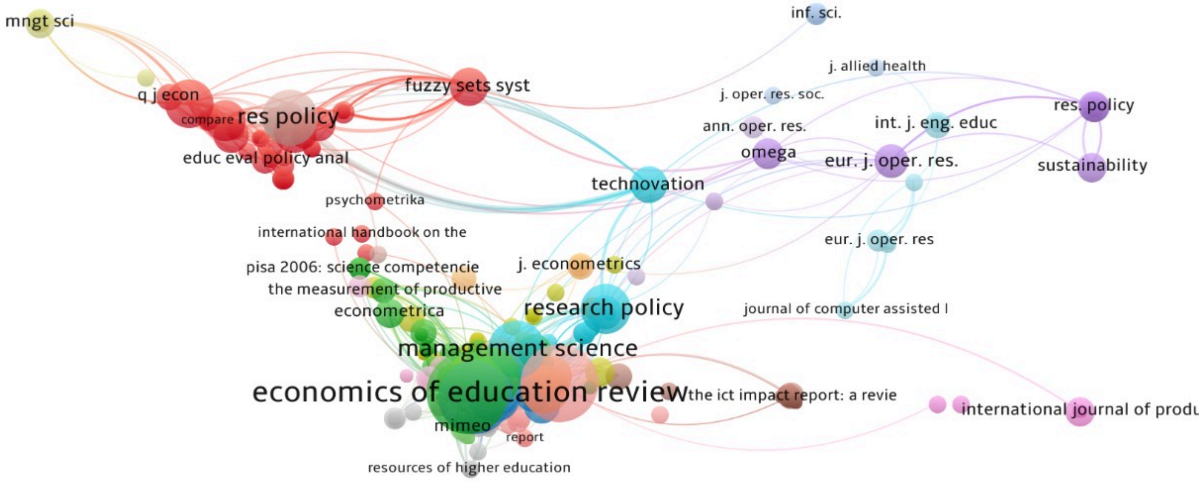


Figure II-6 Journal co-citation network

II.3.2 Co-citation analysis

Co-citation of articles occurs when two articles receive a citation from the same third articles (Cancino et al., 2017). Analysis of co-citation relies on the assumption which two articles cited together are highly related (White and Griffith, 1981), and thus should be concentrated in a cluster solution of a visualization map. It aims to highlight the influential articles and the corresponding reference relationship. The network of author co-citation and journal co-citation are shown in Figure II-5 and Figure II-6, respectively. Each node represents an author (or journal), and the relationship between them (i.e., by co-citations) is indicated by the links between the nodes. The distance between two nodes approximately indicates the relatedness of the two authors (or journals) in terms of co-citations (van Eck and Waltman, 2010). The larger the author's (or journal's) name and the larger the circle, the greater the weight of the node. The weight of each node is determined by the total link strength (TLS) of all the links connected to the node. TLS reflects the correlation between any two nodes in the formed networks; the higher value of TLS, the higher centrality and importance of the node has (Hu et al., 2019). In Figure II-5, the network consists of 1,872 different authors cited by articles in this review. Eric Hanushek is the most influential authors in this research domain since he has been cited the most by articles in this review; followed by Abraham Charnes. In Figure II-6, the network consists of 909 different cited journals. *Economics of Education Review* has the most citations (66 citations), followed by *European Journal of Operational Research* (ISSN: 0377-2217) with 59 citations, and *Education Economics* (Online ISSN: 1469-5782) with 52 citations.

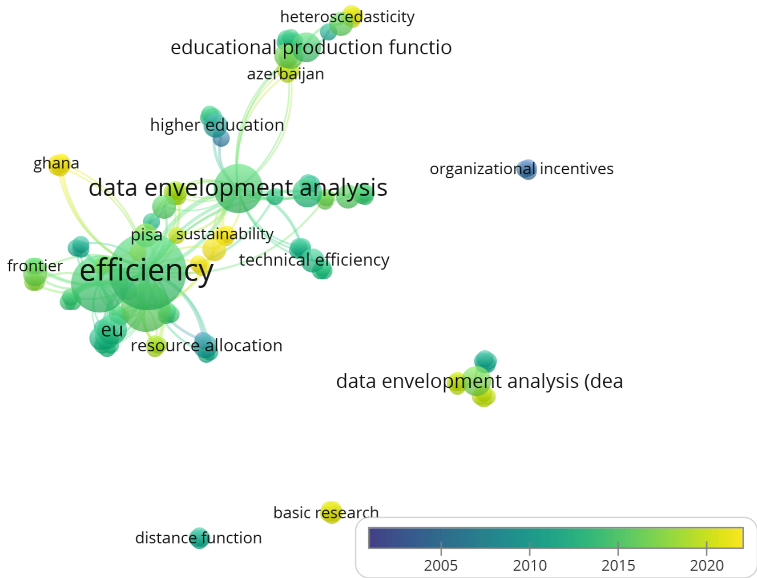


Figure II-7 Keywords co-occurrence network

II.3.3 Co-occurrence analysis

Keywords co-occurrence analysis aims to map the co-occurred keywords and group them into several research clusters. It is performed to explain the structure and internal composition of the research domain as well as to reveal the frontier (Hu et al., 2019). The map can be used to elucidate the knowledge structure of the research theme and it might help to identify the potential research opportunities in the future. Keywords co-occurrence network is shown in Figure II-7. Among 127 keywords, the most frequent keyword is “efficiency” which appeared twenty times, followed by “education” (11 times), and “data envelopment analysis” (8 times).

II.3.4 Bibliographic coupling analysis

Bibliographic coupling analysis uses citations to give information about the similarities between two articles (or authors, or countries). This process relies on the assumption that two articles referencing a third article are highly related; and should be concentrated in a cluster solution of the visualization map. The strength of the bibliographic coupling is determined by the total number of references or citations of other third articles that they share. In this study, two bibliographic coupling analyses, i.e., bibliographic coupling of articles and countries, are performed. The first occurs when two articles cite the same third article; while the latter occurs when publications from two countries cite the same publication from a third country. Figure II-8 illustrates bibliographic coupling of articles; and Figure II-9 represents complex maps of many clusters reflects the diversity and interconnectedness of work being published from various countries. The largest cluster of bibliographic coupling of articles is anchored by Johnes (2006), which is considered as one of the earliest articles which incorporated ICT into the

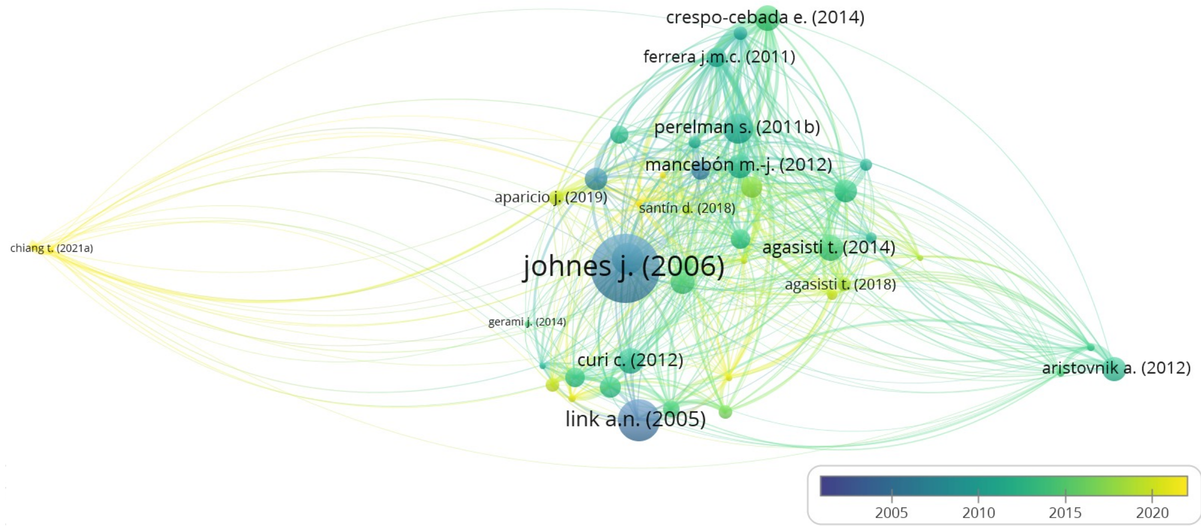


Figure II-8 Bibliographic coupling of articles

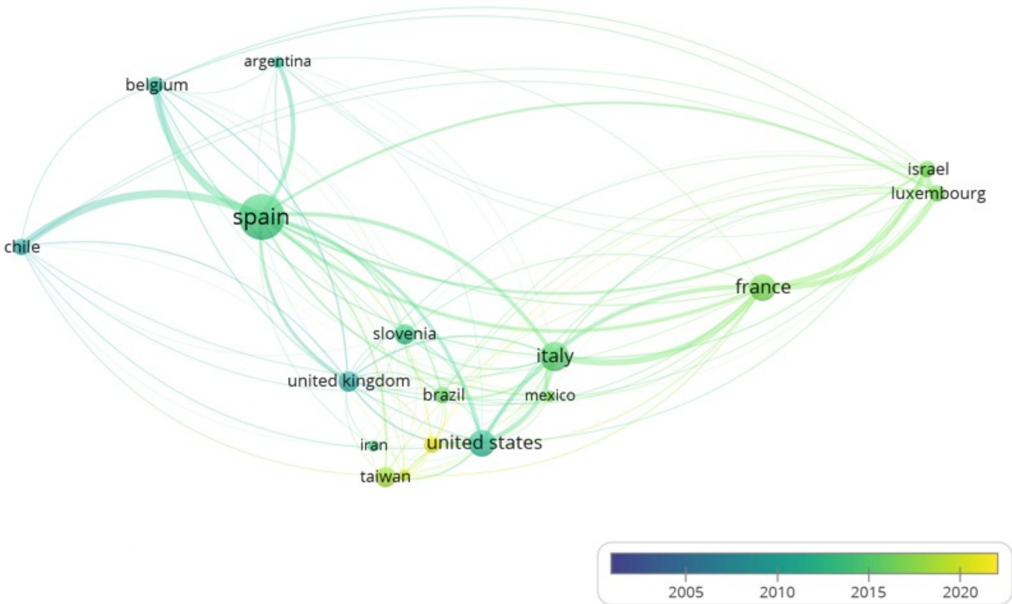


Figure II-9 Bibliographic coupling of countries

model of assessing efficiency of education. In the bibliographic coupling of countries, there are 17 countries contributed to this literature review, with Spain contributed the most, followed by Italy and France. The largest cluster is anchored by Spain; it suggests that Spain has a central influence in this research domain than those other countries are coupled to. The figure also illustrates frequent coupling among other countries, such as France, Italy, the United States, and the United Kingdom.

Table II-1 Level of analysis

Level of Analysis	Observed in
Student	Chiang (2021a, b), Deutsch et al. (2013), Ferrera et al. (2010, 2011), Mongan et al. (2011), Perelman and Santín (2011a, b)
School	Agasisti (2011, 2013), Agasisti and Zoido (2018, 2019), Aparicio et al. (2019), Cordero et al. (2017), Crespo-Cebada et al. (2014), Deutsch et al. (2019), Garcia-Diaz et al. (2016), Gerami et al. (2014), Mancebón et al. (2012), Primont and Domazlicky (2006), Salas-Velasco (2020)
Class (in a school)	Santín and Sicilia (2018)
University	Johnes (2006, 2008, 2013), Ruth et al. (2019), Zoghbi et al. (2013)
University TTO	Curi et al. (2012), Huang et al. (2011), Kim (2013), Link and Siegel (2005), Sutopo et al. (2019)
Region or province	Alves and De Araújo (2018), Aristovnik (2014), Chen et al. (2020)
Country (education system)	Agasisti (2014), Aristovnik (2012, 2013), Carlota and Ignacio (2021), Giménez et al. (2007), Thieme et al. (2012)

II.4 Classification of Articles

Each article is classified according to the levels of analysis, data source, and methods used. Moreover, the role of ICT in the model is also reviewed. Classifying extracted articles from the literature review allows readers to validate what has been studied and also can provide the possibility to find gaps in this research domain.

II.4.1 Level of analysis

The first classification is based on the level of analysis. The analysis of efficiency is conducted at different levels as shown in Table II-1. Most studies (13 studies) focused their analysis at the school level, while only one study was conducted at the class level. Studies investigating the efficiency of university Technology Transfer Office (TTO) are also included as this unit is under the responsibility of the university.[4] Six studies assessed the efficiency at the country level—the highest level of aggregagation, while eight studies are at the lowest level, i.e., student level. Student level of analysis involves a great advantage over other level of analysis since it provides information on students' efficiency independently of their educational system or school; furthermore, the efficiency measurement allows considering separately student's own socio-economic background and their schoolmates' one (or the so-called peer-group effect), two inputs which cannot be simultaneously included with aggregated data.

[4] TTO is a kind of organization that assist university in managing its intellectual assets in ways that facilitate its transformation into a benefit for society (Carlsson and Fridh, 2002). One of the common strategies is technology commercialization.

Table II-2 Data source

Data Source	Observed in
Authors' survey (using questionnaire)	Chiang (2021a, b), Gerami et al. (2014), Ruth et al. (2019), Sutopo et al. (2019)
National database (e.g., from the Ministry of Education, Statistics Agency, etc.)	Alves and De Araújo (2018), Chen et al. (2020), Curi et al. (2012), Garcia-Diaz et al. (2016), Huang et al. (2011), Johnes (2006, 2008, 2013), Kim (2013), Link and Siegel (2005), Mongan et al. (2011), Primont and Domazlicky (2006), Santín and Sicilia (2018), Zoghbi et al. (2013)
International database: a. PISA b. TIMSS c. PIRLS d. Others	 a. Agasisti (2011, 2013, 2014), Agasisti and Zoido (2018, 2019), Aparicio et al. (2019), Aristovnik (2012, 2013, 2014), Carlota and Ignacio (2021), Crespo-Cebada et al. (2014), Deutsch et al. (2013, 2019), Ferrera et al. (2010, 2011), Mancebón et al. (2012), Perelman and Santín (2011a, b), Salas-Velasco (2020), Thieme et al. (2012) b. Giménez et al. (2007) c. Cordero et al. (2015) d. Aristovnik (2012, 2013, 2014)

II.4.2 Data source

The choice of the level of analysis might depend on the availability of the data source, see Table II-2 for the data source used by articles in this review. The international database such as PISA, contains information at the student level that can be utilized by researchers. The data at the student level also can be aggregated at the school level (or perhaps at the country level) by using the weight at the appropriate level of aggregation to guarantee representativeness of the selected sample of students at school or country level. Majority of the studies used the international database, such as PISA (20 studies), TIMSS (1), PIRLS (1), and others, i.e., Eurostat, OECD Education at a Glance, and World Bank. Using this international database allows for international comparison. Researchers might compare efficiency across countries. Apart from the international database, there are also national databases (used by 15 studies), which is provided by, for instance, Ministry of Education of such country or National Statistics Agency. Lastly, five studies conducted independent surveys that were intended solely to accomplish the objective of their research.

II.4.3 Method used

The next classification provides the methods to measure efficiency. Although there are a significant number of studies on the efficiency measurement in education, efficiency is mainly measured by frontier methods. Bogetoft (2012) provided a taxonomy of the frontier methods and showed that the methods can be categorized into four classes: (i) parametric-deterministic, i.e., corrected ordinary least squares (COLS); (ii) parametric-stochastic, i.e., SFA; (iii) non-parametric-deterministic, e.g., DEA; and (iv) non-parametric-stochastic (stochastic DEA). Parametric approach is characterized by being defined a priori except for a finite set of unknown parameters that are estimated from the data. On the other hand, non-parametric approach is less restricted a priori. In the stochastic models, it is allowed that the

Table II-3 Methods used

Methods Used	Observed in
Non-parametric-deterministic: a. DEA	a. Aparicio et al. (2019), Aristovnik (2012, 2013, 2014), Carlota and Ignacio (2021), Chiang (2021a), Gerami et al. (2014), Giménez et al. (2007), Huang et al. (2011), Johnes (2008, 2013), Kim (2013), Primont and Domazlicky (2006), Ruth et al. (2019), Sutopo et al. (2019), Thieme et al. (2012)
b. DEA with bootstrapping procedure	b. Aparicio et al. (2019), Johnes (2006)
c. Two-stage DEA	c. Agasisti (2011, 2013, 2014), Agasisti and Zoido (2018, 2019), Curi et al. (2012), Deutsch et al. (2019), Ferrera et al. (2010), Santín and Sicilia (2018)
d. DEA with hierarchical linear model	d. Mancebón et al. (2012)
e. Fuzzy DEA	e. Aparicio et al. (2019), Chiang (2021b)
f. Order-*m* frontier with bootstrapping procedure	f. Cordero et al. (2015)
Non-parametric-stochastic: g. CCDEA	g. Aparicio et al. (2019)
Parametric-deterministic: h. Two-stage COLS with distance function	h. Deutsch et al. (2013)
Parametric-stochastic: i. SFA	i. Alves and De Araújo (2018), Link and Siegel (2005), Mongan et al. (2011), Zoghbi et al. (2013)
j. SFA with distance function	j. Crespo-Cebada et al. (2014), Ferrera et al. (2011), Johnes (2013), Perelman and Santín (2011a, b)
k. SFA with panel data	k. Chen et al. (2020), Garcia-Diaz et al. (2016), Salas-Velasco (2020)

individual observations may be affected by random (statistical) noise. In the deterministic models, the noise is suppressed, and any variation in the data is considered to contain significant information about the efficiency and the shape of the technology.

In this review, the majority of the articles (32) used non-parametric method, see Table II-3. The findings of De Witte and López-Torres (2017), Rhaiem (2017), and Worthington (2001) also confirmed this result, implying that this method has been preferred to other methods in measuring efficiency of education. The most popular non-parametric-deterministic methods in this review is DEA, proposed by Charnes et al. (1978). The application of DEA is common due to its flexibility and simplicity. It can handle multiple outputs and inputs more simply, and as a non-parametric approach, it does not require any assumption about the functional form. However, one of the shortcomings of DEA is that it cannot determine factors that might affect efficiency (called the determinants of efficiency). This motivates the development of two-stage DEA. In the first stage, the efficiency score is calculated and in the second stage, the explanatory factors potentially affecting the efficiency is identified. In this review, two approaches are used in the second stage, the ordinary least squares (OLS) regression (found in e.g., Agasisti, 2014) and Tobit regression (in Agasisti, 2013 and Ferrera et al., 2010).

Some scholars extended the *deterministic* DEA to be *stochastic*. The review by Olesen and Petersen (2016) pointed out the existence of three main stochastic approaches to DEA: (i) the first direction extends DEA to be able to handle estimated deviations as random deviations; (ii) the second direction extends DEA to be able to handle random noise in the form of either measurement errors or specification errors; and (iii) the third direction extends DEA to be able to regard or conceive the production possibility set as a random set, based on the random variation in the data. In this review, only Aparicio et al. (2019) used this approach. The authors followed the third direction as they used chance constrained DEA (CCDEA) by Cooper et al. (1998) as an additional robustness check. In this model, they assumed that inputs and outputs follow the normal distribution. They observed that the correlation coefficient between efficiency scores obtained from CCDEA and traditional DEA was statistically significant and high (more than 0.7).

Other lesser popular method in the deterministic approach is the use of the corrected OLS (COLS), which is parametric. In this review, it is only found in Deutsch et al. (2013). The authors argued that the maximum likelihood approach used in the SFA does not always converge. In addition, the COLS procedure is easy to implement and generates an estimated production frontier that lies on—at least one producer—or above the data. However, this simplicity comes at a cost: the estimated frontier is parallel to the OLS regression since only the intercept is corrected. It implies that the structure of the frontier is the same as the structure of the central tendency. Moreover, another issue in the COLS is that the statistical disturbances of the frontier function cannot be distinguished from the inefficiency effect of the model; and therefore, it is impossible, in general, to allow for both inefficiency and statistical error in the model.

One of the biggest shortcomings of the deterministic approach is that it assumes all deviations from the frontier are because of inefficiency. It means that it does not distinguish inefficiency from another factor, such as statistical noise, resulting in that it may overestimate the level of inefficiency. Contrarily, in the SFA, introduced independently by Aigner et al. (1977) and Meeusen and van den Broeck (1977), as a stochastic approach, the drawback can be avoided since it does differentiate the deviation as inefficiency and statistical disturbance. In addition, as a parametric approach, one can observe the effect of inputs on the outputs. In this review, there are four articles that used this approach. The drawback of the "ordinary" SFA that cannot deal simultaneously with multiple outputs can be handled flexibly by the distance function, as demonstrated by e.g., Perelman and Santín (2011a, b). The opportunity to analyze panel data to control unobserved heterogeneity further escalates SFA's attractiveness over its non-parametric counterpart. Different from cross-sectional data, panel data contains more information since the unit of analysis (e.g., school or university) is observed repeatedly. Another benefit is that it enables the researcher to observe whether inefficiency has been persistent over time or time-varying. In this review, SFA with panel data is only observed in three articles, i.e., Chen et al. (2020), Garcia-Diaz et al. (2016), and Salas-Velasco (2020).

Apart from those advantages, SFA has a strong assumption about the functional form and technical relationship among inputs and outputs. There are several functions that associate inputs and outputs, i.e., production, cost, revenue, and profit functions (Kumbhakar and Lovell, 2000). In this review, only production function was used; perhaps due to limited information about the expenditure or cost. Eight articles (observed in Chen et al., 2020; Salas-Velasco, 2020; Crespo-Cebada et al., 2014; Johnes 2013; Ferrera et al., 2011; Perelman and Santín, 2011a, b; Link and Siegel, 2005) used translog production function due to its highly flexible nature, which allows the study of interactions in the production process. However, the presence of quadratic and interaction terms in the translog form do not make the results simple to interpret (Felipe, 1988; Johnes and Johnes, 2009); this makes others used other form, such as Cobb-Douglas (observed in Alves and De Araújo, 2018 and Mongan et al., 2011) and linear specifications (observed in Garcia-Diaz, 2016 and Zoghbi et al., 2013).

II.5 ICT-Related Variables

ICT-related variables are categorized as outputs, inputs, and determinants of (in)efficiency, see Table II-4. The selection of variables has been guided by the existent literature, e.g., De Witte and López-Torres (2017) and was constrained by data availability (or data source). More specifically, the educational process is modelled as each unit of analysis (e.g., student, school, or even country) receives a given number of inputs and use them for "producing" as much output as possible.

II.5.1 Output

The output is in general related to the achievements obtained by the students (e.g., test score) or by researchers (e.g., number publications, citations, research grants, patents). However, as observed, it is uncommon to use ICT as output (only two articles in this study). Curi et al. (2012) who assessed the efficiency of TTO operated by French university used number of software applications as an output. Ruth et al. (2019) used number of students enrolled resulting from the use of information system from 2013 to 2017 as an output when assessed the efficiency of technical universities in Ghana. Whereas the use of ICT for student enrollment system is very common in developed countries, however, it is a challenge in developing (or under-developed) countries due to poor infrastructure.

II.5.2 Input

In this study, input is categorized into four levels: (i) student level, (ii) family level, (iii) institution level, and (iv) country level, see Table II-4. In the student level, student engagement with

Table II-4 The role of ICT

The Role of ICT	Observed in
Outputs: a. Number of software applications b. Number of students enrolled resulting from IS usage	a. Curi et al. (2012) b. Ruth et al. (2019)
Inputs: • Student level: c. student engagement with ICT d. time to watch TV, radio, and visit website • Family level: e. resources available at home • Institution level: f. ratio of computers to the number of students g. ratio of computers connected to the internet h. internet available i. frequency of studens who use computers/educational softwares j. quality of school resources k. ICT training l. invention disclosures m. expenditure on ICT • Country level: n. expenditure on ICT o. internet users per 100 people p. international internet bandwidth q. number of computers per 100 students	c. Chiang (2021a, b) d. Deutsch et al. (2013) e. Deutsch et al. (2013, 2019), Giménez et al. (2007), Mongan et al. (2011), Perelman and Santín (2011a) f. Agasisti and Zoido (2018, 2019), Alves and De Araújo (2018), Carlota and Ignacio (2021), Deutsch et al. (2019), Mancebón et al. (2012), Perelman and Santín (2011b), Primont and Domazlicky (2006), Ruth et al. (2019), Zoghbi et al. (2013) g. Agasisti (2011, 2013), Deutsch et al. (2013), Salas-Velasco (2020) h. Garcia-Diaz et al. (2016), Sutopo et al. (2019) i. Mancebón et al. (2012) j. Aparicio et al. (2019), Crespo-Cebada et al. (2014), Deutsch et al. (2013), Ferrera et al. (2010, 2011), Giménez et al. (2007), Perelman and Santín (2011a), Ruth et al. (2019), Salas-Velasco (2020), Santín and Sicilia (2018), Thieme et al. (2012) k. Ruth et al. (2019) l. Kim (2013), Huang et al. (2011), Link and Siegel (2005) m. Gerami et al. (2014), Johnes (2006, 2008, 2013), Ruth et al. (2019), Thieme et al. (2012) n. Aristovnik (2012, 2013) o. Aristovnik (2012, 2013) p. Aristovnik (2012) q. Cordero et al. (2015)
Determinants of (in)efficiency: r. quality of school resources s. internet users t. resources available at home	r. Agasisti and Zoido (2019) s. Chen et al. (2020) t. Agasisti (2014), Deutsch et al. (2013)

ICT was used in Chiang (2021a, b). The author gathered the information using questionnaire collected from students. “I do not use cellphone in the course” and “I learn from YouTube or others” are two examples of the item questions. The authors argued that the problem of using mobile phones in the course still affected the learning process. In addition, self-discipline and adherence to classroom norms, such as arriving in class on time and not using mobile phones, can improve learning efficiency. Deutsch et al. (2013) utilized PISA database to collect information about time to informal learning, such as watching TV programs, visiting websites, and listening to radio programs.

Move to the family level, Deutsch (2013, 2019) and Mongan et al. (2011) used educational resources available at home as an input, including a computer used for study, educational software, and internet access. Giménez et al. (2007) used percentage of students with computer at home. A different perspective was found in Perelman and Santín (2011a) who used possession of video console (Play-Station, X-Box, or similar) at home.

In the institution level, ten articles used ratio of computers to the number of students, while four articles used ratio of computers connected to the internet as inputs. Garcia-Diaz et al. (2016) and Sutopo et al. (2019) used a binary data of availability of the internet. Mancebón et al. (2012) utilized frequency of students who use computer frequently or occasionally to write documents. Mostly (11 articles), scholars used quality of school resources. In PISA, this variable is an index derived from the responses of school principals to seven items related to the availability of educational resources, such as computers for teaching purposes, educational software, calculators, books, audiovisual resources, and laboratory equipment. In articles discussed about measuring the efficiency of TTO, invention disclosure—as a proxy for the set of available technologies for licensing—is used as input (observed in Huang et al., 2011, Kim, 2013, and Link and Siegel, 2005).

In the country level, Aristovnik (2012, 2013) used ICT expenditure as a percentage of GDP from the World Bank and the number of internet users per 100 people as inputs. Aristovnik (2012) used international internet bandwidth (bits per person) and Cordero et al. (2015) used number of computers per 100 students.

II.5.3 Determinants of (in)efficiency

Not only as a driver of output, but one may also want to investigate the role of ICT as determinants of (in)efficiency, i.e., factors that might affect (in)efficiency. In the literature of efficiency measurement, there are two distinct terms used due to different model specification. In the nonparametric data envelopment analysis, scholars use "*determinants of efficiency*" as the factors that can explain efficiency; on the other hand, in the parametric stochastic frontier analysis, the term "*determinants of inefficiency*" is used instead. In this review, only four articles which used ICT as the determinants of (in)efficiency.

In the SFA, ICT can be incorporated into the model as the determinants of inefficiency using the heteroscedastic model. In this review, only one article, i.e., Chen et al. (2020), who used the degree of internet penetration or the ratio of internet users to the local population as determinant of inefficiency.

Since in the deterministic approach one cannot identify which factors affect efficiency, scholars used two-stage DEA by using regression analysis to handle this issue. Deutsch et al. (2013) used material wealth of parents, including possession of cellular phone, TV, and computer. Agasisti (2014) used proportion of students who have regular access to the internet at school and at home as a proxy for the digital literacy of the student population. Agasisti and Zoido (2019) used index of adequateness of instructional material's quality. This index includes shortage of computers and educational software for educational purposes and internet connectivity.

II.6 The Influence and Significance of ICT-Related Variables

This section describes the influence and significance of the ICT-related variables used in the model of efficiency measurement in education.

II.6.1 The influence of ICT as input

In the parametric approach, the relationship between inputs and output, the significance of inputs through the production function, as well as the contribution to the output controlled for other inputs can be investigated.

Alves and De Araújo (2018) reported the number of computers for students was statistically significant at the level of 5% and had positive influence on output measured by the index of development of basic education. They observed that an increase of 1% in the number of computers at school would result in an increase of 0.02% in the index of development of basic education, while holding other predictors constant. However, this variable was found to be not significant in Perelman and Santín (2011b) to influence the academic outcomes (measured by the PISA scores, the OECD education assessment at the age of 15-years old). The ratio of computers connected to the internet also was found not statistically significant in Salas-Velasco (2020). School educational resources as observed in Crespo-Cebada et al. (2014) was statistically significant only in public schools from Navarre, Spain, but not in other regions in Spain. The insignificant condition also was reported in Perelman and Santín (2011a). However, Salas-Velasco (2020) found that this variable had significant positive value to influence the PISA score; while the coefficient found in Ferrera et al. (2011) had negative value. This condition corroborates the findings of previous research which were inconclusive regarding the role of school resources on the academic performance. Some studies showed a positive influence (e.g., Carroll, 1963; Krueger 1999), but others found that there was no direct correlation between more school inputs and better academic outcomes.

At the family level, Mongan et al. (2011) showed that resources available at home are statistically significant at the level of 1%—with a positive value—to affect student's achievement in Argentina measured by standardized test in language. The importance of household characteristics was already revealed by the well-known Coleman Report (Coleman et al., 1966), which found that the difference in school results in the United States was due more to cultural and socioeconomic reasons than to the allocation of educational resources.

On the other hand, in non-parametric approach, the influence of inputs on outputs cannot be investigated; and therefore, the selection of inputs only depends on the literature without knowing whether the selected inputs significantly affect the output(s) or not. However, Ferrera et al. (2010, 2011) argued that inputs must fulfil the requirement of isotonicity (i.e., ceteris paribus, more input implies an equal or higher level of output); thus, the selected inputs should present a significant positive correlation with the output(s) in addition to having theoretical support from previous works. Accordingly, some authors provided a correlation analysis between inputs used and outputs given which will be discussed in the following. Agasisti (2013) showed that the proportion of computers connected to the web is weakly correlated with science and mathematical literacy measured by the PISA scores. Chiang (2021a) reported that student engagement with ICT had strong correlation with interdisciplinary skills, moderate correlation with reflective behavior and recognizing disciplinary perspectives; and very weak correlation with incentive outcomes. Huang et al. (2011) stated that invention disclosure has positive moderate correlation with license income as a solely output in their model. Aristovnik (2012, 2013, 2014) used partial correlation between different variables while controlling for the other variable(s). For instance, PISA scores showed a weak and positive (but not statistically significant) correlation with ICT expenditures (in % of GDP) when controlling for the number of internet users (Aristovnik 2013).

Other approaches had been used beside (partial) correlation analysis as observed in Johnes (2006) and Mancebón et al. (2012). Johnes (2006) applied the test from Pastor et al. (2002) for assessing the relevance of input(s) and/or output(s) included in the DEA. The test is useful, to an extent, in reducing an input-output set to a smaller "significant" set. Mancebón et al. (2012) used the hierarchical linear model (HLM) for selecting the relevant inputs. The result extracted from the HLM regression permits inputs for the subsequent DEA efficiency analysis to be selected in a robust empirical fashion.

II.6.2 The influence of ICT as determinant of (in)efficiency

In this study, only four studies used ICT-related variables as determinants of (in)efficiency, one used the SFA (Chen et al., 2020), and the rest used two-stage DEA (Agasisti, 2014; Agasisti and Zoido, 2019; Deutsch et al., 2013). The degree of internet penetration was used in Chen et al. (2020) as a determinant of inefficiency in their stochastic frontier model. However, this variable was not statistically significant, meaning that expecting to reduce inefficiency by altering the degree of internet penetration is not obvious.

Deutsch et al. (2013) who measured the efficiency of students in Brazil, Chile, Colombia, Mexico, and Uruguay, used material wealth of parents, including possession of cellular phone, TV, and computer, as one of determinants of efficiency in their two-stage DEA model. This variable was only significant—

with negative value—in Chile; while in other countries it was not. Agasisti (2014) reported that proportion of students who have regular access to the internet at school and at home was significant with positive value. Since the coefficient value was 0.0012, it means that increasing the percentage of students who have access to the internet would increase the efficiency score by 0.12%. Agasisti and Zoido (2019) reported that index of school educational resources was statistically significant with small positive value (the coefficient value was 0.0054).

II.7 Discussion and Concluding Remarks

This study systematically reviews articles from the Scopus database on the role of ICT in the efficiency of education. Through the PRISMA procedure, the review collects 41 articles published in 30 different journals. The collected articles are categorized according to three classifications, i.e., the levels of analysis, the data source, and the method used. According to the level of analysis, most studies were conducted at the school level. This finding is a bit different with the findings of others, more general, review studies in the field of efficiency in education, i.e., De Witte and López-Torres (2017) and Rhaiem (2017) in which the majority of the articles are collected at the level of university. The level of analysis might depend on the availability of the data source. Majority of the studies used the international database, such as PISA, TIMSS, and PIRLS, since these databases contain several ICT-related data that can be exploited, such as ICT resources available at home, ratio of computers at school connected to the internet, time spent of students using digital devises outside school, etc. These databases contain information at the student level, school level, and at some extent, the information about the parents; therefore, using these databases benefit the scholars to conduct the analysis at the student or school level. Moreover, these databases also allow for international comparisons as they are well-regarded as authoritative sources of comparison for educational achievement across the world. Literature about international comparisons of efficiency in education is in its very infancy and it is probably one of the most interesting trends of research in the future. Agasisti and Zoido (2019) observed two main reasons behind this limited development of studies, i.e., the lack of reliable datasets and the substantial differences in institutional (country-specific) settings. Using these internationally comparable databases, one could deal with both of those limitations.

This study also provides insights on the role of ICT in efficiency of education. ICT is categorized as output, input, and determinant of (in)efficiency. It is observed that it is uncommon to use ICT as output, primarily because education outcomes are commonly represented as student's or researcher's achievements. On the other hand, most of the articles used ICT as input, which in this review, it is divided into four levels: student, family, institution, and country level. Interestingly, no conclusive empirical evidence has been found on the influence of ICT on education outcomes. Inconclusive result also has been found on the influence of ICT on (in)efficiency. These findings follow De Witte and

Rogge (2014) who argued that ICT is like a double-edged sword: it could have a positive influence on education in some extent; but also, might, say, distract the students in the learning process—as the negative influence. The classic argument of Hanushek (1996, 2003) is about to be highlighted that putting more money into schools does not guarantee per se better education outcomes—in this context the money can be considered as an investment in ICT, such as installing more computers and providing the internet. In the end, we must carefully examine the use of ICT in educational setting as most scholars agree that when properly used, ICT holds great promises to improve teaching and learning in addition to shaping workforce opportunities (Mobi et al., 2015).

The next issue is about the *concept* of ICT. When one tries to estimate the influence of ICT on education outcomes and (in)efficiency, at first, s/he needs to define this concept (i.e., ICT) and how to properly measure it. The big question is: what is ICT? Does it refer to ICT infrastructures or to the actual use? Is the location of the infrastructures (i.e., at school, at home) and what students do (e.g., for school-related tasks or for entertainment purposes) when they use ICT relevant? Is the intensity of use (e.g., every day, one a week, twice a week) an important factor? In this review, majority of the articles which used ICT referred to the physical infrastructure, e.g., number of computers at school, internet availability, resources available at home. It is of interest to see the influence of *other type* of ICT beside physical infrastructure, e.g., the ICT use. Students could use ICT at school or at home (e.g., for educational purposes or just for leisure activities). The international database such as PISA provides this information that can be used.[5] However, things might become even more complex when trying to utilize ICT infrastructure and ICT use because it can give rise to different education outcomes due to the interplay of many factors (e.g., the degree of ICT confidence of the teachers, students and parents, the accessibility of ICT resources at home, school or other relevant environment, peer effects, etc.) (Biagi and Loi, 2013).

Lastly, the method used to measure efficiency is being addressed. Notice that majority of the studies did not employ the recent methods in measuring the efficiency. Therefore, it is highly recommended to use the recent methods for the future research. First, the recent development in the non-parametric method is discussed followed by the parametric method.

Following the findings of De Witte and López-Torres (2017), Rhaiem (2017), and Worthington (2001), the application of the non-parametric method is preferred to other methods. However, as has been mentioned previously, if one desires to investigate factors that might affect efficiency, the two-stage DEA should be used. Here, the estimated efficiency scores are regressed, in an appropriated limited dependent variable parametric regression model on the determinants of efficiency. As pointed

[5] For instance, in PISA 2018 data, there are three variables related to ICT use, i.e., ENTUSE, HOMESCH, and USESCH, which are defined as the indices of ICT use outside of school for leisure activities, ICT use outside of school for school-work activities, and ICT used at school in general, respectively.

out by Simar and Wilson (2007), this procedure is flawed by the fact that usual inference on the obtained estimates of the regression coefficient is not available (see also Daraio and Simar, 2005, for the discussion). The bootstrap procedure is then proposed to obtain more accurate inference.

Another issue is when DEA delivers many efficient decision-making units (DMUs)—it can be school, university, or country. It is then difficult to differentiate among those best performers. The super-efficiency model has this capability. When the *traditional* DEA gives scores 1 (one) for the most efficient DMU; in the super-efficiency model, the DMU allows to have efficiency score more than 1 (one). Accordingly, this particular DMU regards as *super-efficient*. This approach is actually proposed to handle the fact that DEA suffers from being highly vulnerable to potential outliers and measurement error, because every unit is related to the most efficient units. A simple idea to correct for potential outliers was proposed by Andersen and Petersen (1993), in which their model excludes each DMU from its own reference set, so that it is possible to obtain efficiency score that exceeds one. Accordingly, Tone (2002) proposed the slacks-based measure of super-efficiency in DEA that has the important properties, such as unit invariant, monotone decreasing, translation invariant, and reference-set dependent. Therefore, in the future, it is recommend combining the super-efficiency model and the bootstrap procedure to both measure efficiency in a condition when there are many efficient DMUs as well as to investigate the determinants of efficiency.

Apart from the advantages of the DEA, it also has several drawbacks that have been addressed previously. In particular, DEA could not handle panel data easily as in the SFA. Panel data contains more information since the unit of analysis (e.g., school or university) is observed repeatedly. A recent development in the SFA dealing with panel data is the "four-component stochastic frontier model". This model was simultaneously proposed by Colombi et al. (2014), Kumbhakar et al. (2014), and Tsionas and Kumbhakar (2014). This model separates producer effects, persistent and time-varying inefficiency, as well as statistical noise. Accordingly, this model disentangles overall inefficiency into two parts: persistent and time-varying inefficiency. The persistent inefficiency refers to a long-term or structural inability of an education institution to achieve the potential level of academic outputs. Time-varying inefficiency, on the other hand, is a short-run deficit which can be eliminated swiftly without a major structural change. Distinguishing between persistent and time-varying inefficiency is important since they may have different policy implications (Lai and Kumbhakar, 2018). In this review, only three articles are dealt with panel data setting, while unfortunately, only one article used the "four-component stochastic frontier model" (i.e., Salas-Velasco, 2020).

CHAPTER III. ICT IN THE EFFICIENCY MEASUREMENT OF EDUCATION SECTOR IN SOUTH-EAST ASIA: THE STOCHASTIC FRONTIER ANALYSIS APPROACH

This chapter presents a study which primarily aims to measure efficiency of education sector in the South-East Asia (SEA) region. SEA is the geographical south-eastern region of Asia, consisting of the regions that are situated south of China, east of the Indian subcontinent, and northwest of Australia. There are eleven countries in this region, i.e., Brunei Darussalam, Cambodia, East Timor, Indonesia, Laos PDR, Malaysia, Myanmar, Philippines, Singapore, Thailand, Vietnam. Ten of these eleven countries are members of the Association of Southeast Asian Nations (ASEAN), while East Timor is an observer country.[6] The region covers more than 4.5 million km^2 with a combined total population of more than 680 million people (May 2022 est.), about 8.5% of the world's population. The number of populations varies from the fourth world populous country like Indonesia with more than 270 million people to the least populace country like Brunei Darussalam (about 430 thousand people). The region is culturally and ethnically diverse, with hundreds of languages spoken by different ethnic groups.

Even though China remains the goliath of emerging markets—with every fluctuation in its GDP making headlines around the globe—investors are increasingly turning their gaze southward to the ten (plus one) dynamic markets that make up the ASEAN (Vinayak et al., 2014). Within emerging economies, SEA represents one of the fastest growing regions (Asian Development Bank, 2011; World Bank, 2010). The economies are at vastly different stages of development, but all share immense growth potential. ASEAN is a major global hub of manufacturing and trade, as well as one of the fastest-growing consumer markets in the world. In 2018, eight of the ASEAN members are among the world's outperforming economies, with positive long-term prospect for the region (Woetzel et al., 2018). In addition, ASEAN's Secretariat projects that the regional body will grow to become the world's fourth largest economy by 2030 (Gronewold, 2019).

Paralleling these economic developments, there is an extensive interest in the characteristics of the education system they have developed. In particular, scholars have hypothesized high levels of student's performance in this region, the recent developments of education policies, and the contemporary debates and policy issues in these countries (Cheng, 1999). Most of the policy-makers and the public in these countries have been aware of the importance of education to the development of their societies and have initiated important policies to expand and improve their education systems. For

[6] Founded in 1967, ASEAN is a political and economic union of ten member countries and one observer country located in Southeast Asia, which promotes intergovernmental cooperation and facilitates economic, political, security, military, educational, and sociocultural integration between its members and countries in Asia-Pacific.

instance, in Singapore, as the most developed country in SEA, significant education reforms were introduced in 1987 as the Singapore government embarked on a number of reform initiatives to diversify educational provisions and deregulate its school system. The diversification and deregulation policy were used to shift Singapore's schools from the centrally controlled and homogenous school system to an educational system that would provide new elective subjects and enrichment programs; and the policy was followed by a number of new school reforms and initiatives (Abu-Bakar et al., 2006). In Indonesia, the national education system is carried out universally, open to every citizen, regardless of their geographic location, race and ethnicity, religion, socio-economic background, and it addresses the different needs of people at various stages of societal development (Purwadi and Muljoatmodjo, 2000). Malaysia believes that education plays a vital role in achieving the country's vision of attaining the status of a fully developed nation in terms of economic development, social justice and spiritual, moral and ethical strength, towards creating a society that is united, democratic, liberal and dynamic (Ministry of Education Malaysia, 2008). Among the nine strategies for implementing education reform in Thailand, two are related to the promotion of education quality and expansion of lifelong educational opportunity (SEAMEO Secretariat, 2001). Quality improvement has become the ultimate goal in the provision of education in Thailand in addition to maintenance of equity and social justice. The government believes that success in terms of equity in education without quality will not enable Thai people to thrive in a knowledge-based economy and society (Office of the Educational Council, 2004). The government therefore is committed to provide equal access to lifelong education and training to all Thai citizens to ensure that they will be equipped with necessary basic life skills and be employed. In the Philippines, "the State shall protect and promote the right of all citizens to quality education at all levels and shall take appropriate steps to make such education available to all" (Art. XIV, Sec. 1 of Special Education Act of 2004, introduced by Senator Jinggoy Ejercito Estrada) (Ballestamon et al., 2000). As the smallest country in terms of population, Brunei Darussalam clearly sets out its education policy aiming at quality education for all. One of its primary goals is to provide a minimum of 12 years of education for every child, covering 7 years primary and pre-school, 3 years lower secondary school, and 2 years on upper secondary or in a vocational/technical college (Hamid, 2000).

In sum, promoting quality and equity education is a common policy for countries in this region regardless their different levels of development. The education scene in SEA is one of the most dynamic, entrepreneurial, and competitive in the world. The fast growing middle-class propels the inherently diverse ASEAN community to set educational standards for the relatively youthful and growing population that are globally competitive and yet locally grounded. One of ASEAN's goals is to establish a "common space for higher education" and efforts to promote trans-national education to enhance human resource development that foster greater economic, social, and political integration among the members.

Expectations of the country, society, media, and other stakeholders stimulate education institutions to manage their resources more efficiently; hence, countries in SEA are currently under a growing pressure to increase efficiency and improve the quality of its activities. However, comparative studies—especially in education—that critically examine this region has yet to be achieved (Symaco and Chao, 2019); and despite the promising analysis of educational issues in one of the most compact regions in the world, published studies about measuring efficiency in the education sector in this region is quite limited—see Chapter III.1 for more details. Therefore, this study attempts to close the gap in the literature by conducting a cross-country analysis of measuring efficiency in education in this region.

As this study allows for international comparisons, it is expected as a means for benchmarking, comparing education systems, against a set of counterparts in several different countries. It enables not only schools, but also educational system in such country, to observe different combinations of inputs and outputs beyond those that are typical in a given country. It then would help to enlarge the perspective of each institution, helping the process of spurring innovation and new ideas of how to employ resources more efficiently. To accomplish the objective, this study uses the stochastic frontier analysis (SFA) allowing for heteroscedasticity using the recent OECD PISA 2018 data.

This chapter is structured as follows. In Chapter III.1, by reviewing previous studies, the contributions of this study are presented. In Chapter III.2, the empirical model of the efficiency measurement is displayed. Data and variables used in this study are presented in Chapter III.3. Results, including robustness analysis, are presented in Chapter III.4. Finally, Chapter III.5 provides discussion and concluding remarks.

III.1 Literature Review and Contributions

The literature about measuring efficiency of education especially in the SEA region is quite limited. A literature review is conducted in the Scopus database with the following search query: TITLE-ABS-KEY((Indonesia* OR Malaysia* OR Thai* OR Singapore* OR Brunei* OR Lao* OR Myanmar* OR Vietnam* OR Philippin* OR Cambodia* OR Timor*) AND efficien* AND (DEA OR data envelopment analysis OR SFA OR stochastic frontier analysis OR frontier) AND education). The period of time is not limited. Note that all countries in SEA are listed to cover articles which implemented the frontier methods, such as DEA and SFA, to measure efficiency in education in this region. Only articles written in English published in the peer-reviewed journals are included.

The search yields only eleven articles. The first screening is performed by reading the title and abstract to verify the relevance of the extracted articles. In this way, six articles are excluded since they did not discuss efficiency in the education sector, leaving only five articles. This low number of pertinent articles indicates that this research area is under-studied, especially in this SEA region. The second screening is executed by carefully reading the full-text of each article to address the eligibility of the articles. All five articles from the first screening are eligible to be further analyzed as follows.

Castano and Cabanda (2007) evaluated the efficiency performance of Philippine Private Higher Educational Institutions using DEA and SFA over the period 1999–2003. Lavado and Cabanda (2009) measured the efficiency of provinces in the Philippines in utilizing public resources for health and education with budget constraints. The input is social services expenditure per capita in each province and the outputs are health and education outcomes as life expectancy for health outcome as well as functional literacy rate and combined primary and secondary enrollment rates for education outcomes. Johnes and Virmani (2020) used data from the Young Lives study to evaluate the efficiency of education systems in four low- and middle-income countries: Ethiopia, India, Peru, and Vietnam. Using DEA, Salcedo (2020) evaluated the performance efficiency of the teacher education programs in seven campuses of the Pangasinan State University in Philippine from academic year 2012-2013 to 2014-2015. Le et al. (2021) investigated how well a province in Vietnam transforms the family expenditure in education into the achievements of students and suggest benchmarks for policies and investments of a provincial government on how it can improve the education system. They used inverse optimization in DEA-based benchmarking to accomplish the objective.

Detailed information of the eligible articles including methods used, level of analysis, inputs, and outputs are shown in Table III-1. Notice that the studies of measuring efficiency in the education sector were only implemented in the Philippines and Vietnam. SFA was only used in one study, corroborating the dominance of DEA over SFA in the measurement of efficiency literature.

This study contributes to the literature as follows. First, as the result of the previous literature review, the use of SFA for this kind of analysis is quite limited and this study attempts to extend the implementation of SFA in measuring efficiency in education sector in the SEA region. The use of SFA is motivated by the fact that, as a parametric approach, the SFA can investigate the influence of inputs on output, whereas a non-parametric approach cannot. Another advantage of SFA over its non-parametric counterpart is that SFA also can take into account factors that affect inefficiency (i.e., the determinants of inefficiency) in a one stage of calculation, called the heteroscedastic model of SFA. Second, in relation to the role of ICT in the efficiency of education, previous studies did not incorporate ICT into their models; thus, how ICT influences education outcomes (as well as inefficiency) cannot be investigated. Therefore, to investigate the influence of ICT on both education outcomes and inefficiency, this study includes ICT-related variables as inputs as well as determinants of inefficiency. Lastly, it can be considered as the first study which assesses the efficiency across country in the SEA region. Previous studies only focused on one specific country in the SEA region, while in this study, a more detailed analysis is provided in cross country level. Literature about international comparisons of school's efficiency is in its very infancy. Despite the intrinsic interest of international comparisons, there are two

Table III-1 Articles included in the literature review

Articles	Methods	Level of Analysis	Inputs	Outputs
Castano and Cabanda (2007)	DEA and SFA	University (in Philippines the)	• Number of faculty members • Property, plant, and equipment • Operating expenses	• Student enrollment • Graduates per year • Total revenue
Lavado and Cabanda (2009)	DEA	Province (in the Philippines)	Social services expenditure per capita	• Life expectancy • Functional literacy rate • Combined primary and secondary enrollment rates
Johnes and Virmani (2020)	DEA	Cross-country (Ethiopia, India, Peru, and Vietnam)	• The wealth index • Household expenditure per capita • Daily hours spent in class • Daily hours spent in private study • Highest grade completed • Student age	Student's score in the Peabody Picture Vocabulary Test
Salcedo (2020)	DEA	Study program (teacher educa-tion programs in the Philippines)	For curriculum: • Number of programs offered • Total number of units in each program • Total number of hours of teaching practice	For curriculum: • Number of accredited programs • Status of accreditation
Le et al. (2021)	DEA	Province (in Vietnam)	• Inside expenditure pa-id by families for their children's education to educational institutions • Outside expenditure paid by families for their children's education to educational institutions	• Math score in National High-school Graduation Exam • Vietnamese score in National High-school Graduation Exam

main reasons behind the limited development of studies looking to compare schools' results between countries, i.e., (i) the lack of reliable datasets and (ii) the substantial differences in institutional (country-specific) settings (Agasisti and Zoido, 2019). Therefore, to overcome both of these limitations, this study uses the OECD PISA data, which is well-regarded as an authoritative source of comparison for educational achievement across the world.

III.2 Method

This study is conducted at the school level. Study conducted at country level is only partially informative, thus; even though it helps in understanding the differences in average performances, it neither indicates how these performances were generated (i.e., precise characteristics of the educational sector), nor how performances are distributed within the country, (i.e., between schools). For all these reasons, it appears useful to investigate the efficiency of education not at a country level, but instead at the school level; in other words, considering how different are results obtained by different schools.

To estimate efficiency, the stochastic frontier analysis (SFA) allowing for heteroscedasticity is used. As the SFA is the parametric approach, it requires assumptions on the functional form and technical relationship among inputs and output. Some earlier papers assumed a translog production function due to its highly flexible nature (e.g., Perelman and Santín, 2011a, b; Salas-Velasco, 2020), which allows the study of interactions in the production process. However, the presence of quadratic and interaction terms in the translog form do not make the results simple to interpret (Felipe, 1988; Johnes and Johnes, 2009); this makes others used other forms, such as the Cobb-Douglas function (e.g., Alves and De Araújo, 2018; Mongan et al., 2011) and linear specification (e.g., Garcia-Diaz, 2016; Zoghbi et al., 2013). In this study, the linear specification is used since some of variables deal with negative numbers. The original data is kept without imposing any transformations to make it easier to interpret the results. The production function is then defined as follows:

$$y_i = \beta_0 + f(\mathbf{x}_i; \boldsymbol{\beta}) + v_i - u_i, \tag{III-1}$$

where y_i is the observed output for an observation i ($i = 1, 2, \dots N$), β_0 is an intercept, $f(\bullet)$ is the production function, $\mathbf{x}_j$ is a vector of inputs, $\boldsymbol{\beta}$ is the associated vector of parameters to be estimated, v_i is two-sided statistical noise, and u_i is nonnegative inefficiency.

Notice that inefficiency can be regarded as an error; thus, the composed error can be modelled as $\varepsilon_i = v_\mathrm{i} - u_i$. Equation (III-1) can be estimated in two steps. In the first step, assuming that v_i and u_i are distributed independently of $\mathbf{x}_i$, the ordinary least squares (OLS) procedure would provide consistent estimates of $\boldsymbol{\beta}$ but not β_0 because $E[\varepsilon_i] = -E[\varepsilon_i] \leq 0$, where $E[\bullet]$ is the expectation of a random variable. The second step involves the use of maximum likelihood procedure to estimate β_0 and the variances of v_i and u_i (σ_u^2 and σ_v^2, respectively); thus, distributional assumptions are required. Several scholars have proposed some pairs of distributional assumptions for v_i and u_i. While most agreed that v_i follows the normal distribution with zero mean and variance σ_v^2, the distribution of u_i differs across studies. Aigner et al. (1977) argued that u_i is generated from the half-normal distribution, while Meeusen and van den Broeck (1977) assumed u_i follows the exponential distribution. Other commonly adopted distributions are truncated normal (Stevenson, 1980) and gamma (Greene, 1980, 2003).

In this study, the pair of half-normal and normal distributions are used as it is a common practice in the literature of efficiency measurement in education, see e.g., Guarini et al. (2020), Kirjavainen (2012), Scippacercola and D'Ambra (2014). The probability density function of u_i is (Kumbhakar et al. 2015)

$$f(u_i) = \frac{\frac{1}{\sigma}\phi\left(\frac{u_i}{\sigma}\right)}{1-\Phi(0)} = \frac{2}{\sigma}\phi\left(\frac{u_i}{\sigma}\right) = 2\left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{u_i^2}{2\sigma^2}\right), \tag{III-2}$$

where $\phi(\bullet)$ and $\Phi(\bullet)$ are the probability density and probability distribution functions, respectively, for the standard normal distribution. The log likelihood function for each observation i is

$$L_i = -\ln\left(\frac{1}{2}\right) - \frac{1}{2}\ln\left(\sigma_v^2 + \sigma_u^2\right) + \ln\phi\left(\frac{\varepsilon_i}{\sqrt{\sigma_v^2 + \sigma_u^2}}\right) + \ln\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right), \tag{III-3}$$

where

$$\mu_{*i} = \frac{-\sigma_u^2\varepsilon_i}{\sigma_v^2 + \sigma_u^2}, \text{ and} \tag{III-4}$$

$$\sigma_* = \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2 + \sigma_u^2}. \tag{III-5}$$

The log-likelihood function is then the observational sum of Equation (III-3), which can be numerically maximized to obtain estimates of the parameters. After the parameters are estimated, the individual inefficiency and efficiency can be estimated. Jondrow et al. (1982)—later it is called JLMS estimator—proposed the conditional mean of u_i given ε_i as a point estimate of inefficiency u_i as follows

$$E\left[u_i\middle|\hat{\varepsilon}_i\right]_{\text{normal-half normal}} = \frac{\sigma_*\phi\left(\frac{\mu_{*i}}{\sigma_*}\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)} + \mu_{*i}, \tag{III-6}$$

where μ_{*i} and σ_* are defined in Equation (III-4) and (III-5), respectively. Next, the efficiency can be estimated as proposed by Battese and Coelli (1988)—later it is called BC estimator, as follows

$$E\left[\exp(-u_i)\middle|\hat{\varepsilon}_i\right]_{\text{normal-half normal}} = \exp\left(-\mu_{*i} + \frac{1}{2}\sigma_*^2\right)\frac{\Phi\left(\frac{\mu_{*i}}{\sigma_*} - \sigma_*\right)}{\Phi\left(\frac{\mu_{*i}}{\sigma_*}\right)}, \tag{III-7}$$

where μ_{*i} and σ_* are also defined in Equation (III-4) and (III-5), respectively. Maximum likelihood estimates of the parameters are substituted into Equation (III-7) to obtain empirical estimates of efficiency. The estimate then has a value between 0 and 1, with the value equal to 1 indicating full efficiency.

The previous model assumes that u_i and v_i are homoscedastic, that is, both σ_u^2 and σ_v^2 are constants. This is not capable to investigate the influence of determinants of inefficiency, i.e., factors that can explain inefficiency, labeled as $\mathbf{z}_i$, a vector of the determinants of inefficiency. The early literature adopts a two-step procedure to investigate this. The basic SFA is used to estimate inefficiency in the first step, and then regresses inefficiency score on a vector of exogenous variables in the second step. The two-step procedure, however, has long been recognized as biased because the model estimated in the first step is mis-specified (Battese and Coelli, 1995). As explained in Wang and Schmidt (2002), if $\mathbf{x}_i$ and $\mathbf{z}_i$ are correlated then the first step of the two-step procedure is biased. Even when $\mathbf{x}_i$ and $\mathbf{z}_i$ are uncorrelated, ignoring the dependence of the inefficiency on $\mathbf{z}_i$ will cause the first-step inefficiency score to be under-dispersed, so that the results of the second-step regression are likely to be biased downward. Wang (2002) provides Monte Carlo evidence of the bias. Given the undesirable statistical properties of the two-step procedure, the preferred approach is the single-step procedure. This procedure estimates the parameters of the relationship between inefficiency and $\mathbf{z}_i$ together with all the other parameters of the model in the maximum likelihood method.

The single-step procedure accounts for the influence of determinants of inefficiency by parameterizing the distribution function of u_i as a function of the determinants of inefficiency $\mathbf{z}_i$ that are likely to affect inefficiency. If u_i follows a half-normal distribution, then σ_u^2 is the (only) parameter to be para-meterized by $\mathbf{z}_i$. Further, the exponential function is used to ensure a positive estimate of the variance parameter. Therefore, the parameterization is as follows

$$\sigma_{u,i}^2 = \exp(\mathbf{z}_{u,i}{}^{\mathrm{T}}\mathbf{w}_u), \tag{III-8}$$

where $\mathbf{w}_u$ is the corresponding coefficient vector for the determinants of inefficiency. The expected value of u_i is now a function of σ_u^2 as

$$E[u_i] = \sqrt{\frac{2}{\pi}}\sigma_u = \sqrt{\frac{2}{\pi}}\exp\left(\frac{1}{2}\mathbf{z}_{u,i}{}^{T} w_u\right). \tag{III-9}$$

However, the estimated values of $\mathbf{w}_u$ might not be very informative because the relationship between $E[u_i]$ and $\mathbf{z}_u$ is nonlinear, and so the slope coefficients of $\mathbf{w}_u$ are not the marginal effects of $\mathbf{z}_u$. The marginal effect of the kth variable of $\mathbf{z}_{u,i}$ on $E[u_i]$ given the half-normal assumption of u_i is

$$\frac{\partial E[u_i]}{\partial z_{[k]}} = w_{[k]}\frac{\sigma_{u,i}}{2}\left[\frac{\phi(0)}{\Phi(0)}\right] = w_{[k]}\sigma_{u,i}\phi(0). \tag{III-10}$$

III.3 Data

To measure the efficiency of education sector in SEA region, the recent OECD PISA 2018 data is used. PISA is a triennial survey of 15-year-old students that assesses the extent to which they have acquired the key knowledge and skills essential for full participation in society. The assessment focuses on proficiency in reading, mathematics, and science. Some education systems in SEA have a long tradition of participation in the PISA assessments, whereas others only started participating in 2018. Indonesia (IDN) and Thailand (THA) have participated in all cycles since the first assessment in 2000;

Table III-2 Features of the PISA 2018 participation of South-East Asian countries

		BRN	IDN	MYS	PHL	SGP	THA	VIE
Assessment format	Computer	X	X	X	X	X	X	
	Paper							X
Global competence		X	X		X	X	X	
Financial literacy			X					
Optional question-naires	Educational career	X					X	
	ICT	X				X	X	
	Well-being							
	Parent							
	Teacher			X				
Language of assessment		English	Indonesia	English, Malay	English	English	Thai	Vietna-mese

Malaysia (MYS) and Singapore (SGP) joined PISA in 2009; Vietnam (VIE) took part for the first time in PISA 2012; and Brunei Darussalam (BRN) and the Philippines (PHL) did so in PISA 2018. Cambodia participated in PISA for Development, a project whose goal was to encourage and facilitate PISA participation by interested and motivated low-and middle-income countries.

In 2018 wave, students in all countries in this region, except in Vietnam, took the computer-based assessment, which allows education systems to take full advantage of the assessment. PISA also provides information that is potentially related to the assessment result, such as variables representing student background, school environment, or educational provision. This information comes from the responses given to different questionnaires completed by students, school principals, or parents. All those countries in 2018 wave distributed the mandatory student and school questionnaires. PISA 2018 wave also offered countries four optional questionnaires for students (i.e., the educational career questionnaire, the information and communication technology (ICT) familiarity questionnaire, the well-being questionnaire, and the financial literacy questionnaire); an optional questionnaire for parents; and an optional questionnaire for teachers (both for reading teachers and for teachers of all other subjects). In the region, Brunei Darussalam and Thailand distributed the educational career questionnaire; Brunei Darussalam, Singapore and Thailand distributed the ICT questionnaire; and Malaysia distributed the teacher questionnaire. In addition, PISA also offers the possibility of assessing financial literacy and each cycle explores a new “innovative domain”, such as problem solving (in 2012 wave), collaborative problem solving (in 2015 wave), and global competence (in 2018 wave). In PISA 2018, Brunei Darussalam, Indonesia, the Philippines, Singapore, and Thailand took part in the global competence assessment, and only Indonesia evaluated financial literacy, see Table III-2.

Apart from the assessment results, due to the difference in the provision of optional questionnaires, only the mandatory questionnaires (i.e., student and school questionnaire) are used in this study. Among seven countries participated in PISA 2018 wave, this study excludes Vietnam due to the following reasons. First, Vietnam participated in PISA 2018 wave using paper-based instruments, whereas others using computer-based instruments. Worldwide, only students in Argentina, Jordan, Lebanon, Moldova, Republic of Northern Macedonia, Romania, Saudi Arabia, and Ukraine still took the paper test. By the time the ranking was published, the international comparability of Vietnam's performance could not be fully ensured despite getting high score. According to the Ministry of Education and Training of Vietnam, Vietnam scores 505 points in the reading test (13rd in the world), 496 in mathematics (24th), and 543 in science (4th) (Việt Nam News, 2019). One strand of literature suspected Vietnam's extraordinary performance in the PISA assessment despite being the one of the poorest of all participating countries and a centralized education system (Asadullah et al., 2020). Glewwe (2016) scrutinized the representativeness of the PISA sample for Vietnam (e.g., better socio-economic status of participating children) as a source of its surprising performance. He stressed the sample of children in the PISA assessment may not be representative of all children in Vietnam. The conventional drivers of educational outcomes are also unable to explain Vietnam's superior performance in PISA (Asadullah et al., 2020). Some studies concluded that further research is needed to study more about the so called "Vietnam's paradoxical performance" phenomenon. For this reason, OECD did not report comparisons of Vietnam's performance in PISA with other countries.[7]

In PISA 2018 wave, more than six hundred thousand students were examined, representing about thirty-two million students of seventy-nine participating countries. In those six SEA countries, there are 47,579 students from 1,286 schools who were sampled. Indonesia has the highest number of sampled students and schools as Indonesia is the most populous country in this region.

Variables used in this study are extracted from student and school questionnaire. Since this study is conducted at the school level, so that variables which are at student level have to be weighted using W_FSTUWT, the final student weight provide by PISA, to get school level variables. Since PISA assessment focuses on three kinds of proficiency, i.e., in mathematics, science, and reading (language), there are three models to be considered in this study, namely, Model 1 in which the output (education outcome) reflects the student's proficiency in mathematics, Model 2 for science, and Model 3 for language. The education outcomes are proxied by the weighted plausible values (PVs) provided by PISA—in this study, later it is called the PISA scores. Rather than a single measure of education outcome, PISA provides five PVs for each domain.[8] In this study, the first PV for each domain is used because these values provide both unbiased point and sampling variance estimates; while the other PVs will be used in the robustness analysis (see Chapter III.4.4).[9]

---

[7] See: Annexes A4 and A6 in OECD (2019a).

[8] See Wu (2005) for a detailed discussion about the role of PVs in large-scale surveys.

[9] The use of one PV or even five PVs does not really make a substantial difference in large samples (see OECD, 2009, p. 44 for details).

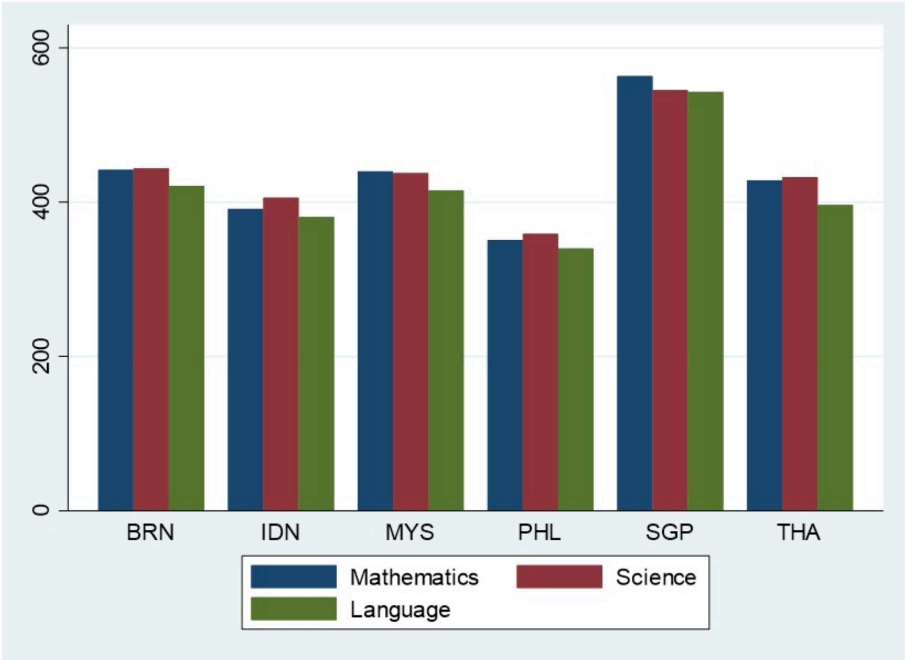


Figure III-1 School's average performance in mathematics, reading, and science in PISA 2018 wave

School's average performance in mathematics, science, and reading (language) for each country is displayed in Figure III-1. On average across six SEA countries, schools scored 425.4 points in mathematics, 429.5 points in science, and 406.2 points in reading. Countries with a similar performance are mostly located in Latin America and Southeast Europe, such as Bulgaria, Colombia, Romania, Serbia, and Uruguay (OECD, 2019b). Singapore has the highest points in all domains, as 563.5, 545.5, and 543.2 points in mathematics, science, and language, respectively; whereas the Philippines has the lowest points in all domains (350.9, 358.9, and 340.1 in mathematics, science, and language, respectively). Other than the Philippines, Indonesia's points are below the SEA average points in all domains, while Thailand's point is below in language domain.

The selection of inputs used in the analysis has been guided by the existent literature in the field of education economics. Some references to be the justification of input selection are reported in Table III-3 which also provides the list of the inputs adopted and their description. Inputs are classified in three categories which reflect the main group of variables: (i) student's characteristics, including index of economic, social, and cultural status (ESCS); (ii) school's characteristics, including leaning time per week in mathematics (MMINS), science (SMINS), or language (LMINS), school size (SCHSIZE), school type (SCHTYE), school location (SCHLOC), proportion of female students in a school (PROP_girl), proportion of native students in a school (PROP_nat), student-teacher ratio (STRATIO), proportion of fully certified teacher (PROATCE), index of educational material shortage (EDUSHORT), and index of educational staff shortage (STAFFSHORT); and (iii) ICT-related variables, including the ratio of computers to the total number of students for educational purposes (COMPRATIO)

Table III-3 Description of inputs

Inputs	Descriptions	References
ESCS	Index of economic, social, and cultural status	Crespo-Cebada et al. (2014); Ferrera et al. (2011); Perelman and Santín (2011a); Salas-Velasco (2020)
M(S/L)MINS	Mathematics (science/language) learning time per week (in minutes)	Dolton et al. (2003)
SCHSIZE	School size or number of enrolled students	Barnett et al. (2002); Bradley and Taylor (1998); Hanushek and Luque (2003); Mora et al. (2010)
SCHTYPE	Type of school: public, private government-independent (if the school gets less than 50% of total funding from government, includes departments, local, regional, state, and national, and private government-dependent (if the school gets more than 50% of their total funding from the government)	Crespo-Cebada et al. (2014); Garcia-Diaz et al. (2016); Kiryavainen (2012); Perelman and Santín (2011a)
SCHLOC	School location: located in rural area (fewer than 3,000 people), in a small town (3,000 to about 15,000 people), in a town (15,000 to about 100,000 people), in a city (100,000 to about a million people), and close to the center of a city with over a million people or elsewhere in a city with over a million people	Kiryavainen (2012); Perelman and Santín (2011a)
PROP_girl	Proportion of female students in a school	Crespo-Cebada et al. (2014); Kiryavainen (2012); Mongan et al. (2011); Perelman and Santín (2011a); Zoghbi et al. (2013)
PROP_nat	Proportion of native students in a school (i.e., students who had at least one parent born in the country	Crespo-Cebada et al. (2014); Perelman and Santín (2011a); Zoghbi et al. (2013)
STRATIO	Student-teacher ratio	Agasisti (2014); Agasisti et al. (2019); Agasisti and Zoido (2019)
PROATCE	Proportion of fully certified teacher	André et al. (2020), Grosskopf et al. (2014)
EDUSHORT (IRT)	Index of educational material shortage	Crespo-Cebada et al. (2014); Ferrera et al. (2011); Perelman and Santín (2011a); Salas-Velasco (2020)
STAFFSHORT (IRT)	Index of educational staff shortage	Courtney et al. (2022); Lima (2017); Shahini (2021)
COMPRATIO*	Ratio of computers to the total number of students for educational purposes	Perelman and Santín (2011b); Zoghbi et al. (2013)
WEBCOMP*	Ratio of computers at school to the number of these computers that were connected to the internet	Salas-Velasco (2020)

*also serve as the determinants of inefficiency

and the ratio of computers that were connected to the internet (WEBCOMP). Two variables are dummy variables (SCHTYPE and SCHLOC); two variables are derived based on item response theory (IRT) scaling, i.e., EDUSHORT and STAFFSHORT.[10] Two variables (COMPRATIO and WEBCOMP) are included as determinants of inefficiency.

ESCS is probably, just after student achievement scores, the most used variable in reports and in secondary analysis of data from PISA (Avvisati, 2020). It helps address relevant questions about educational opportunity and inequalities in learning outcomes. In PISA, ESCS is defined as a measure of students' access to family resources (financial capital, social capital, cultural capital and human capital) which determine the social position of the student's family/household (Avvisati, 2020). It is a composite score based on three indicators: highest parental occupation, parental education, and home possessions. The rationale for using these three components was that the socio-economic status has usually been seen as based on education, occupational status, and income (Sirin, 2005; Willms and Tramonte, 2019). As no direct income measure has been available from the PISA data, the existence of household items has been used as a proxy for family wealth. All three components were standardized for OECD countries and partner countries/economies with an OECD mean of zero and a standard deviation of one. ESCS has been widely known to explain the educational outcomes, see for example Crespo-Cebada et al. (2014), Ferrera et al. (2011), Perelman and Santín (2011a), Salas-Velasco (2020).

Dolton et al. (2003) who investigated the effect of student's learning time on examination performance at a university in Spain found that within the formal system of teaching in Spain, both formal study and self-study were significant determinants of the exam scores. School size indicates the total number of students in the school. The influence of this variable in the educational process has also been tested in previous studies. Some papers suggest that schools with more students have better results (Barnett et al. 2002; Bradley and Taylor 1998), whereas others find that size has no influence on student results (Hanushek and Luque, 2003), and yet others conclude that smaller school sizes reduce the dropout rate and the proportion of early school-leaving (Mora et al., 2010).

There are three types of school considered in PISA, i.e., public, private government-independent, and private government-dependent.[11] Regarding this variable, literature shows that it may importantly affect the performance of the students (e.g., Crespo-Cebada et al., 2014; Garcia-Diaz et al., 2016; Kirjavainen, 2012). In PISA, location of the school is divided into five categories, i.e., located in rural area (fewer than 3,000 people), in a small town (3,000 to about 15,000 people), in a town (15,000 to about 100,000 people), in a city (100,000 to about a million people), and close to the center of a city with over a million people or elsewhere in a city with over a million people. School location has been

---

[10] For details on how each IRT-derived variable was constructed, see the *PISA 2018 Technical Report* (available online at: https://www.oecd.org/pisa/data/pisa2018technicalreport/).

[11] Schools are categorized as private government-independent if they are not public school, but they get less than 50% of total funding from government (includes departments, local, regional, state, and national); whereas private government-dependent if they are not public school and get more than 50% of their total funding from the government.

proposed as determinants of student achievement by several studies (e.g., Kirjavainen, 2012; Perelman and Santín, 2011a).

PROP_girl is an index of the proportion of female students in the school that is based on the enrolment data provided by the school principals. It is computed by dividing the number of female students by the total number of students at the school. Gender might influence the level of student's achievement as has been shown in Rudd (1984), Rodger and Ghosh (2001), and Smith and Naylor (2001). The next independent variable is the proportion of native students in a school. Students' immigration status was included since it could give effect to the educational outcome (e.g., Cortes, 2006; Schnepf, 2008). The student-teacher ratio might affect the educational output as shown by Agasisti (2014), Agasisti et al. (2019), Agasisti and Zoido (2019). The proportion of fully certified teachers (PROATCE) is computed using school principals' responses on the number of teachers and the number of teachers who are fully certified. It is commonly used predictor of student achievement. It is often considered the most reliable among various measures of teacher quality (Darling-Hammond, 2000). André et al. (2020) showed that this variable is positively affect the average of mathematics test result of schools in Sweden.

The next two variables which are based on IRT scaling, i.e., EDUSHORT and STAFFSHORT, are about school resources, measuring school principals' perceptions of potential factors hindering instruction at school. EDUSHORT is derived from four items: a lack of educational material; inadequate or poor-quality educational material; a lack of physical infrastructure; inadequate or poor-quality physical infrastructure; whereas STAFFSHORT is derived from four items: a lack of teaching staff; inadequate or poorly qualified teaching staff; a lack of assisting staff; inadequate or poorly qualified assisting staff. These variables were scaled to have a mean of 0 and a standard deviation of 1 across OECD countries (with equally weighted countries). A score of 0 is expected for an average student in an OECD country. Negative values on the index do not imply that students responded negatively to the underlying question; rather, students with negative scores are those who responded less positively than the average student across OECD countries, but not necessarily negative with regards to the underlying question. Likewise, students with positive scores are those who responded more positively than the average student in OECD countries.

The last two variables are related to ICT, i.e., COMPRATIO and WEBCOMP. COMPRATIO was calculated as the number of computers for educational purposes divided by the total number of students in the school. This variable has been used in the study of Zoghbi et al. (2013) to estimate the efficiency of higher education institutions in Brazil and Perelman and Santín (2011b) to investigate the performance of Spanish schools proxied by the PISA scores of mathematics and language. WEBCOMP was calculated as the number of computers connected to the internet divided by the total number of

Table III-4 Descriptive statistics

Variable	Total	BRN	IDN	MYS	PHL	SGP	THA
Number of students	47,579	6,828	12,098	6,111	7,233	6,676	8,633
Number of schools	1,286	55	397	191	187	166	290
Mean of ESCS	-1.054	-0.164	-1.472	-0.763	-1.423	0.133	-1.283
Mean of MMINS	265.588	227.153	255.908	241.651	316.567	310.268	243.548
Mean of SMINS	268.881	266.655	229.840	259.092	314.986	320.285	270.208
Mean of LMINS	246.406	209.662	245.801	260.905	317.275	265.430	188.110
Mean of SCHSIZE	1,218.428	867.273	571.055	1,097.058	2,300.610	1,165.133	1,392.249
Mean of PROP_girl	0.498	0.482	0.497	0.493	0.508	0.495	0.502
Mean of PROP_nat	0.924	0.890	0.967	0.946	0.928	0.743	0.961
Mean of STRATIO	16.444	9.305	16.936	11.620	25.600	11.257	17.504
Mean of PROATCE	0.818	0.939	0.591	0.932	0.906	0.857	0.882
Mean of EDUSHORT	0.246	0.022	0.712	-0.038	0.706	-1.067	0.381
Mean of STAFFSHORT	0.012	-0.167	0.364	0.199	-0.248	-0.698	0.085
Mean of COMPRATIO	0.522	0.898	0.362	0.372	0.303	1.083	0.533
Mean of WEBCOMP	0.835	0.970	0.806	0.859	0.526	0.991	0.935

computers for educational purposes. This variable has been used in the study of Salas-Velasco (2020) to evaluate the performance of Spanish secondary schools.

The descriptive statistics of numerical inputs is shown in Table III-4. Several important findings are discussed as following. Since ESCS is scaled to have mean of zero and standard deviation of one across senate-weighted OECD countries, the fact that the mean of ESCS of six SEA countries is -1.054 indicates that the economic, social, and cultural condition of the sampled students is below the average of students in the OECD countries. It is no surprise that Singapore has the highest average value of ESCS as having the highest GDP per capita in the SEA region. The number of girls and boys in a school is quite balance (the average value of PROP_girl is 49.84%). However, there are 17 schools which haves no girl (8 in Singapore and 3 in Brunei Darussalam, Malaysia, and Thailand). In terms of school size, there is a school in Thailand which has the lowest value, i.e., 3, meaning that the school only has three students; oppositely, there is a school in the Philippines that has 11,990 students, the highest among all. In terms of proportion of fully certified teachers, the average value is 81.830%. There are 521 sampled schools in which all the teachers are fully certified while only 56 sampled schools in which no fully certified teacher. Regarding the ICT-related variables, there are twenty sampled schools which do not have computer (7 in Malaysia, 6 in the Philippines and Indonesia, and 1 in Singapore) and sixty-four sampled

Table III-5 Tabulation of non-numerical inputs

Non-numerical inputs	Total (%)	BRN (%)	IDN (%)	MYS (%)	PHL (%)	SGP (%)	THA (%)
SCHTYPE:	1,248 (100%)	55 (100%)	359 (100%)	191 (100%)	187 (100%)	166 (100%)	290 (100%)
Private government independent	119 (9.54%)	14 (25.45%)	49 (13.65%)	11 (5.76%)	17 (9.09%)	13 (7.83%)	15 (5.17%)
Private government-dependent	93 (7.45%	0 (0%)	59 (16.43%)	1 (0.52%)	17 (9.09%)	0 (0%)	16 (5.52%)
Public	1,036 (83.01%)	41 (74.55%)	251 (69.92%)	179 (93.72%)	153 (81.82%)	153 (92.17%)	259 (89.31%)
SCHLOC:	1,233 (100%)	55 (100%)	351 (100%)	190 (100%)	187 (100%)	160 (100%)	290 (100%)
Rural	196 (15.90%)	6 (10.91%)	68 (19.37%)	38 (20.00%)	13 (6.95%)	0 (0%)	71 (24.48%)
Small town	259 (21.01%)	28 (50.91%)	122 (34.76%)	34 (17.89%)	17 (9.09%)	0 (0%)	58 (20.00%)
Town	275 (22.30%)	14 (25.45%)	66 (18.80%)	41 (21.58%)	72 (38.50%)	0 (0%)	82 (28.28%)
City	256 (20.76%)	7 (12.73%)	68 (19.37%)	63 (33.16%)	58 (31.02%)	0 (0%)	60 (20.69%)
Large city	247 (20.03%)	0 (0%)	27 (7.69%)	14 (7.37%)	27 (14.44%)	160 (100%)	19 (6.55%)

schools whose computers do not have access to the internet. The rest of the numerical inputs are constructed using the IRT scaling methodology. When the mean value is below zero, it means that it is below the average value of OECD countries, vice versa.

The tabulation of non-numerical inputs is displayed in Table III-5. The proportion of public school participated in the survey is 83.01% (or 1,036 schools), the most among others. In Brunei Darussalam and Singapore, there is no private government-dependent schools which participated in PISA 2018 wave. About 15.90% of participated schools are located in the rural area; 259 schools are located in the small town; and about 22.30%, 20.76%, and 20.03% of participated schools are located in town, city, and large city, respectively. In Singapore, all schools participated in PISA 2018 wave only located in large city, while in Brunei Darussalam, there is no schools located in the large city.

III.4 Results

III.4.1 Parameters estimation

This section describes parameters estimation using SFA allowing for heteroscedasticity. The linear specification is used as the production function. Notice that there are three models with different PISA score domain: mathematics (Model 1), science (Model 2), and reading (Model 3). Result is displayed in Table III-6. The sign of the coefficient can be interpreted as follows. A positive coefficient indicates that as the value of the input increases, the expected value of the output also tends to increase,

Table III-6 Parameters estimation

Inputs	Model 1		Model 2		Model 3	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
Constant	460.787**	15.161	449.435**	13.930	472.657**	14.446
ESCS	54.862**	2.143	49.270**	2.025	52.833**	2.024
M(S/L)MINS	0.010	0.019	0.042**	0.014	-0.095**	0.017
SCHSIZE	0.000	0.001	0.000	0.001	-0.002	0.001
PROP_girl	37.201**	9.337	41.985**	8.597	67.058**	8.677
PROP_nat	-35.416**	12.207	-25.009**	11.300	-54.939**	11.523
STRATIO	-0.803**	0.214	-0.725**	0.196	-0.463**	0.198
PROATCE	-6.863	5.178	-11.697**	4.769	-8.310**	4.841
EDUSHORT	-9.068**	1.587	-7.816**	1.458	-8.125**	1.475
STAFFSHORT	4.103**	1.516	4.725**	1.396	4.137**	1.411
COMPRATIO	2.508	2.803	2.854	2.441	3.067	2.617
WEBCOMP	42.195**	5.855	35.555**	5.201	27.191**	6.028
SCHTYPE:						
Private gov.-dep.	4.265	6.733	1.573	6.199	6.816	6.273
Public	29.813**	4.969	24.256**	4.608	25.856**	4.639
SCHLOC:						
Small town	-3.567	4.514	-0.526	4.153	1.465	4.201
Town	-9.741**	4.695	-9.952**	4.318	-4.362	4.377
City	-11.840**	5.088	-11.667**	4.676	-4.872	4.749
Large city	23.686**	5.658	16.942**	5.170	35.583**	5.271
σ_u^2:						
Constant	4.806**	2.183	4.949**	2.445	3.423	4.794
COMPRATIO	-13.352**	6.429	-17.718**	7.899	-16.437	12.260
WEBCOMP	2.967	2.166	2.687	2.360	3.875	4.494
σ_v^2	7.481**	0.045	7.323**	0.044	7.350**	0.044

*significant at the level of 10%
**significant at the level of 5%

vice versa. The value of the coefficient signifies how much the expected value of the output alters given a one-unit shift in the particular input while holding other inputs constant. This property is crucial because it allows to assess the effect of each variable in isolation from the others. Not only the sign, but the significancy of the coefficients also has to be considered since only the significant inputs have influence on the output (in Table III-6, it is marked by the asterisk * and double asterisk ** when the particular input is statistically significant at the level of 10% and 5%, respectively).

The anticipated positive value of the index of economic, social, and cultural status (ESCS) in all models indicates as the higher the economic, social, and cultural status of the student (aggregated to school level), the higher the PISA score will be obtained. This finding confirms the result of other studies, e.g., Perelman and Santín (2011a), Salas-Velasco (2020), and Ulkhaq (2021). The positive sign is also found in the share of female students, meaning that as the proportion of female students in a school increase, the PISA score for all proficiencies tends to increase as well. This finding somewhat partially follows Mancebón et al. (2012) who found that girls perform better at language, but worse for mathematics and science, at which boys achieve better results in PISA 2006.

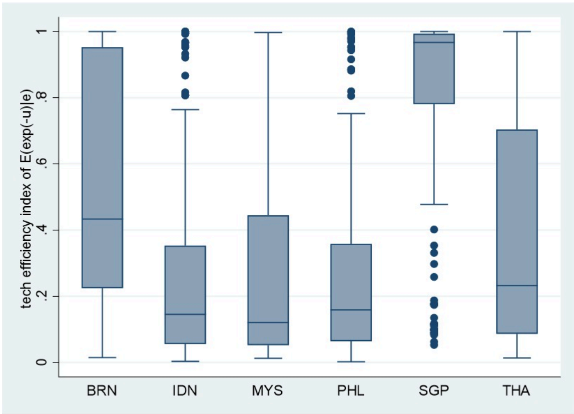
On the other hand, the negative sign is found in student-teacher ratio, meaning that if the ratio increases, the school's performance tends to decrease. This finding is consistent with those obtained by e.g., Franta and Konecny (2009) and Mizala et al. (2002) with regard to the negative relationship between this ratio and educational outcomes. This finding is as expected and confirms the conventional wisdom that smaller classes are more conductive to better learning (Chakraborty et al., 2001). The negative sign is also found in the index of school educational resources. Proportion of native students is also found to be significant with negative value in all models. School size is found to be not significant in all models. It is suspected that the relationship of school size and output is *monotonic* since there may come a point where schools become "too big" (Bradley and Taylor, 1998). Schools may also become more difficult to manage when they become very big, giving rise to disciplinary problems (Haller, 1992).

Regarding the ICT-related variables as inputs, only the ratio of computers connected to the internet is significant in all models. This finding contrasts with Salas-Velasco (2020) who found that this variable is not significant to influence student's performance. Subsequently, the ratio of computers to the number of students in a school is not significant. This result confirms the finding of Perelman and Santín (2011b) who mentioned that this ratio had no influence on education outcomes. In relation to the school policy, it indicates that to get higher PISA score, the school has to increase the number of computers connected to the internet and not only number of computers per se. This is in line with the results of Garcia-Diaz et al. (2016) who found that internet access at school positively and significantly influences education outcomes.

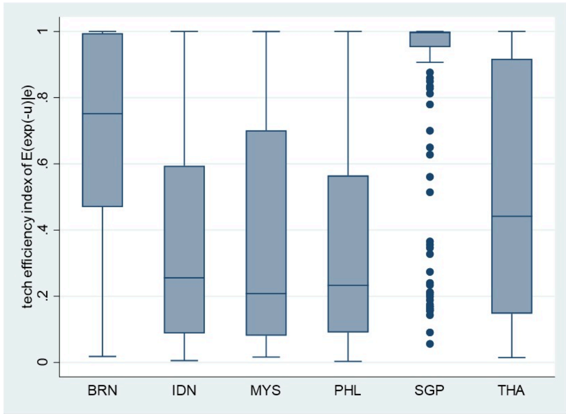
III.4.2 Efficiency estimation

Individual efficiency (for each school) is obtained by using the BC estimator. The distributions of efficiency scores, by country, are shown in Figure III-2. Singapore has the highest average of efficiency scores in all models (i.e., 80.05% in mathematics, 86.05% in science, and 85.61% in language test score), indicating that on average, about 15%~20% of the potential output (proxied by the PISA score) is lost due to inefficiency. On the other hand, Indonesia has the lowest average of efficiency scores in all models (i.e., 24.88% in mathematics, 36.04% in science, and 37.28% in language test score). The average efficiency scores across all countries are found at 38.47% (in mathematic), 48.83% (in science), and 49.82% (in language), implying that, when considering the best-performing schools in the sample, on average, the other schools can improve their PISA score by about 50% to 62% with the currently available resources. Such figure suggests the room for considerable efficiency improvements.

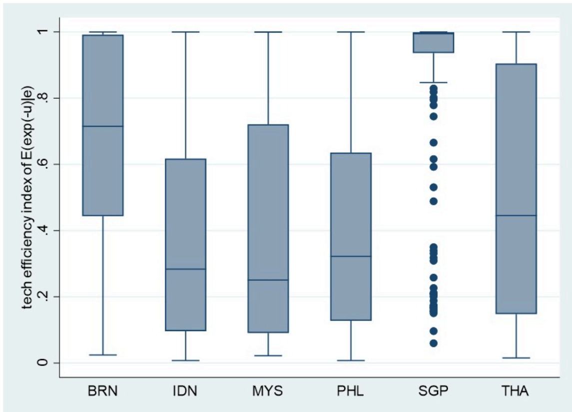
Standard deviation, as well as the range of minimum and maximum values, indicates that relevant variation both between and within countries can be detected, see again Figure III-2. This visual representation is also useful for showing countries with particular characteristics to be investigated or discussed, with the aim of understanding the particular status of high/low efficiency score and their



(a) Model 1



(b) Model 2



(c) Model 3

Figure III-2 The distributions of efficiency scores, by country

variations. For instance, Singapore has the lowest standard deviation (and also the highest PISA scores) in Model 2 and Model 3, implying that the education system in this country could give a more *uniform* education outcomes measured by the PISA scores of science and reading. Detailed information of the average scores and standard deviations can be seen in Table III-7, in columns: Average score and Std. Deviation.

In the subsequent analysis, following Agasisti and Zoido (2019), schools according to the ESCS score are categorized. The "advantaged" schools are when they have high proportion of relatively better-off students. Formally, they are defined so when their students' average ESCS is above the 75th percentile of the within-country ESCS distribution. The "disadvantaged" schools are those in which the average students' ESCS is below the 25th percentile of the within-country ESCS distribution. While residually, the "average" schools are those whose students' average ESCS is between 25th and 75th percentile of the within country ESCS distribution. However, the finding of this study—see Table III-7—contrasts with the finding of Agasisti and Zoido (2019) as they reported that "advantaged" schools are on average more efficient than their "disadvantaged" counterparts. In this study, this phenomenon is only observed in several countries. The "advantaged" schools in Brunei Darussalam and Malaysia have

Table III-7 Efficiency scores, by country

Country	Average score	Std. Deviation	Disadvantaged schools (a)	Average schools (b)	Advantaged schools (c)	Difference (d)
Panel A: Model 1						
BRN	53.49%	0.361	37.40%	53.49%	75.89%	102.93%
IDN	24.88%	0.268	32.92%	24.88%	24.29%	-26.22%
MYS	28.34%	0.304	30.09%	28.34%	34.37%	14.22%
PHL	27.66%	0.294	36.18%	27.66%	35.85%	-0.92%
SGP	80.05%	0.320	87.22%	80.05%	76.41%	-12.39%
THA	38.72%	0.351	49.09%	38.72%	38.35%	-21.89%
Panel B: Model 2						
BRN	68.23%	0.323	58.08%	68.23%	81.82%	40.88%
IDN	36.04%	0.314	44.40%	36.04%	36.50%	-17.78%
MYS	39.31%	0.351	39.54%	39.31%	47.49%	20.11%
PHL	36.14%	0.328	45.58%	36.14%	47.65%	4.55%
SGP	86.05%	0.281	91.81%	86.05%	83.44%	-9.11%
THA	50.84%	0.362	61.79%	50.84%	51.67%	-16.37%
Panel C: Model 3						
BRN	67.10%	0.322	56.47%	67.10%	81.42%	44.18%
IDN	37.28%	0.312	46.47%	37.28%	37.20%	-19.95%
MYS	40.02%	0.344	40.33%	40.02%	47.83%	18.60%
PHL	41.25%	0.325	51.43%	41.25%	49.55%	-3.65%
SGP	85.61%	0.281	91.43%	85.61%	82.98%	-9.24%
THA	50.22%	0.357	61.34%	50.22%	50.34%	-17.92%

Notes: (a) “disadvantaged” schools are those in which the average students’ ESCS is below the 25th percentile of the within-country ESCS distribution; (b) “advantaged” schools are defined so when their students’ average ESCS is above the 75th percentile of the within-country ESCS distribution; (c) “average” schools are those whose students’ average ESCS is between 25th and 75th percentiles of the within-country ESCS distribution; (d) the difference is computed as ((c–a)/a) × 100%.

higher efficiency scores compared to the “disadvantaged” schools. However, Indonesia, Singapore, and Thailand experience the opposite condition, meaning that the "disadvantaged” schools turn out having higher efficiency scores than their counterparts. Inconclusive result is found in the case of the Philippines as the “advantaged” schools perform better in terms of efficiency in science test score, but worse in mathematics and language test scores.

For further analysis, the characteristics of the most and the least efficient schools are examined. The most and the least efficient schools are defined as those whose the efficiency score is in the 90th and 10th percentile of scores’ distribution, respectively. The objective of this analysis is to characterize the frontier or the efficient schools, especially by checking whether schools in specific countries are more likely than others to influence the efficiency benchmark or lagging well behind the efficiency standards. The result of this analysis is shown in Table III-8 for the least efficient schools and Table III-9 for the most efficient schools.

Almost 50% of schools in Singapore belong to the frontier group as the most efficient school in all models. To corroborate the dominance of this country in the SEA region, only one school belongs to the least efficient school (in terms of language test score). Next, schools in Thailand account for about

Table III-8 The characteristics of the least efficient schools

Panel A: The least efficient schools in Model 1												
Country	**n**	**% (country)**	**% (frontier)**	**ESCS**	**MMINS**	**SCHSIZE**	**PROP_girl**	**PROP_nat**	**STRATIO**	**PROATCE**	**EDUSHORT**	**STAFFSHORT**
BRN	1	1.82%	0.93%	-0.138	235.225	1557.000	1.000	0.966	11.449	0.779	-1.421	-0.587
IDN	41	10.33%	37.96%	-1.450	248.275	619.927	0.510	0.950	17.646	0.552	1.002	0.486
MYS	24	12.57%	22.22%	-0.718	231.871	1522.167	0.507	0.952	13.365	0.875	0.113	0.174
PHL	29	15.51%	26.85%	-1.419	292.801	4335.862	0.509	0.945	26.891	0.984	1.259	0.019
SGP	0											
THA	13	4.48%	12.04%	-1.367	223.352	1706.077	0.454	0.955	19.985	0.982	0.792	-0.236
Total	108	44.70%	100.00%	-1.257	253.465	1957.639	0.507	0.950	19.401	0.794	0.826	0.194
Panel B: The least efficient schools in Model 2												
Country	**n**	**% (country)**	**% (frontier)**	**ESCS**	**SMINS**	**SCHSIZE**	**PROP_girl**	**PROP_nat**	**STRATIO**	**PROATCE**	**EDUSHORT**	**STAFFSHORT**
BRN	1	1.82%	0.93%	-0.138	255.180	1557.000	1.000	0.966	11.449	0.779	-1.421	-0.587
IDN	38	9.57%	35.19%	-1.487	217.077	607.790	0.502	0.942	16.708	0.551	1.095	0.386
MYS	25	13.09%	23.15%	-0.669	249.107	1594.360	0.509	0.955	13.411	0.880	0.087	0.189
PHL	32	17.11%	29.63%	-1.449	294.299	4304.281	0.509	0.935	26.999	0.984	1.326	0.046
SGP	0											
THA	12	4.14%	11.11%	-1.189	246.457	1701.250	0.468	0.972	19.725	0.970	0.908	0.290
Total	108	45.73%	100.00%	-1.241	250.989	2061.704	0.507	0.947	19.280	0.804	0.886	0.220
Panel C: The least efficient schools in Model 3												
Country	**n**	**% (country)**	**% (frontier)**	**ESCS**	**LMINS**	**SCHSIZE**	**PROP_girl**	**PROP_nat**	**STRATIO**	**PROATCE**	**EDUSHORT**	**STAFFSHORT**
BRN	2	3.64%	1.85%	0.155	230.728	1250.000	0.791	0.876	9.191	0.890	0.133	-0.600
IDN	37	9.32%	34.26%	-1.471	235.428	614.973	0.502	0.941	17.062	0.538	1.027	0.377
MYS	24	12.57%	22.22%	-0.727	253.878	1631.500	0.507	0.953	13.629	0.875	0.120	0.232
PHL	27	14.44%	25.00%	-1.437	277.581	4264.444	0.505	0.944	26.672	0.984	1.212	0.006
SGP	1	0.60%	0.93%	0.801	271.057	903.000	0.433	0.436	11.652	0.781	-1.421	-1.455
THA	17	5.86%	15.74%	-1.266	175.028	1481.706	0.465	0.963	19.671	0.979	1.188	0.532
Total	108	46.42%	100.00%	-1.214	240.802	1904.093	0.503	0.942	18.916	0.802	0.858	0.241

Table III-8. The characteristics of the least efficient schools (continued)

Panel A: The least efficient schools in Model 1			
Country	**COMPRATIO**	**WEBCOMP**	**Math. score**
BRN	0.095	1.000	451.629
IDN	0.094	0.974	356.592
MYS	0.099	0.997	431.319
PHL	0.067	0.990	349.949
SGP			
THA	0.100	0.996	374.822
Total	0.088	0.986	374.488
Panel B: The least efficient schools in Model 2			
Country	**COMPRATIO**	**WEBCOMP**	**Science score**
BRN	0.095	1.000	438.970
IDN	0.087	0.979	374.523
MYS	0.094	1.000	436.135
PHL	0.065	0.931	353.344
SGP			
THA	0.087	0.996	403.702
Total	0.082	0.972	386.349
Panel C: The least efficient schools in Model 3			
Country	**COMPRATIO**	**WEBCOMP**	**Language score**
BRN	0.077	1.000	481.497
IDN	0.088	0.978	346.281
MYS	0.093	1.000	408.223
PHL	0.062	0.989	334.834
SGP	0.159	1.000	495.221
THA	0.099	0.997	369.995
Total	0.085	0.989	364.800

Notes: **n** represents the most (or the least) efficient schools for each country; **% (country)** represents the proportion of the most (or the least) efficient schools for each country; **% (frontier)** represents the proportion of the country's schools in the group of the most (or the least) efficient ones.

22%~23% of the most efficient group while having about 11%~15% of the least efficient schools. Schools in Brunei Darussalam account for an additional 10.28% as the most efficient ones and just having just about 1%~2% of the least efficient schools. The highest proportion of the least efficient schools is found in Indonesia as having about 34%~37% of the least efficient ones.

Other interesting insights from Table III-8 and Table III-9 deal with the different profiles of the most and the least efficient schools. For instance, while schools in Brunei Darussalam, Malaysia, and Singapore with the efficiency score in the 90th percentile have higher than average scores in the corresponding test scores (i.e., 425.4, 429.5, and 406.2 in mathematics, science, and language, respectively), the most efficient schools in Indonesia, Malaysia, the Philippines, and Thailand also belong to this group despite of their relatively low test scores.

Next, it is attempted to correlate the efficiency score with the student's performance. As shown in Figure III-3, it is possible to individuate how efficiency is able to capture a different perspective than pure performance. On the horizontal axis, the efficiency score is reported, while the vertical axis has the PISA 2018 scores on the basis of country-average data. The PISA score is then regressed on the efficiency

Table III-9 The characteristics of the most efficient schools

Country	n	% (country)	% (frontier)	ESCS	MMINS	SCHSIZE	PROP_girl	PROP_nat	STRATIO	PROATCE	EDUSHORT	STAFFSHORT
Panel A: The most efficient schools in Model 1												
BRN	11	20.00%	10.28%	0.384	248.518	642.818	0.491	0.876	9.056	0.980	-0.699	-0.667
IDN	7	1.76%	6.54%	-2.116	193.083	99.714	0.495	0.971	6.088	0.199	0.821	1.296
MYS	5	2.62%	4.67%	-0.689	227.097	714.000	0.587	0.977	8.751	1.000	-0.343	0.198
PHL	6	3.21%	5.61%	-0.926	351.096	648.667	0.511	0.943	12.921	0.869	0.254	-0.538
SGP	53	31.93%	49.53%	0.104	316.425	1194.377	0.513	0.738	11.254	0.815	-1.215	-0.829
THA	25	8.62%	23.36%	-1.538	253.675	466.120	0.497	0.946	14.289	0.855	0.004	-0.048
Total	107	68.14%	100.00%	-0.491	284.484	842.860	0.509	0.838	11.376	0.813	-0.621	-0.427
Panel B: The most efficient schools in Model 2												
Country	**n**	**% (country)**	**% (frontier)**	**ESCS**	**SMINS**	**SCHSIZE**	**PROP_girl**	**PROP_nat**	**STRATIO**	**PROATCE**	**EDUSHORT**	**STAFFSHORT**
BRN	11	20.00%	10.28%	0.384	289.177	642.818	0.491	0.876	9.056	0.980	-0.699	-0.667
IDN	7	1.76%	6.54%	-2.116	140.402	99.714	0.495	0.971	6.088	0.199	0.821	1.296
MYS	6	3.14%	5.61%	-0.634	301.378	711.500	0.580	0.981	9.005	1.000	-0.269	0.194
PHL	6	3.21%	5.61%	-0.926	351.250	648.667	0.511	0.943	12.921	0.869	0.254	-0.538
SGP	53	31.93%	49.53%	0.096	324.362	1190.660	0.513	0.736	11.251	0.816	-1.201	-0.794
THA	24	8.28%	22.43%	-1.545	277.462	474.333	0.510	0.943	14.528	0.849	-0.017	0.010
Total	107	68.32%	100.00%	-0.485	298.409	845.037	0.512	0.838	11.390	0.813	-0.618	-0.394
Panel C: The most efficient schools in Model 3												
Country	**n**	**% (country)**	**% (frontier)**	**ESCS**	**LMINS**	**SCHSIZE**	**PROP_girl**	**PROP_nat**	**STRATIO**	**PROATCE**	**EDUSHORT**	**STAFFSHORT**
BRN	11	20.00%	10.28%	0.384	228.546	642.818	0.491	0.876	9.056	0.980	-0.699	-0.667
IDN	7	1.76%	6.54%	-2.116	199.040	99.714	0.495	0.971	6.088	0.199	0.821	1.296
MYS	6	3.14%	5.61%	-0.850	253.026	830.333	0.571	0.976	9.706	1.000	-0.077	0.494
PHL	6	3.21%	5.61%	-0.926	340.463	648.667	0.511	0.943	12.921	0.869	0.254	-0.538
SGP	52	31.33%	48.60%	0.115	272.596	1209.596	0.514	0.735	11.332	0.811	-1.211	-0.817
THA	25	8.62%	23.36%	-1.538	216.342	466.120	0.497	0.946	14.289	0.855	0.004	-0.048
Total	107	68.06%	100.00%	-0.502	252.820	852.290	0.509	0.839	11.444	0.813	-0.596	-0.395

Table III-9. The characteristics of the least efficient schools (continued)

Panel A: The most efficient schools in Model 1			
Country	**COMPRATIO**	**WEBCOMP**	**Math. score**
BRN	2.421	0.992	497.903
IDN	2.688	0.571	353.811
MYS	1.301	0.857	458.244
PHL	1.404	0.373	388.819
SGP	1.660	0.981	559.484
THA	1.651	0.890	423.652
Total	1.772	0.894	493.661
Panel B: The most efficient schools in Model 2			
Country	**COMPRATIO**	**WEBCOMP**	**Science score**
BRN	2.421	0.992	510.025
IDN	2.688	0.571	370.554
MYS	1.282	0.881	479.982
PHL	1.404	0.373	404.624
SGP	1.664	1.000	539.957
THA	1.678	0.927	424.522
Total	1.776	0.913	488.953
Panel C: The most efficient schools in Model 3			
Country	**COMPRATIO**	**WEBCOMP**	**Language score**
BRN	2.421	0.992	483.121
IDN	2.688	0.571	345.607
MYS	1.261	0.774	441.045
PHL	1.404	0.373	386.941
SGP	1.669	0.981	540.462
THA	1.651	0.890	384.475
Total	1.771	0.888	471.190

Notes: **n** represents the most (or the least) efficient schools for each country; **% (country)** represents the proportion of the most (or the least) efficient schools for each country; **% (frontier)** represents the proportion of the country's schools in the group of the most (or the least) efficient ones.

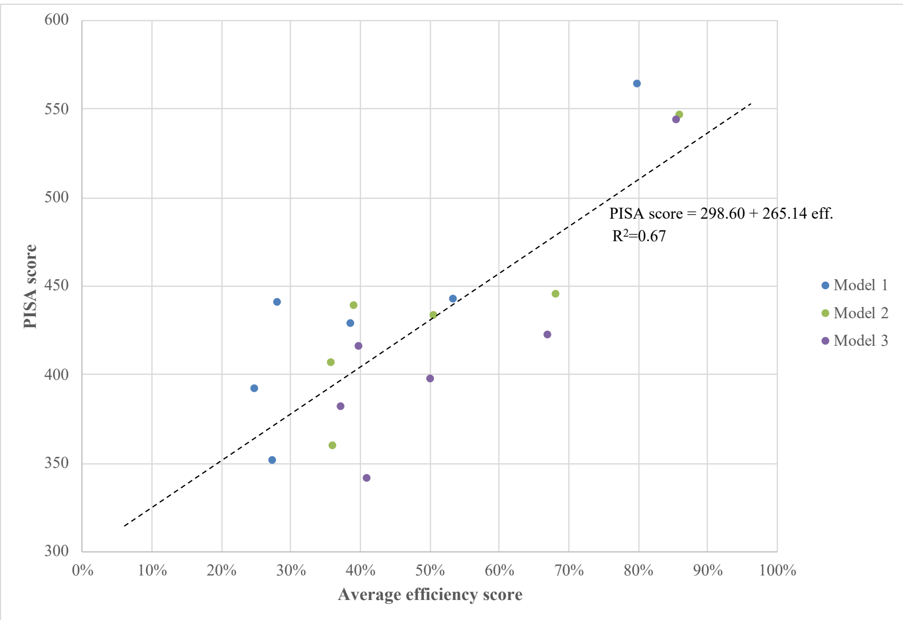
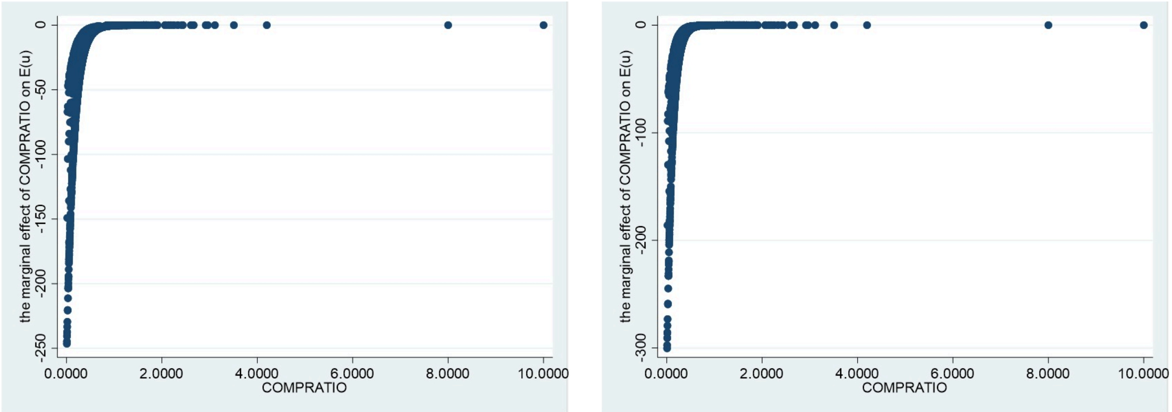


Figure III-3 The relation of average efficiency score and the PISA score, by country



(a) Model 1 (b) Model 2

Figure III-4 The marginal effect of COMPRATIO of inefficiency on $E[u]$

score. Results show that the efficiency score is statistically significant at the level of 5% with positive value, implying that the efficiency score is a good determinant of the PISA score. Notice that we also have a good value of R^2.

III.4.3 The influence of determinants of inefficiency

To investigate the influence of ICT on inefficiency, all ICT-related variables are included as the determinants of inefficiency. The estimation result is shown in Table III-6 under parameter σ_u^2. In all models, the ratio of computers connected to the internet (WEBCOMP) has no influence on inefficiency. The ratio of computers to the number of students (COMPRATIO) is significant in Model 1 and Model 2. Interpreting the sign of the determinants of inefficiency is not straightforward since the relationship between inefficiency and its determinants is nonmonotonic (Wang, 2002); it implies that depending on the values of the determinants, the influence on inefficiency can change directions in the sample. The marginal effect of determinants of inefficiency on the expected value of inefficiency is shown in Figure III-4. The figure only shows the determinants that are significant (i.e., COMPRATIO) in Model 1 and Model 2. The graphs indicate that for all observations, the marginal effect of the ratio of computers to the number of students is negative; thus, increasing this variable would decrease, on average, the level of inefficiency, or on the other words, increase efficiency. The size of the negative effect is larger when the value of this ratio is smaller. When this ratio continues to rise, the marginal effect moves toward 0 but not 0 (the line y = 0 acts as a horizontal asymptote). This observation indicates that none of the sample behaves that the ratio of computers has positive influence on inefficiency score. A closer investigation reveals that as one value increases in the ratio of computers to the number of enrollments, the level of efficiency increases, on average, by 37.295 (in Model 1) and 33.603 (in Model 2).

III.4.4 Additional checking

The results are rather stable across the different quality proxies. To verify the validity of the results, robustness checks are performed to analyze the impacts on the results. Specifically, first, the test whether the sign and significancy of inputs and determinants of inefficiency differs when another PISA score (called PV2) is used as education outcomes is performed. Second, by still using PV2, the individual efficiency is re-estimated; and correlation analysis between the previous result (as the baseline) and the new efficiency scores is also performed.

In the first examination, another PISA score as alternative education outcomes (called PV2) with similar inputs is used. In the literature of academic performance, student proficiencies actually cannot be observed. They are like missing data that must be inferred from the observed item responses (in PISA, they are item questions in the PISA assessment). There are several possible alternative approaches for making the inference. PISA uses the imputation methodology referred to as PVs. They are a selection of likely proficiencies for students that attained each score.[12] In this test, if the output is changed with other similar value which measures (as a proxy of) student proficiencies, it is expected that the result would not change that much. If so, the model is said to be not robust. Result of the first robustness analysis is shown in Table III-10. Note that the sign and significancy of inputs used are not changed much. The values of the coefficients, if one observes, are slightly similar; the difference is trivial.

In the second test, by still using PV2, the individual efficiency for each school is computed. A correlation analysis between the baseline model and the new efficiency scores is then performed. The result is shown in Table III-11. The result shows that the efficiency scores based on PV2 is very similar to the baseline model. Moreover, there are strong positive correlations (the coefficient correlations are more than 0.9) Consequently, the individual efficiency scores are robust to different PISA scores.

Next, an additional test to formally examine the existence of inefficiency in the model by using the generalized likelihood ratio (LR) is conducted. It is important since if the evidence for inefficiency is not found, then the model reduces to a standard regression model for which a simple OLS estimation would suffice. The LR statistic can be constructed based on the log-likelihood values of the OLS regression and the stochastic frontier model as follows.

$$\mathrm{LR} = -2[L(\mathrm{H}_0) - L(\mathrm{H}_1)], \tag{III-11}$$

where $L(\mathrm{H}_0)$ and $L(\mathrm{H}_1)$ are the log-likelihood values of the OLS regression and the stochastic frontier model, respectively. In this model, the LR test amounts to testing the hypothesis that there is no inefficiency (σ_u^2 = 0). The complication of the test is that the null hypothesis is on the boundary of the parameter value's permissible space, and therefore the LR test statistic does not have a standard chi-

[12] For details on how to construct PVs, see the *PISA 2018 Technical Report* (available online at: https://www.oecd.org/pisa/data/pisa2018technicalreport/).

Table III-10 Robustness analysis – parameters estimation using different PISA scores

Inputs	Model 1		Model 2		Model 3	
	Baseline	PV2	Baseline	PV2	Baseline	PV2
Constant	460.787**	470.138**	449.435**	446.043**	472.657**	479.715**
ESCS	54.862**	55.175**	49.270**	48.148**	52.833**	52.489**
M(S/L)MINS	0.010	0.002	0.042**	0.049**	-0.095**	-0.097**
SCHSIZE	0.000	0.000	0.000	0.000	-0.002	-0.002
PROP_girl	37.201**	36.745**	41.985**	37.554**	67.058**	66.907**
PROP_nat	-35.416**	-40.462**	-25.009**	-23.673**	-54.939**	-59.956**
STRATIO	-0.803**	-0.770**	-0.725**	-0.718**	-0.463**	-0.459**
PROATCE	-6.863	-6.748	-11.697**	-10.434**	-8.310**	-8.015**
EDUSHORT	-9.068**	-9.604**	-7.816**	-7.963**	-8.125**	-8.223**
STAFFSHORT	4.103**	4.174**	4.725**	5.147**	4.137**	4.183**
COMPRATIO	2.508	1.943	2.854	3.217	3.067	2.652
WEBCOMP	42.195**	39.216*	35.555**	37.898**	27.191**	27.325**
SCHTYPE:						
Private gov.-dep.	4.265	6.825	1.573	3.867	6.816	5.084
Public	29.813**	31.103**	24.256**	22.343**	25.856**	24.597**
SCHLOC:						
Small town	-3.567	-3.837	-0.526	-2.451	1.465	1.065
Town	-9.741**	-10.625**	-9.952**	-10.840**	-4.362	-5.089
City	-11.840**	-11.232**	-11.667**	-11.701**	-4.872	-5.147
Large city	23.686**	23.303**	16.942**	17.335**	35.583**	35.322**
σ_u^2:						
Constant	4.806**	4.622**	4.949**	4.932**	3.423	2.633
COMPRATIO	-13.352**	-11.642**	-17.718**	-17.513**	-16.437	-13.392
WEBCOMP	2.967	2.908	2.687	2.733	3.875	4.477
σ_v^2	7.481**	7.480**	7.323**	7.331**	7.350**	7.360**

*significant at the level of 10%
**significant at the level of 5%

Table III-11 Robustness analysis – correlation of different efficiency scores obtained from different PISA scores

	Model 1		Model 2		Model 3	
	Baseline	PV2	Baseline	PV2	Baseline	PV2
Baseline	1		1		1	
PV2	0.996	1	0.999	1	0.9863	1

square distribution. Coelli (1995) shows that, in such cases, the test has a mixture of chi-square distributions. This LR statistic a mix chi-square distribution with the degree of freedom equal to 3. According to the calculation, the value of LR is 26.009 for Model 1, 20.712 for Model 2, and 14.750 for Model 3. For the 5% level of significance, the critical value of the statistic is 7.045 (Kodde and Palm, 1986). Given the values of the LR statistics of all models which are way more than the critical value, the result indicates that the null hypothesis is rejected, meaning that indeed inefficiency does present. As such, there is a confidence towards the presence of inefficiency and ultimately, the stochastic frontier model is confirmed.

III.5 Discussion and Concluding Remarks

By implementing the SFA with heteroscedastic model, the efficiency scores of schools in the SEA region participated in the recent OECD PISA 2018 data is estimated. It can be considered as the first study which measures the efficiency at the school level analysis across country in the SEA region. The result reveals that Singapore has the (relatively) best performance among other SEA countries in terms of efficiency in education proxied by the PISA scores of mathematics, science, and language literacy. Singapore is widely recognized as one of the "high-performing education systems" (HPES)—a term used to describe education systems that excelled in PISA's league tables in the most recent years (Lee et al., 2014)—and has become the object of envy and emulation in many countries (see e.g., Barber and Mourshed, 2007; Darling-Hammond, 2010). At the top of the class on many of the international comparative measures on academic achievement, Singaporean students have surpassed many of their counterparts in traditional educational centers in North America and Europe and even Japan which was the first Asian country to modernize its education system (Luke et al. 2005).[13]

The average efficiency scores across all countries are found at 38.47% (in mathematics), 48.83% (in science), and 49.82% (in language), implying that, when considering the best-performing schools in the sample, on average, the other schools can improve their PISA score by about 50% to 62% with the currently available resources. This finding suggests the room for considerable efficiency improvements, especially when recalling that the best-in-class are schools operating in developing countries (thus, the relative efficiency scores are not computed comparing schools with counterparts in the developed countries).[14]

Among the thirteen inputs used in this study, seven inputs are statistically significant (at the level of 5%) to influence the education outcomes in all models, i.e., the index of economic, social, and cultural status, the share of girls, the proportion of native students, student-teacher ratio, index of educational material shortage, index of educational staff shortage, and the ratio of computers connected to the internet. Factors that might influence inefficiency are investigated by including ICT-related variables into the model. The result shows that the ratio of computers to the total number of students is significantly influencing inefficiency in Model 1 and Model 2, while the ratio of computers connected to the internet has no influence in all models.

One policy lesson that might be derived from this study is that even in the countries where schools' mean efficiency is low (i.e., Indonesia, Malaysia, and the Philippines), there are some schools belong to the most efficient group by utilizing their available resources. In this sense, conducting

---

[13] It is encouraged to view Deng and Gopinathan (2016) who offered an explanation for the education success of Singapore.

[14] According to the World Bank and the Department of Statistics of Singapore, Singapore is arguably the developing country despite of its high-income economy.

benchmarking analyses within each country is useful since it allows seeing (and measuring) the degree of internal (country-specific) heterogeneity in efficiency results. Therefore, international comparisons are meaningful because they set higher targets for all schools, independently of the geographical and institutional context where they operate (Agasisti and Zoido, 2019). In such a perspective, the international benchmarking is a great opportunity to enlarge the knowledge of practices and actions that make easier the transformation of inputs (in this study, they are student's and school's characteristics as well as investment in ICT) into output, i.e., students' academic achievement. In relation to the ICT infrastructure, it indicates that to get higher PISA score, the school has to increase the number of computers that connected to the internet, but to get higher efficiency, the school has to add the number of computers per se (not necessary connected to the internet).

CHAPTER IV. ICT AS A DRIVER OF EDUCATION OUTCOMES AND EFFICIENCY OF SCHOOLS IN SOUTH-EAST ASIA: THE TWO-STAGE SUPER-EFFICIENCY MODEL

This chapter presents a study which aims to measure efficiency of schools in South-East Asia by means of a two-stage super-efficiency model. The super-efficiency model allows for a decision-making unit (DMU)—in this case it is a school—to have an efficiency score higher than one. The idea is to leave out one DMU to be evaluated from the solution set. In the first stage of the analysis, the non-parametric super-efficiency data envelopment analysis (DEA)-based model is used to estimate the efficiency of schools in the SEA region. After finding out the efficiency scores, in the second stage of the analysis, a bootstrapped quantile regression is applied to examine factors that might explain the efficiency (called the determinants of efficiency).

The motivations of applying the bootstrapped quantile regression are as follows. Previous studies used the classic regression analysis with ordinary least square (OLS) estimation to examine the influence of determinants of efficiency. This procedure has several drawbacks. First, it cannot answer an important question: "Do the determinants influence efficiency levels differently for high-efficient schools and for those with low or average efficiency levels?" (Nwaogbe et al., 2018). The OLS estimates the mean (or the expected value) rate of change of the dependent variable as a conditional function of one or several independent variables. This feature can lead to inaccurate estimates of coefficients or to the omission of important relationship. A quantile regression, however, extends this estimation to any part of the dependent variable's distribution, i.e., to any selected quantile (or percentile), thereby facilitating a clearer interpretation of the relationship between variables that may otherwise have weak or no relation (Arshad et al., 2018). As such, quantile regression allows the simultaneous study of changes in specific portions of the distribution of the dependent variable to independent variables independently of the change and variability experienced by the rest of the distribution. This allows comparison of how some percentiles of the efficiency levels may be more affected by certain determinants of efficiency than other percentiles.

Next, as mentioned by Simar and Wilson (2007), the OLS regression procedure is flawed by the fact that usual inference on the obtained estimates of the regression coefficients is not available. Then, they proposed a bootstrap algorithm to obtain more accurate inference. In addition, the bootstrap procedure can be used to correct for the biases resulting from the correlation between the inputs or outputs of the first stage and the regressors of the second stage. Finally, one should also take into account the skewness of those efficiency scores (recall that the distribution of the efficiency score is not symmetric, see Gajewski et al., 2009; Santín and Sicilia, 2017; Sowlati and Paradi, 2004); thus, it is suggested to use the quantile regression which relies on the conditional quantiles rather than the conditional means as in OLS. As discussed in Angrist et al. (2006), quantile regression can explain changes in the distribution shape and spread, i.e., the skewness of the efficiency scores. Taking all these together, this study uses the bootstrapped quantile regression in the second stage of analysis. Due to these benefits of the bootstrapped quantile regression that is more robust to outliers than the OLS regression, to account for the issues of bias-correction (via bootstrap), as well as the skewness of efficiency scores, it is suggested that this procedure would present more insightful information compared to the conventional one.

This chapter is structured as follows. In Chapter IV.1, a literature review is conducted to present previous studies of the application of super-efficiency model in the education sector. The contributions of the study are also presented built upon the findings of the literature review. The methods used in this study are briefly described in Chapter IV.2, while data and variables used are shown in Chapter IV.3. The findings of this study are presented in Chapter IV.4; and finally, Chapter IV.5 provides discussion and concluding remarks.

IV.1 Literature Review and Contributions

This study aims to measure the efficiency of schools in South-East Asia by the means of two-stage super-efficiency model. As presented in Chapter III, literature about measuring efficiency of education in South-East Asia is quite limited. In this study, it extends one stream of literature of the application of two-stage super-efficiency model in the education sector. A literature review is then conducted in the Scopus database with the following search query: TITLE-ABS-KEY((“super-efficien*” OR “super efficien*”) AND (education OR universit* OR school)). The article type is restricted to peer-reviewed research article published in a journal and written in English. Only articles published in 1993 and afterwards are included since the idea of super-efficiency was proposed in 1993 by Andersen and Peterson. The search yields 67 articles. The first screening is performed by reading the title and abstract to verify the relevance of the extracted articles. In this way, 37 articles are excluded since the studies are not about the application of super-efficiency in education. The second screening is executed by carefully reading the full text of each article to address the eligibility of the articles. There are three articles whose full text cannot be accessed; thus, they are excluded. Five articles are excluded in this second screening procedure. The third screening is performed to investigate whether the articles included second-stage analysis (i.e., addressing the determinants of efficiency) or just performed one-stage analysis (i.e., only measuring the efficiency). Doing this way, only 8 articles included in this third screening procedure as briefly described in the following.

Zhang et al. (2022) examined the influence of the innovation ability of universities (IAU) on the efficiency of university–industry knowledge flow and investigated whether the level of provincial innovative agglomeration moderates the relationship between IAU and the efficiency of the university–industry knowledge flow. This study used the radial DEA model allowing for super-efficiency to measure knowledge research efficiency and knowledge transformation efficiency and then studied the influencing mechanism of the two kinds of efficiency using the spatial Tobit model with panel data from 2008 to 2017. The sample included 104 universities in China. Zhang and Wang (2022) measured the knowledge innovation efficiency (KIE) and knowledge transformation efficiency (KTE) of industry–university–research knowledge flow using the super-efficiency radial DEA. In the second stage, the authors aimed to study the impact mechanism of innovative city pilot policy on the knowledge flow dual efficiency (i.e., KIE/KTE) by adopting spatial difference-in-difference (SDID).

Zhou and Zhu (2021) measured the efficiency of scientific and technological (S&T) transformation in the Yangtze River Economic Belt, China, using the super-efficiency non-radial DEA. In the second stage, the authors used panel regression model to identify the influence of GDP, industrial structure, openness, human resources, scientific research projects, and international cooperation on the technology transformation efficiency of cities in that area. Chen and Shu (2021) explored S&T innovation performance of world-class universities in China from 2014 to 2019, based on the super-efficiency radial DEA model and the Malmquist index. In the second stage, the mixed OLS, fixed effect and random effect Tobit panel model was used to investigate the influence of the following determinants on the efficiency score: input factors quality index, the matching structure of scientific research elements index, the university and government relationship index, the industry-academia-research collaboration index, and the regional economic environment factor.

Wohlrabe et al. (2019) assessed the efficiency of 50 *elite* US universities using DEA, FHD, and two robust models: the order-m and order-α approaches. Only the two last approaches allow for super-efficiency. The authors then used the OLS regression analysis in the second stage. Türkan and Özel (2017) measured the efficiency of 43 state universities in Türkiye using the radial DEA; then, factors affecting efficiency are examined by Tobit and beta regression analysis. Agha et al. (2011) evaluated the relative technical efficiencies of academic departments at the Islamic University in Gaza during the years 2004-2006 using the radial DEA model allowing for super-efficiency. Further, multiple linear regression was used to develop a relationship between super-efficiency and input and output variables. Lastly, Lee (2009) evaluated the competitive effect of charter schools on hosting school districts using the super-efficiency radial DEA and regression analysis to obtain a DID estimator to measure the effect of charter school enrollment on charter hosting districts.

According to the literature review, there are three methods to measure efficiency allowing for super-efficiency, i.e., the radial and non-radial DEA as well as the order-m and order-α approaches which belong to the partial frontier analysis. In this study, the two models of full frontier analysis (i.e., radial and non-radial DEA) are implemented to measure school's efficiency in South-East Asia. In the second stage, the bootstrapped quantile regression is used to examine the influence of determinants of efficiency. The bootstrapped quantile regression is proposed to handle issues of robustness to outliers, bias-correction, and skewness of efficiency scores as have been discussed previously. Therefore, this study contributes to the literature of efficiency in education by extending the use of bootstrapped quantile regression in the second stage of two-stage super-efficiency model. To justify this contribution, a literature review is again conducted by adding the word "quantile" in the previous search query. It results no article. Other than in education sector, the practices of bootstrapped quantile regression in the second stage of efficiency measurement (but not in the super-efficiency model) can be seen in studies on agriculture (e.g., Frýd and Sokol, 2021), environmental or energy efficiency (e.g., Ibrahim et al., 2021; Moutinho et al., 2017; Qi et al., 2019), banking (e.g., Le et al., 2022), and aviation industry (e.g., Nwaogbe et al., 2018). The next contribution related to the role of ICT in education is that none of the previous studies incorporated ICT into their models as both input in the first stage and determinant of efficiency in the second stage.

IV.2 Methods

This study uses the two-stage super-efficiency model which allows for a DMU to have efficiency score more than one. In the first stage, the non-parametric DEA model is used to measure efficiency of schools in South-East Asia. Then, the slacks-based measure of super-efficiency (SSBM) is applied to rank and differentiate the efficient schools, leading to the super-efficient schools. In the second stage, the bootstrapped quantile regression is applied to investigate the influence of determinants of efficiency. The use of the bootstrapped quantile regression is due to the benefits that this procedure is better to handle for the issues of bias-correction, skewness of the efficiency scores, and robustness to the outliers.

IV.2.1 First stage of the analysis

DEA is a non-parametric technique to assess the efficiency of a DMU. In particular, it measures the ability of a DMU to minimize inputs to produce given outputs or, equivalently, to obtain maximum outputs from given inputs (Kumbhakar and Lovell, 2000). Consequently, a DMU is fully efficient if it produces the maximum possible outputs from a fixed level of inputs (in an output orientation), or if it uses the minimum possible inputs to produce a given level of outputs (in an input orientation). It is a non-parametric approach that requires very few assumptions in estimating efficiency compared to the parametric approach such as the SFA. In SFA, one has to define a functional form a priori and estimate the finite set of unknown parameters from the data. In addition, due to the use of the maximum likelihood method, the distribution of the inefficiency must be defined a priori. In DEA, these issues are not required. In addition, DEA can handle multiple outputs simpler that its parametric counterparts.

DEA can deal with both constant returns-to-scale (CRS), also called the CCR model after Charnes et al. (1978); and variable returns-to-scale (VRS), also called the BCC model after Banker et al. (1984). The CRS model is based on the assumption that constant returns-to-scale exists at the efficient frontiers whereas VRS model assumes variable returns-to-scale frontiers. There are two different specifications of the radial DEA model, i.e., input-oriented (IO) and output-oriented (OO). In IO model, DMUs minimize inputs while maintaining the same level of output. Conversely, in OO model, DMUs maximize their level of outputs while keeping inputs constant. Basically, the difference is the ability that a DMU has to control for the input or output quantity. If it can control input, then the IO version is preferable, the opposite is true in the OO case. In this study, the OO model is employed because schools strive to maximize education outcomes and cannot easily reduce their inputs at least in the short term. This study uses the VRS model with the output-oriented approach. This model is widely used in the literature for measuring efficiency in education (see e.g., Agasisti, 2014; Agasisti and Zoido, 2019; Aristovnik, 2013; Santín and Sicilia, 2018).

There are two models of DEA, i.e., the "radial model" and the "non-radial model". Historically, the radial measure, initially proposed by Charnes et al. (1978), was the first DEA model; whereas the non-radial model, represented by the SBM model by Tone (2001) was a latecomer.[15] For instance, in the IO case, the radial model deals mainly with proportionate reduction of input resources. In other words, if the DMU has two inputs, this model aims at obtaining the maximum rate of reduction with the same proportion, i.e., a radial contraction in the two inputs that can produce the current outputs. In contrast, the non-radial model puts aside the assumption of proportionate contraction in inputs and aims at obtaining maximum rates of reduction in inputs that may discard varying proportions of the original input resources.

In the radial DEA model with the VRS environment, the production possibility P_{VRS} is defined as

$$P_{VRS} = \{(x, y) | x \geq X\boldsymbol{\lambda}, y \leq Y\boldsymbol{\lambda}, \mathbf{e}\boldsymbol{\lambda} = 1, \boldsymbol{\lambda} \geq \mathbf{0}\}, \tag{IV-1}$$

where X is $m \times N$ matrix of inputs, Y is $s \times N$ matrix of outputs, m is the number of inputs, s is the number of outputs, N is the number of DMUs, $\mathbf{e}$ is a row vector with unity for all elements, and $\boldsymbol{\lambda}$ is the non-negative intensity vector. Then the radial VRS-OO model evaluates the efficiency of DMU$_o$ (o = 1, 2, …, N) by solving the following linear program:

$$\begin{aligned} & \max\ \omega \\ \text{subject to}\quad & X\boldsymbol{\lambda} \leq x_o \\ & \omega y_o - Y\boldsymbol{\lambda} \leq \mathbf{0} \\ & \mathbf{e}\boldsymbol{\lambda} = 1 \end{aligned}$$

15 The additive DEA model can also directly measure non-radial inefficiency but is unable to report the efficiency of the unit in a scalar value. As such, the SBM is regarded as the successor of the additive model.

$$\boldsymbol{\lambda} \geq \mathbf{0}. \tag{IV-2}$$

The $\omega \geq 1$ describes the output enlargement rate; meaning that the higher the value, the less efficient the DMU. The efficiency score $\theta\,(0 \leq \theta \leq 1)$ is obtained through $\theta = 1/\omega$. A DMU with the full ratio efficiency ($\theta = 1$) and with no slacks in any optimal solution is called an efficient DMU.

In the radial measure, an optimal solution is obtained if it satisfies two conditions, i.e., having efficiency score θ equal to 1 as well as having no slacks. Therefore, it is important to observe both the efficiency score and the slacks. Tone (1993) attempted to unify θ and slacks into a scalar measure. On the other hand, Charnes et al. (1985) developed the additive model of DEA which has no scalar measure. Tone (2001) then proposed the non-radial SBM model which deals with slacks of each input/output individually, independently, and integrate them into an efficiency measure. Moreover, the model has some important properties compared to the radial DEA as follows:

- *Unit invariant*: the measure is invariant with respect to the unit measurement of each input and output item.
- *Monotone*: the measure is monotone decreasing in each input and output slack.
- *Translation invariant*: the measure is invariant under parallel translation of the coordinate system applied.
- *Reference-set dependent*: the measure is determined only by consulting the reference-set of the observed DMU.

In the SBM model with VRS environment, the production possibility set is the same as that of the radial VRS model (Equation IV-1). In this sense, the DMU under consideration, DMU$_o$ (x_o, y_o), is expressed as $x_o = X\boldsymbol{\lambda} + s^-$ and $y_o = Y\boldsymbol{\lambda} - s^+$, where s^- and s^+ are defined as vector of input excesses and output shortfalls, respectively (both are called the slacks). In order to estimate the efficiency of DMU, the fractional program is formulated as follows

$$\min \rho_{SBM-VRS} = \frac{1 - \dfrac{1}{m}\sum_{i=1}^{m} \dfrac{s_i^-}{x_{io}}}{1 + \dfrac{1}{s}\sum_{r=1}^{s} \dfrac{s_r^+}{y_{ro}}}$$

$$\begin{aligned} \text{subject to} \quad & x_o = X\boldsymbol{\lambda} + s^- \\ & y_o = Y\boldsymbol{\lambda} - s^+ \\ & \mathbf{e}\boldsymbol{\lambda} = 1 \\ & \boldsymbol{\lambda} \geq \mathbf{0}, s^- \geq \mathbf{0}, s^+ \geq \mathbf{0}, \end{aligned} \tag{IV-3}$$

where $0 \leq \rho \leq 1$ is an index that satisfies the unit invariant and monotone properties. The fractional program in Equation (IV-3) can be transformed into a linear program using the Charnes-Cooper transformation (Charnes and Cooper, 1962). Let us multiply a scalar variable ($t > 0$) to both the denominator and nominator of the objective function of Equation (IV-3). It will not change ρ. Then t is adjusted so that the denominator becomes 1; this term is then moved to constraints. The objective is to minimize the numerator; thus, it becomes

$$\min \tau_{SBM-VRS} = t - \frac{1}{m}\sum_{i=1}^{m}\frac{ts_i^-}{x_{io}}$$

$$\text{subject to} \quad 1 = t + \frac{1}{s}\sum_{r=1}^{s}\frac{ts_r^+}{y_{ro}}$$
$$x_o = X\boldsymbol{\lambda} + s^-$$
$$y_o = Y\boldsymbol{\lambda} - s^+$$
$$\mathbf{e}\boldsymbol{\lambda} = 1$$
$$\boldsymbol{\lambda} \geq \mathbf{0}, s^- \geq \mathbf{0}, s^+ \geq \mathbf{0}, t > 0. \qquad \text{(IV-4)}$$

The model in Equation (IV-4) is a nonlinear programming problem. It can be transformed into a linear program by defining $S^- = ts^-$, $S^+ = ts^+$, and $\boldsymbol{\Lambda} = t\boldsymbol{\lambda}$ as follows

$$\min \tau_{SBM-VRS} = t - \frac{1}{m}\sum_{i=1}^{m}\frac{S_i^-}{x_{io}}$$

$$\text{subject to} \quad 1 = t + \frac{1}{s}\sum_{r=1}^{s}\frac{S_r^+}{y_{ro}}$$
$$tx_o = X\boldsymbol{\Lambda} + S^-$$
$$ty_o = Y\boldsymbol{\Lambda} - S^+$$
$$\mathbf{e}\boldsymbol{\lambda} = 1$$
$$\boldsymbol{\Lambda} \geq \mathbf{0}, S^- \geq \mathbf{0}, S^+ \geq \mathbf{0}, t > 0. \qquad \text{(IV-5)}$$

Let an optimal solution of Equation (IV-5) be (γ^*, t^*, $\boldsymbol{\Lambda}^*$, S^{-*}, S^{+*}). Then, the optimal solution of the SBM-VRS is defined by $\{\rho^* = \gamma^*, \boldsymbol{\lambda}^* = \boldsymbol{\Lambda}^*/t^*, s^{-*} = S^{-*}/t^*, s^{+*} = S^{+*}/t^*\}$. A DMU is SBM-efficient is $\rho^* = 1$. Then the SBM with the VRS-OO approach can defined by neglecting the numerator of the objective function of Equation (IV-6) as follows

$$\min \rho_{SBM-VRS-OUT} = \frac{1}{1 + \frac{1}{s}\sum_{r=1}^{s}\frac{s_r^+}{y_{ro}}}$$

$$\text{subject to} \quad 1 = t + \frac{1}{s}\sum_{r=1}^{s}\frac{S_r^+}{y_{ro}}$$
$$x_o = X\boldsymbol{\lambda} + s^-$$
$$y_o = Y\boldsymbol{\lambda}$$
$$\mathbf{e}\boldsymbol{\lambda} = 1$$
$$\boldsymbol{\lambda} \geq \mathbf{0}, s^- \geq \mathbf{0}. \qquad \text{(IV-6)}$$

To rank the efficient DMUs, the SBM of super-efficiency (SSBM) is applied. The radial super-efficiency model by Andersen and Peterson (1993) under the VRS environment suffers from having no feasible solution under certain condition. So far, the SBM of super-efficiency (SSBM) in DEA under the VRS environment is proved that this model is always feasible and has a finite optimum (Cooper et al., 2006). For SSBM with VRS model, let us define a production possibility set $P \setminus (x_o, y_o)$ spanned by (X, Y) excluding (x_o, y_o) as

$$P_{SSBM-VRS} \setminus (x_o, y_o) = \left\{ (\bar{x}, \bar{y}) \middle| \bar{x} \geq \sum_{j=1,\neq 0}^{N} \boldsymbol{\lambda}_j x_j, \bar{y} \leq \sum_{j=1,\neq 0}^{N} \boldsymbol{\lambda}_j y_j, \mathbf{e}\boldsymbol{\lambda} = 1, \bar{y} \geq \mathbf{0}, \boldsymbol{\lambda} \geq \mathbf{0} \right\}. \quad \text{(IV-7)}$$

Further, a subset $\bar{P}_{SSBM-VRS} \setminus (x_o, y_o)$ of $P_{SSBM\text{-}VRS} \setminus (x_o, y_o)$ is defined as

$$\bar{P}_{SSBM-VRS} \setminus (x_o, y_o) = P_{SSBM-VRS} \setminus (x_o, y_o) \cap \{\bar{x} \geq x_o \text{ and } \bar{y} \leq y_o\}. \quad \text{(IV-8)}$$

By the assumption of $X > 0$ and $Y > 0$, $\bar{P}_{SSBM-VRS} \setminus (x_o, y_o)$ is not empty.

The SSBM with VRS model is defined as the optimal objective function φ^* of the following

$$\varphi^*_{SSBM-VRS} = \min \varphi = \frac{\frac{1}{m}\sum_{i=1}^{m} \frac{\bar{x}_i}{x_{io}}}{\frac{1}{s}\sum_{r=1}^{s} \frac{\bar{y}_r}{y_{ro}}}$$

$$\begin{aligned} \text{subject to} \quad & \bar{x} \geq \sum_{j=1,\neq 0}^{N} \boldsymbol{\lambda}_j x_j \\ & \bar{y} \leq \sum_{j=1,\neq 0}^{N} \boldsymbol{\lambda}_j y_j \\ & \bar{x} \geq x_o \\ & \bar{y} \leq y_o \\ & \sum_{j=1,\neq 0}^{N} \boldsymbol{\lambda}_j = 1 \\ & \bar{y} \geq \mathbf{0}, \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned} \quad \text{(IV-9)}$$

The fractional program in Equation (IV-9) can be transformed into a linear program using the Charnes-Cooper transformation (Charnes and Cooper, 1962) as follows

$$\delta^*_{SSBM-VRS} = \min \delta = \frac{1}{m}\sum_{i=1}^{m} \frac{\tilde{x}_i}{x_{io}}$$

$$\begin{aligned} \text{subject to} \quad & 1 = \frac{1}{s}\sum_{r=1}^{s} \frac{\tilde{y}_r}{y_{ro}} \\ & \tilde{x} \geq \sum_{j=1,\neq 0}^{N} \Lambda_j x_j \\ & \tilde{y} \leq \sum_{j=1,\neq 0}^{N} \Lambda_j y_j \\ & \tilde{x} \geq t x_o \end{aligned}$$

$$\tilde{y} \leq t y_o$$

$$\sum_{j=1,\neq 0}^{N} \boldsymbol{\lambda}_j = 1$$

$$\tilde{y} \geq \mathbf{0}, \boldsymbol{\Lambda} \geq \mathbf{0}, t > 0. \qquad \text{(IV-10)}$$

Let an optimal solution of Equation (V-10) be (δ^*, $\tilde{x}^*$, $\tilde{y}^*$, t^*, $\boldsymbol{\Lambda}^*$). Then, the optimal solution of the SSBM-VRS is defined by $\varphi^* = \delta^*$, $\boldsymbol{\lambda}^* = \boldsymbol{\Lambda}^*/t^*$, $\bar{x}^* = \tilde{x}^*/t^*$, $\bar{y}^* = \tilde{y}^*/t^*$. Then the SSBM with the VRS-OO approach can defined as follows

$$\varphi^*_{SSBM-VRS-OO} = \min \varphi = \frac{1}{\frac{1}{s}\sum_{r=1}^{s}\frac{\bar{y}_r}{y_{ro}}}$$

subject to $\bar{x} \geq \sum_{j=1,\neq 0}^{N} \boldsymbol{\lambda}_j x_j$

$$\bar{y} \leq \sum_{j=1,\neq 0}^{N} \boldsymbol{\lambda}_j y_j$$

$$\bar{x} = x_o$$

$$0 \leq \bar{y} \leq y_o$$

$$\sum_{j=1,\neq 0}^{N} \boldsymbol{\lambda}_j = 1$$

$$\boldsymbol{\lambda} \geq \mathbf{0}. \qquad \text{(IV-11)}$$

IV.2.2 Second stage of the analysis

This second stage of analysis is devoted to investigating the influence of the determinants of efficiency by the means of the bootstrapped quantile regression. The quantile regression was introduced by Koenker and Bassett (1978) and has become an increasingly important tool in statistical analysis. Suppose the random variable y_i (i = 1, 2, … T) is a random sample generated by a linear regression $Y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$, where Y is the dependent variable, $\mathbf{x}$ represents the vector of independent variables, $\boldsymbol{\beta}$ is the corresponding vector of parameters, and ε_i is a random error whose conditional quantile distribution has a zero mean. The κth quantile, $0 < \kappa < 1$ of the explained variables has the form of

$$Q_\kappa (y_i|\mathbf{x}_i) = \mathbf{x}_i'\mathbf{b}, \qquad \text{(IV-12)}$$

where **b** estimate shows the quantile regression κth and solves the minimization problem

$$\arg\min_b \left[\sum_{i\in(i:y_i \geq \mathbf{x}_i'\mathbf{b})} \kappa |y_i - \mathbf{x}_i'\mathbf{b}| + \sum_{i\in(i:y_i < \mathbf{x}_i'\mathbf{b})} (1-\kappa)|y_i - \mathbf{x}_i'\mathbf{b}| \right]. \qquad \text{(IV-13)}$$

Since κ is equal to the different values, different parameter estimations will be obtained. As Equation (IV-13) shows, the quantile regression minimizes the sum of absolute errors and, therefore, this method is more robust against outliers compared to the OLS (Guan, 2003).

Next, in the bootstrapped quantile regression procedure, the methodological advantages of the quantile regression are amplified by the bootstrapping technique. The bootstrap is a robust statistical procedure which could be employed for a small sample analysis without relying on the error terms normality assumption. In this method, the standard deviation of the parameter coefficients is created by the bootstrap re-sampling method. The bootstrapped standard error is estimated following these three steps (Efron and Tibshirani, 1986)

1. Draw a large number R from the bootstrapped samples (p^*) by a resample method with replacement, say p_1^*, p_2^*, … p_R^*.
2. Estimate the bootstrapped parameter coefficient (b^*) from each of the bootstrapped samples, say b_1^*, b_2^*, … b_R^*.
3. Calculate the standard error from the distribution of the bootstrapped parameter coefficients $SE(b^*)$ as (Efron and Tibshirani, 1986)

$$SE(b^*) = \sqrt{\frac{\sum_{k=1}^{R}\left(b_k^* - b^*(m)\right)^2}{R-1}}, \qquad \text{(IV-14)}$$

where b_k^* is the bootstrapped estimate of the parameter coefficients from the bootstrapped sample p_i^*, and $b^*(m)$ is the mean value of the bootstrapped parameter coefficient in all the bootstrapped samples and is defined as

$$b^*(m) = \frac{\sum_{k=1}^{R} b_k^*}{R}. \qquad \text{(IV-15)}$$

IV.3 Data

The data is taken from the recent PISA 2018 data. Among the seven countries participating in PISA 2018, this study excludes Vietnam (see again Chapter III.3 for the discussion of this issue). Since this study is conducted at the school level, the variables which are at student level are weighted using W_FSTUWT, the final student weight provide by PISA, to get school level variables. The output (education outcomes) is proxied by the plausible value (PV) of mathematics (PVMATH), science (PVSCIE), and reading (PVREAD). As inputs, four variables are used, i.e., (i) the (inverse) of student-teacher ratio (INVSTRATIO) which measures the quantity of human resources; (ii) the ratio of computers at school to the total number of students for educational purposes (COMPRATIO); (iii) the

Table IV-1 Average value of inputs and outputs by country

Variables	Total	BRN	IDN	MYS	PHL	SGP	THA
PVMATH	432.367	443.438	395.175	440.258	357.714	562.063	424.324
PVSCIE	412.792	422.414	384.233	414.144	345.105	544.099	395.753
PVREAD	435.500	446.364	408.920	437.512	362.726	544.757	432.394
INVSTRATIO	0.080	0.119	0.074	0.100	0.051	0.093	0.072
COMPRATIO	0.531	0.910	0.333	0.380	0.298	1.097	0.537
WEBCOMP	0.885	0.970	0.893	0.874	0.633	0.997	0.942
ESCS	-0.977	-0.161	-1.447	-0.778	-1.336	0.126	-1.287

ratio of computers that were connected to the internet (WEBCOMP); these latest two inputs act as indirect measures of schools' facilities related to ICT; and (iv) to control for students' background, the index of school's economic, social, and cultural status (ESCS) is included.

After defining the outputs and inputs, the missing data is dropped (around 18%). At the end, the dataset comprises 1,051 schools from six countries. The average values of variables used in this study are shown in Table IV-1. The average PISA score in mathematics is 432.367, while in science and language are 412.792 and 435.5, respectively. Singapore has the highest scores in all domains, whereas the Philippines has the lowest scores in all domains. Since ESCS is scaled to have mean of zero and standard deviation of one across senate-weighted OECD countries, the fact that the mean of ESCS of these six countries is -0.977 indicates that the economic, social, and cultural condition of the sampled schools is below the average values in the OECD countries. Singapore has the highest average value of ESCS. In terms of (inverse of) student-teacher ratio, the Philippines has the lowest score (i.e., 0.051, meaning that there is one teacher for every 20 students) while Brunei Darussalam has the highest score. Regarding the ICT-related variables, Singapore has both the highest ratios (i.e.,1.097 and 0.997 for COMPRATIO and WEBCOMP, respectively), while the Philippines has the lowest scores in both ratios.

As the determinants of efficiency, seven variables is included. In this sense, the causality about the relationship between efficiency and the determinants of efficiency cannot be verified as the separability condition is not tested (Daraio et al., 2018). Still, the statistical associations showed in this study can add the bulk of limited existence evidence for the efficiency of schools in this region. The first group of determinants is school's characteristics, including the proportion of girls (PROP_girl) and school size (SCHSIZE). Taken together, these variables aim at capturing if school's efficiency (the ability to maximize students' achievement given the available resources) is influenced or not by a set of school's characteristics. The second group of determinants is school's resources, including proportion of fully certified teacher (PROATCE), the index of educational material shortage (EDUSHORT), and the index of educational staff shortage (STAFFSHORT). While the efficiency model takes the *quantity* of human and material resources into account, these three variables aim at identifying the influence of the *quality* of the resources on school's efficiency. ICT infrastructure (COMP-

Table IV-2 Descriptive statistics of the determinants of efficiency

Variables	Mean	Median	Std. Dev.	Min.	Max.
PROP_GIRL	0.498	0.456	0.145	0	1
SCHSIZE	1,285.470	942	1,363.633	21	11,990
PROATCE	0.833	0.987	0.292	0	1
EDUSHORT	0.173	0.1	1.189	-1.421	2.96
STAFFSHORT	-0.043	0.0055	1.052	-1.455	4.044
COMPRATIO	0.531	0.357	0.545	0.008	8
WEBCOMP	0.885	1	0.249	0.01	1

RATIO and WEBCOMP) is also included to investigate the influence of these variables on school's efficiency.[18] The descriptive statistics of the determinants of efficiency is shown in Table IV-2.

IV.4 Results

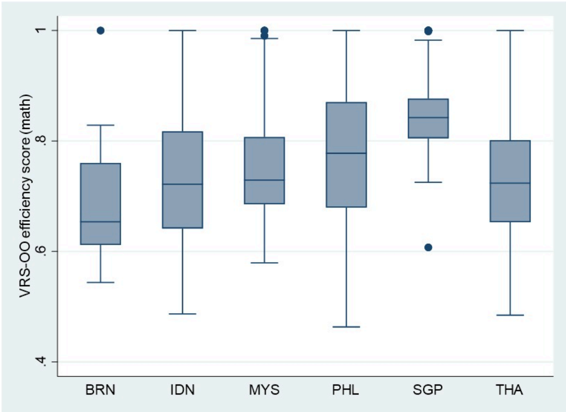
Four cases are generated in this study, each with different outputs and identical inputs. The output of the first case is PVMATH, or the PISA score of mathematics, while the outputs of the second and the third cases are the PISA score of science (PVSCIE) and reading (PVREAD), respectively. Lastly, in the fourth case, all PISA domains are used as outputs. Doing this way, it is attempted to observe different behaviors that might take place.

IV.4.1 Result of the first stage

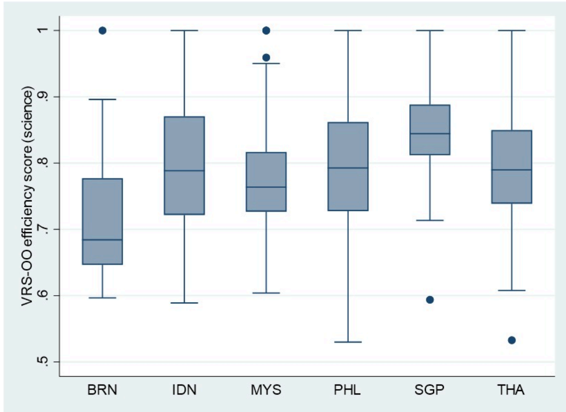
In the first stage, the radial and non-radial DEA are used to measure the efficiency of schools in six countries in South-East Asia. The distributions of the efficiency score for the radial model in all cases are shown as box plots in Figure IV-1. Notice that the boxes represent the 25th and 75th percentiles of efficiency scores' distribution, the whiskers are the upper and lower "adjacent values", respectively, the points are outliers, and the median is represented as the horizontal line inside the boxes. The medians of the efficiency scores are found at 0.749, 0.794, 0.769, and 0.800 for the first, second, third, and fourth case, respectively. It seems that when it combines all the PISA scores, the efficiency of schools in these six counties are higher than when it only considers only a single PISA score. Such figure suggests room for improvements since on average, the inefficient schools can improve their test scores by about more than 20% with the currently available resources when considering the efficient schools in this sample.

At a country level, Singapore has the highest average value of the efficiency score for the radial model in all cases (see Table IV-3). Interestingly, despite of the low PISA score, the Philippines is at the second place after Singapore in the first, third, and fourth case (Thailand is at the second place in the second case). Compared to the average score obtained from all countries, in terms of mathematics

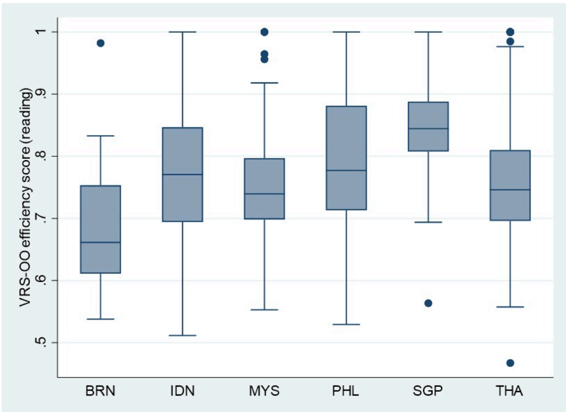
[18] Recall Chapter III.3 for the description of the variables.



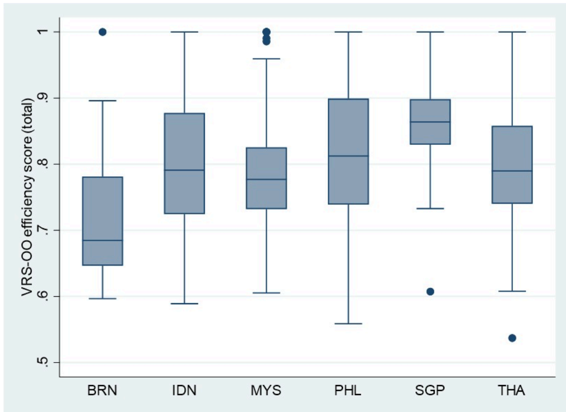
(a) Case 1 (mathematics)



(b) Case 2 (science)



(c) Case 3 (reading)



(d) Case 4 (all domains)

Figure IV-1 The distribution of efficiency scores for the radial model, by country

literacy, only Singapore and the Philippines are higher, while in science and all domains, only Singapore is higher than the average score, and lastly, in reading literacy, Malaysia, the Philippines, and Singapore get higher score than the average score of all countries.

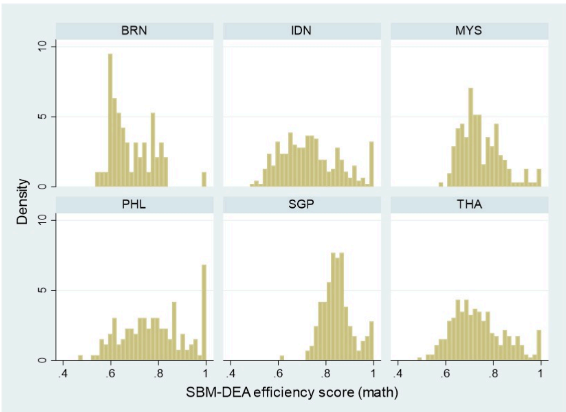
In terms of standard deviation, the Philippines has high value, implying a higher degree of heterogeneity of school's efficiency. In other hand, Singapore has the lowest standard deviation, representing the information that this country might maintain *uniform* education system across its schools. This figure also highlights the presence of a wider distribution of the two most efficient countries.

Schools in the Philippines have the lowest efficiency scores in terms of mathematics and science among other schools in the sample, while in terms of reading and all literacies, schools in Thailand have the lowest scores. Notice that every country has the most efficient score represented by the efficiency score of more than one but Brunei Darussalam (the most efficient school in this country has the efficiency score of 0.982 in terms of reading literacy).

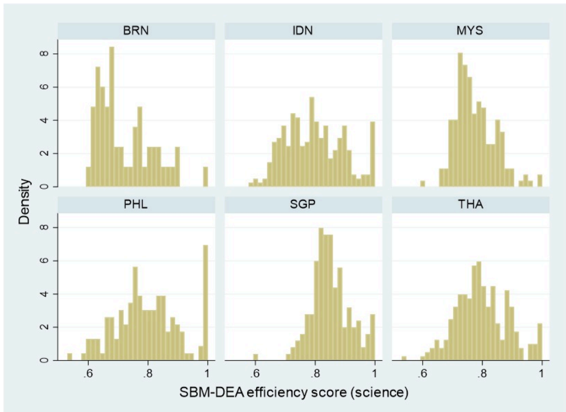
Moving to the non-radial model, the distributions of the efficiency score in all cases are shown in Figure IV-2. One might notice that visually, the *patterns* are different for each country. Some counties

Table IV-3 Descriptive statistics of the efficiency scores for the radial model

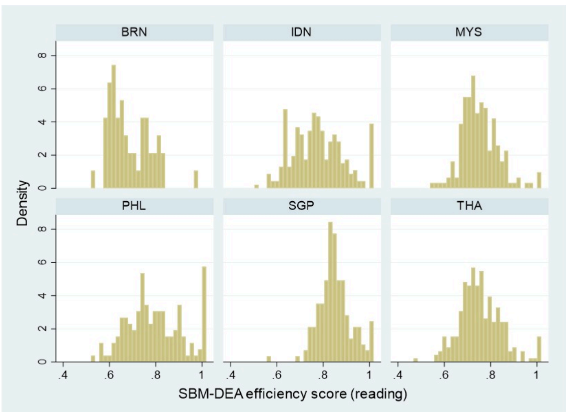
Country	Average				Standard Deviation			
	Math.	Science	Reading	All	Math.	Science	Reading	All
BRN	0.686	0.721	0.686	0.722	0.092	0.091	0.089	0.091
IDN	0.733	0.798	0.777	0.802	0.123	0.098	0.109	0.100
MYS	0.747	0.777	0.752	0.786	0.086	0.068	0.079	0.074
PHL	0.781	0.800	0.795	0.816	0.133	0.110	0.118	0.113
SGP	0.848	0.852	0.851	0.866	0.065	0.066	0.070	0.064
THA	0.736	0.799	0.754	0.802	0.110	0.087	0.090	0.088
All countries	0.758	0.799	0.777	0.807	0.115	0.092	0.103	0.095
Country	**Min.**				**Max.**			
	Math.	Science	Reading	All	Math.	Science	Reading	All
BRN	0.544	0.597	0.538	0.597	1	1	0.982	1
IDN	0.487	0.589	0.512	0.589	1	1	1	1
MYS	0.579	0.604	0.553	0.605	1	1	1	1
PHL	0.463	0.530	0.529	0.559	1	1	1	1
SGP	0.607	0.594	0.564	0.607	1	1	1	1
THA	0.485	0.533	0.467	0.537	1	1	1	1
All countries	0.463	0.530	0.467	0.537	1	1	1	1



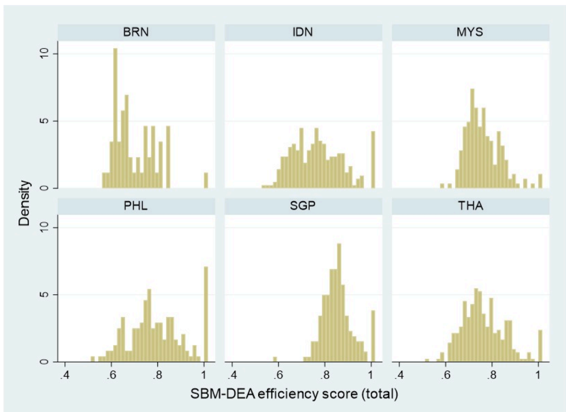
(a) Case 1 (mathematics)



(b) Case 2 (science)



(c) Case 3 (reading)



(d) Case 4 (all domains)

Figure IV-2 The distribution of efficiency scores for the non-radial model, by country

Table IV-4 Descriptive statistics of the efficiency scores for the non-radial model

Country	Average				Standard Deviation			
	Math.	**Science**	**Reading**	**Total**	**Math.**	**Science**	**Reading**	**Total**
BRN	0.686	0.721	0.686	0.701	0.092	0.091	0.089	0.090
IDN	0.733	0.798	0.777	0.770	0.123	0.098	0.109	0.100
MYS	0.747	0.777	0.752	0.761	0.086	0.068	0.079	0.075
PHL	0.781	0.800	0.795	0.794	0.133	0.110	0.118	0.119
SGP	0.848	0.852	0.851	0.855	0.065	0.066	0.070	0.066
THA	0.736	0.799	0.754	0.764	0.110	0.087	0.090	0.095
All	0.758	0.799	0.777	0.780	0.115	0.092	0.103	0.102
Country	**Min.**				**Max.**			
	Math.	**Science**	**Reading**	**Total**	**Math.**	**Science**	**Reading**	**Total**
BRN	0.544	0.597	0.538	0.562	1	1	0.982	1
IDN	0.487	0.589	0.512	0.538	1	1	1	1
MYS	0.579	0.604	0.553	0.580	1	1	1	1
PHL	0.463	0.530	0.529	0.522	1	1	1	1
SGP	0.607	0.594	0.564	0.591	1	1	1	1
THA	0.485	0.533	0.467	0.511	1	1	1	1
All	0.463	0.530	0.467	0.511	1	1	1	1

have skewed distributions. The data that skews to the right is usually a result of a lower boundary in a data set; in this case, it is an indication of low efficiency (for example, Brunei Darussalam has the lowest average efficiency score, i.e., 0.686, 0.721, 0.686, and 0.701 in respect to the PISA score of mathematics, science, reading, and all domains, respectively). In details, the descriptive statistics of the efficiency score is shown in Table IV-4.

In subsequent analysis, we differentiate the most efficient schools by SBM of super-efficiency model (SSBM) of Tone (2002).[20] In terms of mathematics, there are 47 super-efficient schools, while in science, reading, and all domains, there are 45, 50, and 59 super-efficient schools, respectively. The descriptive statistics of the super-efficient schools is shown in Table IV-5. Even though this class of school is considered as the *best performers* among others, the discrepancy in terms of PISA score (all domains) is still high, which is reflected in the standard deviation. The super-efficient schools suffer from the low point of ESCS: the average values of ESCS in all cases are below the average values of ESCS in the OECD countries. The condition of ICT infrastructure at the super-efficient schools is also not quite promising since there are still many schools having low value of both ratios: COMPRATIO and WEBCOMP.

IV.4.2 Result of the second stage

In the second stage, the influence of the determinants of efficiency is examined. The dependent variable is the efficiency score while the independent variables are the determinants of efficiency. The bootstrapped quantile regression is used for investigating the influence of the determinants of efficiency

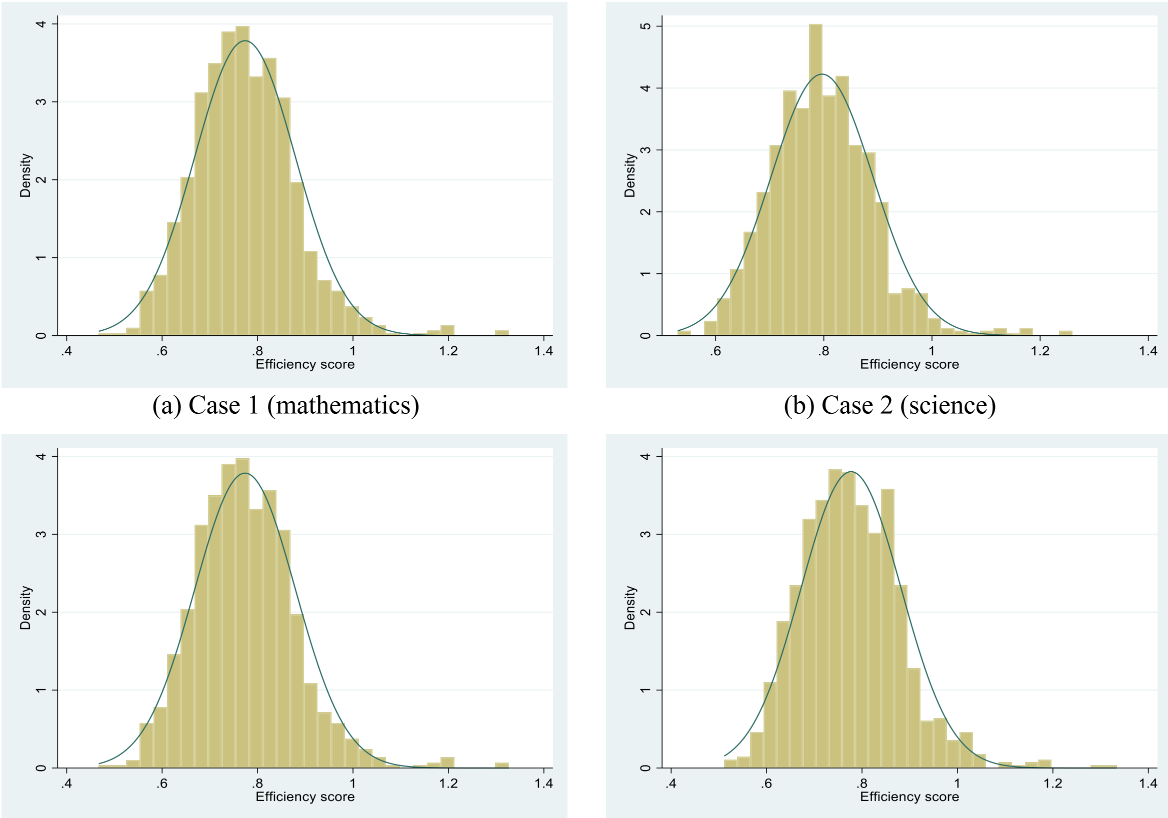
[20] The SBM-DEA model cannot discriminate inefficient DMUs for they will get the same efficiency score of 1.

Table IV-5 Descriptive statistics of the super-efficient schools

Variables	Average	Median	Standard Deviation	Min.	Max.
Panel A – Case 1 (Mathematics)					
Variables	**Average**	**Median**	**Standard Deviation**	**Min.**	**Max.**
Efficiency score	1.093	1.043	0.110	1.005	1.442
PV1MATH	500.302	459.001	125.149	345.119	687.987
PV1SCIE	491.194	443.240	113.978	350.631	676.031
PV1READ	475.568	430.334	125.358	311.676	688.732
INVSTRATIO	0.068	0.057	0.038	0.030	0.151
COMPRATIO	0.277	0.190	0.320	0.016	1.239
WEBCOMP	0.611	0.683	0.409	0.018	1.000
ESCS	-0.682	-0.853	1.305	-2.718	0.909
Panel B – Case 2 (Science)					
Variables	**Average**	**Median**	**Standard Deviation**	**Min.**	**Max.**
Efficiency score	1.092	1.089	0.079	1.002	1.260
PV1MATH	495.269	459.001	119.201	345.119	682.387
PV1SCIE	499.661	471.147	110.508	350.631	681.954
PV1READ	479.940	448.951	124.569	311.676	690.755
INVSTRATIO	0.068	0.058	0.037	0.030	0.151
COMPRATIO	0.300	0.201	0.347	0.016	1.380
WEBCOMP	0.621	0.683	0.399	0.018	1.000
ESCS	-0.812	-1.119	1.378	-2.718	0.921
Panel C – Case 3 (Reading)					
Variables	**Average**	**Median**	**Standard Deviation**	**Min.**	**Max.**
Efficiency score	1.089	1.045	0.097	1.000	1.327
PV1MATH	485.713	439.269	117.657	345.119	685.256
PV1SCIE	484.790	441.104	107.844	350.631	681.954
PV1READ	475.198	431.599	122.450	311.676	690.755
INVSTRATIO	0.070	0.053	0.042	0.030	0.208
COMPRATIO	0.375	0.221	0.411	0.016	1.380
WEBCOMP	0.598	0.595	0.419	0.018	1.000
ESCS	-0.955	-1.120	1.343	-2.817	0.921
Panel D – Case 4 (All domains)					
Variables	**Average**	**Median**	**Standard Deviation**	**Min.**	**Max.**
Efficiency score	1.064	1.026	0.086	1.000	1.334
PV1MATH	504.979	461.857	121.050	345.119	687.987
PV1SCIE	499.794	444.117	112.100	350.631	681.954
PV1READ	487.446	449.203	122.905	311.676	690.755
INVSTRATIO	0.073	0.067	0.040	0.030	0.208
COMPRATIO	0.407	0.227	0.425	0.016	1.380
WEBCOMP	0.646	0.850	0.398	0.018	1.000
ESCS	-0.801	-0.956	1.337	-2.817	0.921

due to the benefits compared to the traditional OLS. Figure IV-3 shows the distribution of the efficiency score, and it indicates the skewed distributions. The correlation matrix reported in Table IV-6 suggests no severe multicollinearity between the independent variables so the implementation of the quantile regression is justified.

The estimation result is shown in Table IV-7 with five quantile results of $Q_{0.1}$, $Q_{0.25}$, $Q_{0.5}$, $Q_{0.75}$, and $Q_{0.9}$. The number of bootstrap replications is set to 500. PROP_GIRL is significant (at least at the level of 10%) with positive value in all quantiles but not in the upper quantile ($Q_{0.9}$). In $Q_{0.9}$, this variable is only significant at the third case. It seems that among the super-efficient schools, the proportion of



(a) Case 1 (mathematics)

(b) Case 2 (science)

(c) Case 3 (reading)

(d) Case 4 (all domains)

Figure IV-3 The distribution of efficiency scores for the non-radial super-efficiency model

Table IV-6 Correlation matrix of the determinants of efficiency

Variables	PROP_ GIRL	SCHSIZE	PRO- ATCE	EDU- SHORT	STAFF- SHORT	COMP- RATIO	WEB- COMP
PROP_GIRL	1.000						
SCHSIZE	0.034	1.000					
PROATCE	0.045	0.195**	1.000				
EDUSHORT	-0.056	0.009	-0.068**	1.000			
STAFFSHORT	-0.062**	-0.145**	-0.072**	0.554**	1.000		
COMPRATIO	0.0006	-0.191**	-0.0005	-0.314**	-0.145**	1.000	
WEBCOMP	0.051	-0.140**	-0.002	-0.199**	-0.056*	0.145**	1.000

*significant at the level of 10%
**significant at the level of 5%

female students does not influence their efficiency. On the other hand, this variable does matter for the low and middle efficient schools (this variable is significant in $Q_{0.1}$, $Q_{0.25}$, $Q_{0.5}$, and $Q_{0.75}$ in all cases).

The number of students is statistically significant with very small value in all cases and all quantiles. It indicates that even though the influence is significant, the effect is negligible due to the very small value. EDUSHORT negatively influences school's efficiency in all cases and all quantiles, while STAFFSHORT is not significant in the lower and middle quantiles, but it is found significant in the upper quantile with positive value. Similar condition to STAFFSHORT is found in PROATCE. It seems

Table IV-7 Parameters estimation of the bootstrapped quantile regression

Variables	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.9}$
Panel A – Case 1 (Mathematics)					
Constant	0.676** (0.032)	0.702** (0.026)	0.790** (0.023)	0.886** (0.030)	1.020** (0.057)
PROP_GIRL	0.108** (0.035)	0.117** (0.027)	0.110** (0.030)	0.069* (0.039)	0.018 (0.046)
SCHSIZE	0.00002** (0.000005)	0.00002** (0.000004)	0.00002** (0.000004)	0.00003** (0.000001)	0.00003** (0.000001)
PROATCE	0.012 (0.017)	0.018 (0.014)	-0.019 (0.014)	-0.024 (0.016)	-0.050* (0.025)
EDUSHORT	-0.020** (0.004)	-0.024** (0.005)	-0.037** (0.005)	-0.035** (0.005)	-0.032** (0.006)
STAFFSHORT	0.001 (0.005)	0.003 (0.004)	0.004 (0.006)	0.008* (0.005)	0.014** (0.006)
COMPRATIO	0.008 (0.012)	0.016** (0.006)	0.009 (0.009)	-0.0001 (0.006)	-0.008 (0.009)
WEBCOMP	-0.150** (0.023)	-0.139** (0.018)	-0.124** (0.015)	-0.129** (0.023)	-0.139** (0.046)
Panel B – Case 2 (Science)					
Constant	0.709** (0.027)	0.736** (0.022)	0.793** (0.025)	0.884** (0.032)	1.004** (0.054)
PROP_GIRL	0.087** (0.030)	0.119** (0.022)	0.131** (0.033)	0.083** (0.037)	0.048 (0.035)
SCHSIZE	0.00001** (0.000002)	0.000011** (0.000002)	0.000013** (0.00003)	0.000013** (0.000004)	0.000014** (0.000004)
PROATCE	0.016 (0.015)	-0.003 (0.013)	-0.009 (0.012)	-0.015 (0.015)	-0.042** (0.019)
EDUSHORT	-0.015** (0.004)	-0.014** (0.004)	-0.018** (0.004)	-0.024** (0.005)	-0.023** (0.005)
STAFFSHORT	0.005 (0.005)	0.001 (0.004)	0.005 (0.004)	0.011** (0.005)	0.010** (0.005)
COMPRATIO	0.009 (0.013)	0.011** (0.005)	0.005 (0.005)	-0.008 (0.010)	-0.007 (0.010)
WEBCOMP	-0.099** (0.018)	-0.081** (0.012)	-0.082** (0.018)	-0.081** (0.022)	-0.109** (0.046)
Panel C – Case 3 (Reading)					
Constant	0.648 (0.031)	0.682** (0.028)	0.793** (0.025)	0.873** (0.027)	1.031** (0.050)
PROP_GIRL	0.189** (0.049)	0.162** (0.032)	0.140** (0.024)	0.103** (0.035)	0.100** (0.032)
SCHSIZE	0.000015** (0.000003)	0.000013** (0.000003)	0.000014** (0.000003)	0.000016** (0.000004)	0.000013** (0.000006)
PROATCE	0.0001 (0.014)	0.012 (0.014)	-0.017 (0.015)	-0.022 (0.015)	-0.064** (0.028)
EDUSHORT	-0.013 (0.004)	-0.013** (0.005)	-0.021** (0.004)	-0.025** (0.005)	-0.030** (0.005)
STAFFSHORT	-0.002 (0.006)	0.001 (0.005)	0.001 (0.005)	0.006 (0.005)	0.010** (0.005)
COMPRATIO	0.003 (0.012)	0.013* (0.007)	0.006 (0.007)	-0.007 (0.006)	-0.010 (0.009)
WEBCOMP	-0.108 (0.016)	-0.097** (0.021)	-0.113** (0.017)	-0.099** (0.018)	-0.163** (0.041)

Table IV-7 Parameters estimation of the bootstrapped quantile regression (continued)

Variables	$Q_{0.1}$	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$	$Q_{0.9}$
Panel D – Case 4 (All domains)					
Constant	0.694** (0.027)	0.710** (0.023)	0.793** (0.023)	0.866** (0.031)	1.026** (0.054)
PROP_GIRL	0.125** (0.038)	0.128** (0.025)	0.128** (0.032)	0.101** (0.030)	0.053 (0.035)
SCHSIZE	0.000017** (0.000003)	0.000015** (0.000003)	0.000017** (0.000003)	0.000018** (0.000004)	0.000018** (0.000006)
PROATCE	0.008 (0.014)	0.011 (0.015)	-0.016 (0.013)	-0.013 (0.016)	-0.058** (0.025)
EDUSHORT	-0.016** (0.004)	-0.016** (0.004)	-0.025** (0.004)	-0.029** (0.005)	-0.031** (0.005)
STAFFSHORT	0.002 (0.005)	0.001 (0.005)	0.004 (0.004)	0.011* (0.006)	0.012** (0.005)
COMPRATIO	0.006 (0.011)	0.013** (0.006)	0.012 (0.008)	-0.003 (0.006)	-0.009 (0.008)
WEBCOMP	-0.133** (0.017)	-0.105** (0.013)	-0.111** (0.016)	-0.099** (0.024)	-0.141** (0.044)

Notes: numbers in parentheses denote the bootstrapped standard error.

*significant at the level of 10%
**significant at the level of 5%

that the *quality* of human resources (i.e., teacher and educational staff) plays an important role in determining efficiency at super-efficient schools.

Regarding the ICT infrastructure, COMPRATIO and WEBCOMP behave differently. The ratio of computers to the total number of students is not significant in all quantiles but is significant in $Q_{0.25}$ with positive value. It is an interesting finding since the low, middle, and highly efficient schools behave similarly, but the lower-middle efficient schools see this ratio as a significant factor that influence school's efficiency. On the other hand, the ratio of computers connected to the internet is found to be significant in all cases and all quantiles.

To distinguish if the conditional distribution of efficiency is quantile dependent, the inter-quantile tests based on the pair *t*-test is employed. The variance-covariance matrixes of the corresponding parameters are estimated from the bootstrap procedure with 500 replications. Table IV-8 shows tests of all quantile pairs for each determinant of efficiency only for Case 4. The null hypothesis about the equality of coefficients for different quantiles of school size is rejected. It means that the effect of school size is statistically similar in all schools, regardless of their efficiency condition. The other determinants behave differently depending on their quantiles. For example, the effect of WEBCOMP equals in all pairs but does not in the lower quantiles (i.e., $Q_{0.1}$ vs $Q_{0.25}$) even though the estimated coefficients look very similar in these two quantiles (see again Table IV-7). If one observes further, in this pair, all the effects of the determinants but WEBCOMP are statistically equal, meaning that only the ratio of compu-

Table IV-8 Results of pair *t*-tests for inter-quantile parameter differences in Case 4 (all domains)

Variables	$Q_{0.1}$ vs $Q_{0.9}$	$Q_{0.25}$ vs $Q_{0.9}$	$Q_{0.50}$ vs $Q_{0.9}$	$Q_{0.75}$ vs $Q_{0.9}$	$Q_{0.1}$ vs $Q_{0.75}$
Constant	5.88**	5.96**	4.50**	3.57**	4.95**
PROP_GIRL	-1.54	-2.04**	-1.95*	-1.67*	-0.54
SCHSIZE	0.20	0.44	0.20	-0.03	0.34
PROATCE	-2.53**	-2.69**	-1.74*	-2.15**	-1.16
EDUSHORT	-2.53**	-2.38**	-1.02	-0.32	-2.25**
STAFFSHORT	1.66*	1.80*	1.46	0.28	1.40
COMPRATIO	-1.13	-2.09**	-2.12**	-0.86	-0.76
WEBCOMP	-0.17	-0.84	-0.70	-1.14	1.36
Variables	$Q_{0.25}$ vs $Q_{0.75}$	$Q_{0.50}$ vs $Q_{0.75}$	$Q_{0.1}$ vs $Q_{0.5}$	$Q_{0.25}$ vs $Q_{0.5}$	$Q_{0.1}$ vs $Q_{0.25}$
Constant	4.83**	3.05**	3.76**	3.68**	0.71
PROP_GIRL	-0.79	-0.93	0.07	0.01	0.08
SCHSIZE	0.74	0.45	0.01	0.52	-0.54
PROATCE	-1.43	0.21	-1.57	-1.88*	0.20
EDUSHORT	-2.25**	-0.92	-1.87*	-2.12**	0.04
STAFFSHORT	1.58	1.38	0.42	0.73	-0.20
COMPRATIO	-2.08**	-2.09**	0.49	-0.18	0.67
WEBCOMP	0.26	0.65	1.13	-0.41	1.92*

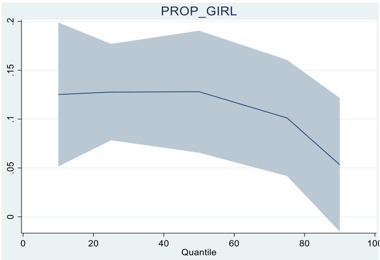
Notes: the numbers denote the *t*-value.

*significant at the level of 10%
**significant at the level of 5%

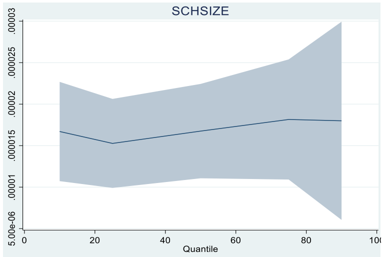
ters connected to the internet differs schools in these quantiles. It is of interest to see that the effect of this determinant statistically equals in all other pairs.

To visually represent the change in the estimated coefficients, Figure IV-4 shows a visual appreciation of the bootstrapped quantile regression results only for the fourth case. The blue line denotes the bootstrapped quantile regression estimates, while the grey interval is the 95% confidence interval. According to Baum (2013), the graph illustrates how the effects of each determinant vary over quantiles. One can view that the influence of PROP_GIRL decreases as the quantile increases. The *stable* effect is found in the school size while the *increasing* effect is found in the index of educational staff shortage.

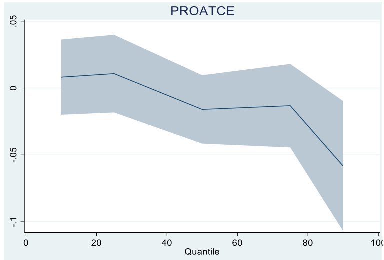
The final analysis is to give further evidence of the heterogeneity across the countries. Table IV-9 reports parameters estimation in smaller samples (i.e., by country) only for Case 4 and in particular the upper quantile $Q_{0.9}$ (to show how the most-efficient schools behave in each country). Note that the issue of a small sample can be easily handled by using the bootstrap procedure (Simar and Wilson, 2007). Looking at the statistics, some similarities and differences emerge. The proportion of female students have no influence to affect the most-efficient schools in all countries. School size only has influence on school's efficiency in Indonesia and Singapore with very small magnitude. The proportion of certified teachers is significant with positive value in the super-efficient schools in the Philippines but with negative value in Indonesia and Malaysia. In terms of ICT infrastructure, the ratio of computers to the total number of students is found to be statistically significant only in Indonesia. Next, the share of the computers connected to the internet is statistically significant in three countries, i.e., Indonesia, Malaysia, and the Philippines.



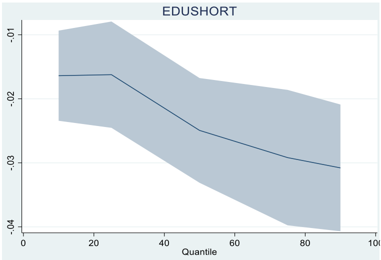
(a) PROP_GIRL



(b) SCHSIZE

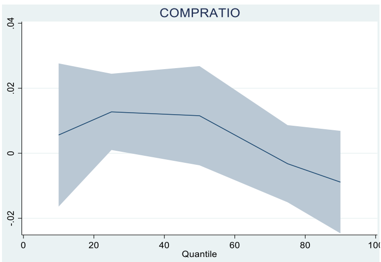


(c) PROATCE

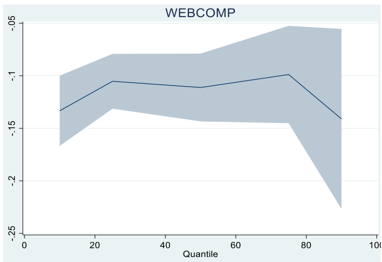


(d) EDUSHORT

(e) STAFFSHORT



(f) COMPRATIO



(g) WEBCOMP

Figure IV-4 The influence of determinants of efficiency in Case 4 (all domains)

Table IV-9 Parameters estimation of the bootstrapped quantile regression for Case 4 and $Q_{0.9}$, by country

Variables	BRN (N = 53)	IDN (N = 238)	MYS (N = 172)	PHL (N = 135)	SGP (N = 157)	THA (N = 252)
Constant	0.718	1.187**	1.246**	0.631**	0.972**	0.875**
PROP_GIRL	0.221	0.030	0.058	0.249	0.073	0.071
SCHSIZE	0.00007	0.00009**	0.00002	0.00001	0.0001**	0.0000009
PROATCE	-0.044	-0.087**	-0.267**	0.217**	-0.048	0.055
EDUSHORT	-0.0096	-0.014	-0.033	-0.015	-0.0002	-0.036**
STAFFSHORT	-0.127	0.005	-0.043*	0.020	0.006	0.036**
COMPRATIO	0.027	-0.109**	-0.015	-0.0396	-0.026	-0.003
WEBCOMP	-0.089	-0.307**	-0.180**	-0.085*	-0.157	-0.100

*significant at the level of 10%
**significant at the level of 5%

IV.5 Discussion and Concluding Remarks

This study analyzes the efficiency of schools in six South-East Asia countries participated in the recent OECD PISA 2018 using two-stage super-efficiency model. In the first stage, the efficiency is estimated using the radial and non-radial super-efficiency model based on DEA, while at the second stage, the bootstrapped quantile regression is applied to investigate the influence of the determinants of efficiency. To date, this might be the first attempt of such an efficiency analysis in education that employs two-stage super-efficiency model using the bootstrapped quantile regression in the second stage.

To observe different behaviors that might take place, four cases are generated. The first case uses the PISA score of mathematics as input, while the second, third, and fourth case use the PISA score of science, reading, and all domains respectively. The inputs are the same for all cases, i.e., the (inverse) of student-teacher ratio, the index of school's economics, social, and cultural status, the ratio of computers to the total number of students, and the ratio of computers connected to the internet. Confirming study in Chapter III, the results in this chapter reveal that Singapore has the (relatively) best performance among the other countries in all cases and all two DEA models.

The factors that might influence school's efficiency are examined in the second stage by the aid of bootstrapped quantile regression using quantiles of 10%, 25%, 50% (median), 75%, and 90%. Seven determinants are included, i.e., PROP_GIRL, SCHSIZE, PROATCE, EDUSHORT, STAFFSHORT, COMPRATIO, and WEBCOMP. Notice that among those determinants, only school size behaves similarly in all quantiles and all cases (i.e., it positively influences efficiency with a very small value). The proportion of female students at school significantly influences efficiency in all quantiles only in Case 3, while the ratio of computers connected to the internet does not affect school's efficiency in all quantiles only in Case 3. Other determinants behave differently depending on the quantiles, indicating that the conditional distribution of efficiency is quantile dependent.

The results presented here suggest a number of policy implications for South-East Asian schools, indicating different courses of action for schools with higher and lower efficiency levels. Lower efficiency schools clearly benefit more from the number of female students in a school than higher efficiency schools. This confirms the finding of Agasisti and Zoido (2019), who found that this determinant was positively correlated with the school's efficiency in developing countries; and finding of Santín and Sicilia (2018) who found that this variable was not statistically significant to influence efficiency of schools in Spain, a developed country. The proportion of certified teachers, on the other hand, belongs to the factor that does not affect the lower efficiency schools but affects the higher efficiency schools. Similar condition also happens in STAFFSHORT, meaning that the quality of human resources plays an important role in determining efficiency at higher efficiency schools.

Lastly, this study also provides a shred of evidence that heterogeneity exists across countries (see Table IV-9). Some differences that take place could reflect factors that are beyond the managerial efficiency of schools and could be related to welfare regimes or other institutional factors that vary across countries in South-East Asia which have different traditions and settings. This finding can be a basis for the future studies which consider more explicitly the role of supranational elements that might shape the educational provision and school's productivity (Agasisti and Zoido, 2019).

CHAPTER V. THE INFLUENCE OF ICT ON EDUCATION OUTOCOMES AND INEFFICIENCY: THE "FOUR-COMPONENT STOCHASTIC FRONTIER MODEL" APPROACH

This chapter provides a panel data analysis to investigate both the influence of ICT on education outcomes and inefficiency. There is limited literature which investigates simultaneously the impact (or influence) of ICT on education outcomes and inefficiency. Previous studies only dealt with one issue separately, either the influence of ICT on education outcomes or on inefficiency.

The impact of ICT on education outcomes has become a controversial issue as there are both positive and negative arguments about the effectiveness of ICT for teaching and learning. On the one hand, the contribution of ICT to the improvement of teaching and learning processes is higher in the schools that have integrated ICT (Sangrà and González-Sanmamed, 2010). Spiezia (2010) found a positive and significant effect of the frequency of computer use on student performances as measured in the PISA 2006 data of science score. By applying a three-level hierarchical linear model to the PISA data from 2000 to 2012 waves, Zhang and Liu (2016) revealed that school-level ICT-related variables had positive influences on learning outcomes when national GDP, school type, and school ICT investment, were controlled for. Srijamdee and Pholphirul (2020), who investigated the impacts of ICT familiarity on education outcomes in developing countries, found that using ICT for educational proposes can help improve Thai students' PISA scores. As evidence from a developed country, Fernández-Gutiérrez et al. (2020) showed an increase in the use of ICT at school in Autonomous Communities in Spain did have positive influence on PISA scores in science. Winkler et al. (2021) showed that the use of smart personal assistants (such as Amazon's Alexa or Google's Assistant) had a positive effect on skill development, more precisely on the development of problem-solving skills.

On the other hand, many scholars argued that the return of using ICT in education is not significantly positive in terms of increasing education outcomes (see, e.g., Angrist and Lavy, 2002; Mora et al., 2018). ICT may also discourage students' effort and logical thinking if educational systems do not fit technology to their ICT-instructional needs (Wheeler et al., 2002). ICT might further provide a lot of wrong or incomplete information which can diminish students' learning (Gómez-Fernández and Mediavilla, 2020). Therefore, the effectiveness of ICT on educational systems will depend on the net effect of these likely positive and negative contributions to education outcomes (Gimenez and Vargas-Montoya, 2021).

In terms of efficiency, there are two opposite sets of observations in the literature about the influence of ICT on the efficiency in education (De Witte and Rogge, 2014). On the one hand, some scholars found that ICT could reduce educational costs. Other advantages are improving the delivery of education and the learning process, the presence of greater flexibility and autonomy for the students' learning, as well as supporting more interaction and a reduction in the teachers' workload (Grimes and Warschauer, 2008; Lei and Zhao, 2008; Venable et al., 2011). On the other hand, when ICT is not well integrated in the curriculum, due to pedagogical barriers, it might hinder students from learning (Fu, 2013). There are also some barriers that obstruct the use of ICT in education from the teacher perspective, such as a lack of teacher collaboration and pedagogical support, a lack of in-service training on the use of ICT, insufficient time to master new educational software or to integrate ICT during a class period, limited knowledge and experience of ICT in teaching contexts, as well as several technical problems related to ICT in the classroom that frequently happened.

Taking all this into account, this chapter investigates the influence of ICT on both education outcomes and inefficiency by using OECD PISA data from 2009 to 2018 waves of 24 OECD countries. The "four-component stochastic frontier model" is used to accomplish the objective. This model disentangles overall inefficiency into two parts: persistent and time-varying inefficiency. The persistent inefficiency refers to a long-term or structural inability of an education institution to achieve the potential level of academic outputs. Time-varying inefficiency, on the other hand, is a short-run deficit which can be eliminated swiftly without a major structural change. Distinguishing between persistent and time-varying inefficiency is important since they may have different policy implications (Lai and Kumbhakar, 2018). In this specification, the ICT-related variables are treated as inputs that might influence the education outcomes and determinants of time-varying inefficiency.

This chapter is structured as follows. In Chapter V.1, by reviewing previous study using literature review, the contributions of this study are presented. In Chapter V.2, the empirical model is displayed. Data and a description of the variables used study are presented in Chapter V.3. Results and robustness analysis are presented in Chapter V.4. Finally, Chapter V.5 provides discussion and concluding remarks.

V.1 Literature Review and Contributions

According to the critical discussions provided in Chapter II.7, it is recommended to conduct this research using the "four-component model" of SFA. This model disentangles overall inefficiency into persistent and time-varying inefficiency. Distinguishing between persistent and time-varying inefficiency is important since they may have different policy implications (Lai and Kumbhakar, 2018). To investigate previous studies that applied this "four-component model" in the education sector, the literature review is performed in the Scopus database. The following search query is used: TITLE-ABS-KEY((education OR universit* OR higher education OR school) AND (SFA OR "stochastic frontier analysis") AND (persistent OR time*) AND efficien*). The article type is restricted to peer-reviewed research article published in a journal and written in English. Only articles published in 2014 and afterwards are considered since this model was proposed in 2014. The search yields 21 articles. The first screening is performed by reading the title and abstract to verify the relevance of the extracted articles. In this way, 9 articles are excluded since the studies are not conducted in the education sector and are not related to the efficiency measurement. The second screening is executed by carefully reading the full text of each article to address the eligibility of the articles. For a practical reason, articles whose full text cannot be accessed are also excluded. Only 5 articles extracted in this second screening procedure. Lastly, the manual forward and backward chaining of the extracted articles are performed to minimize the risk of missing articles. This procedure results in the addition of only 1 article. Finally, there are six articles that satisfied the inclusion criteria as briefly described in the following.

Titus et al. (2017) examined cost efficiency at 252 public master's institutions in the United States over a nine-year (2004–2012) period. They employed a slightly modified version of the "four-component stochastic frontier model" of Kumbhakar et al. (2014) by taking into account spatial interdependency to decompose cost efficiency into long-term stable (persistent) and short-term (time-varying) efficiency. Gralka (2018) investigated cost efficiency at 73 German public universities covering the years from 2001 to 2013. Using the translog cost function, the author exposed that persistent efficiency was lower than transient efficiency and was the main cause of the overall efficiency potential of the institutions, indicating operational problems and showing that a substantial increase in the efficiency level can be generated only through a comprehensive change in policy. Agasisti and Gralka (2019) compared efficiency of 55 Italian and 70 German public universities covering the years from 2001 to 2011. They tried to answer the specific research question: are efficiency differences between countries more related to the individual performance of the universities or instead to the higher education system's structure? To answer the question, they used the "four-component stochastic frontier model" that distinguishes between persistent and time-varying inefficiency, while controlling for institution-specific heterogeneity. The result showed that the two countries exhibit a high and similar time-varying institutional efficiency. Instead, the remaining inefficiency and the gap between the two higher education sectors was driven by long-term structural inefficiency. The country-specific characteristics seemed to influence the universities to a strong and disadvantageous extent.

Salas-Velasco (2020) evaluated the performance of Spanish secondary schools using PISA data of 2003 and 2012. The result showed that schools are moderately inefficient, and that inefficiency was presumably not caused by something unexpected within each year such as greater difficulty in hiring teachers, but rather by persistent factors such as classroom management: schools with better classroom disciplinary climate tend to be less inefficient in educational production. Titus (2020) examined the financial context of bachelor's degree production efficiency among public master's colleges and universities in the United States. The results show that bachelor's degree production is positively and non-linearly related to doctoral degree production; while persistent efficiency is positively related to tuition revenue, state appropriations, and Pell grant revenue (a federal need-based grants to undergraduate students named after former US Senator Pell) and negatively related to federal grant and contract revenue. Badunenko et el. (2021) evaluated the inefficiency of adult education programs in Belgium. The results indicated that the overall inefficiency amounts to 12%, suggesting that, given the available resources, the outputs (measured by exam scores, class attendance rates, and exam participation) could increase by 12%. Decomposing the overall inefficiency reveals that about 5 percentage points of the inefficiency are on average due to structural differences between the programs, whereas about 7 percentage points are at the discretion of the adult education management.

From this literature review, only one article which incorporated ICT into the model, i.e., Salas-Velasco (2020) who used the ratio of computers at school that connected to the internet as input. It is apparent that more research should be conducted to explore more the influence of ICT in measuring efficiency in education, especially by employing the "four-component stochastic frontier model".

Recall from the systematic literature review in Chapter II, according to the level of analysis, there is limited study conducted at cross-country that incorporated ICT in the measurement of efficiency in education. Previous studies only performed their analysis inside their countries, such as in China (Chen et al., 2020), Spain (Crespo-Cebada et al., 2014; Ferrera et al., 2011; Perelman and Santín, 2011a, b; Salas-Velasco, 2020), Brazil (Alves and de Araújo, 2018; Zoghbi et al., 2013), Mexico (Garcia-Diaz et al., 2016), and the United Kingdom (Johnes, 2013). From the literature review conducted in this chapter, only the study of Agasisti and Gralka (2019) which was conducted to compare the education institutions in Germany and Italy allowed for international comparisons. However, the study did not incorporate ICT into the model. Combining all these together, this study tries to fill this gap by using the OECD PISA data, which is well-regarded as an authoritative source of comparison for educational achievement across the world. This PISA data contains several ICT-related data that can be exploited.

The contributions of this study are then as follows. This study extends the application of the "four-component stochastic frontier model" in the education sector. There is limited literature which applied this model in the education sector. Recall in the critical discussions provided in Chapter II.7, majority of the studies incorporating ICT only discussed the physical infrastructure, e.g., number of computers at school, internet availability, resources available at home. Therefore, this study provides a comprehensive analysis on the role of ICT in efficiency of education by including three under-studied ICT-related variables, categorized as ICT use, i.e., the index of time spent by student in using ICT (i) at school, (ii) outside school for entertainment purposes, and (iii) at home for school-related tasks.

V.2 Method

To investigate the influence of ICT on both education outcomes and inefficiency, the "four-component stochastic frontier model" is applied. One of the drawbacks of the "standard" SFA applied to the panel data is that the producer effects (fixed or random) are assumed to be parts of inefficiency.

The model that separates producer effects from inefficiency exploits the panel structure of the data better, since it can control for unobserved heterogeneity. Apart from that, the standard SFA applied in the panel data also does not separate inefficiency into persistent and time-varying term. The persistent inefficiency refers to a long-term or structural inability of the producer to achieve the potential level of output. Since the persistent inefficiency is time-invariant it can only be changed in the long-run through some restructuration. Time-varying inefficiency, on the other hand, is a short-run deficit which can be eliminated swiftly without a major structural change; therefore, it can be manipulated more quickly than persistent inefficiency.

The "four-component stochastic frontier model" is specified as follows:

$$y_{it} = \alpha_0 + f(\mathbf{x}_{it};\boldsymbol{\beta}) + \tau_i - \eta_i + v_{it} - u_{it}, \tag{V-1}$$

where y_i is the output, α_0 is an intercept, $f(\mathbf{x}_{it};\boldsymbol{\beta})$ is the production technology; $\mathbf{x}_{it}$ is a vector of inputs; $\boldsymbol{\beta}$ is the associated vector of parameters to be estimated; τ_i is a random producer effects that capture unobserved heterogeneity, η_i is non-negative persistent inefficiency, u_{it} is non-negative time-varying inefficiency, and v_{it} is two-sided statistical noise. The subscript i (i = 1, 2, …, N) and t (t = 1, 2, …, T) refer to school—as this study is conducted at the school level—and time, respectively.

Scholars have proposed several methods to estimate Equation (V-1). Colombi et al. (2014) who derived a closed form expression of the likelihood function of the composed error term $\varepsilon_{it} = \tau_i - \eta_i + v_{it} - u_{it}$, proposed a single-step classical maximum likelihood method based on distributional assumptions on those four components. They assumed that each component is distributed independently and identically and are also independent of each other. Filippini and Greene (2016) suggested using the maximum simulated likelihood approach, while Tsionas and Kumbhakar (2014) proposed a Bayesian solution in iterated two-steps. This study used a multi-step procedure proposed by Kumbhakar et al. (2014) which is easier and simpler to follow and implement.

Equation (V-1) can be rewriten as:

$$y_{it} = A + f(\mathbf{x}_{it};\boldsymbol{\beta}) + B_i + \Gamma_{it}, \text{ where} \tag{V-2}$$

$$A = \alpha_0 - E(\eta_i) - E(u_{it}), \tag{V-3}$$

$$B_i = \tau_i - \eta_i + E(\eta_i), \text{ and} \tag{V-4}$$

$$\Gamma_{it} = v_{it} - u_{it} + E(u_{it}). \tag{V-5}$$

With this specification, B_i and Γ_{it} will have zero mean and constant variance. This model then can be estimated in three steps as follows.

Since Equation (V-2) resembles the standard panel regression model; thus, in Step 1, the standard panel regression is used to estimate $\boldsymbol{\beta}$. This procedure also gives predicted values of B_i and Γ_{it}, denoted by $\hat{B}_i$ and $\hat{\Gamma}_{it}$. In Step 2, by assuming τ_i to follow normal distribution, η_i to follow half-normal distribution (which means $E[\eta_i]$ = $\sqrt{2/\pi}\sigma_\eta$), and ignoring the difference between the true and predicted values of B_i, which is the standard in multi-step procedure, Equation (V-4) can be estimated using the standard SFA applied in the cross-sectional data (see Chapter III.2). The procedure will give predicted value of persistent inefficiency, denoted by $\hat{\eta}_i$ (using the JLMS estimator). The persistent efficiency (PE) can be estimated by using the BC estimator. In Step 3, by assuming v_i to follow normal distribution, u_{it} to follow half-normal distribution ($E[u_{it}]$ = $\sqrt{2/\pi}\sigma_\eta$), and ignoring the difference between the true and predicted values of Γ_{it}, Equation (V-5) can be estimated also by using the standard SFA applied in the cross-sectional data. Using the JLMS estimator, time-varying inefficiency can be estimated, denoted by $\hat{u}_{it}$. Time-varying efficiency (TE) can be estimated using the BC estimator. The overall efficiency (OE) is then obtained from the product of PE and TE as:

$$\text{OE} = \text{PE} \times \text{TE}. \qquad \text{(V-6)}$$

The previous model assumes that u_i and v_i are homoscedastic, that is, both σ_u^2 and σ_v^2 are constants. This is not capable to investigate the influence of determinants of inefficiency, i.e., factors that can explain inefficiency. To model the determinants of inefficiency, see again Chapter III.2.

V.3 Data

To examine the influence of ICT on education outcomes and inefficiency, four waves of PISA data from 2009 to 2018 are used. This dataset provides international comparative data on the performance of 15-year-old students in different competencies (i.e., mathematics, sciences, and language). PISA also provides information that is potentially related to the assessment result, such as variables representing student background, school environment, or educational provision. This information comes from the responses given to different questionnaires completed by students, school principals, or parents. Student, school, and ICT familiarity questionnaire are included as those are the most relevant for this study. Only OECD countries always participated in all editions of PISA from 2009 to 2018 waves that offered student, school, and ICT familiarity questionnaires are included. It yields 24 countries, i.e., Australia (AUS), Austria (AUT), Belgium (BEL), Switzerland (CHE), Chile (CHL), Czech Republic (CZE), Germany (DEU), Denmark (DNK), Estonia (EST), Finland (FIN), Greece (GRC), Hungary (HUN), Ireland (IRL), Iceland (ISL), Israel (ISR), Italy (ITA), Japan (JPN), Republic of Korea (KOR), Latvia (LVA), New Zealand (NZL), Poland (POL), Slovak Republic (SVK), Slovenia (SVN), and Sweden (SWE).

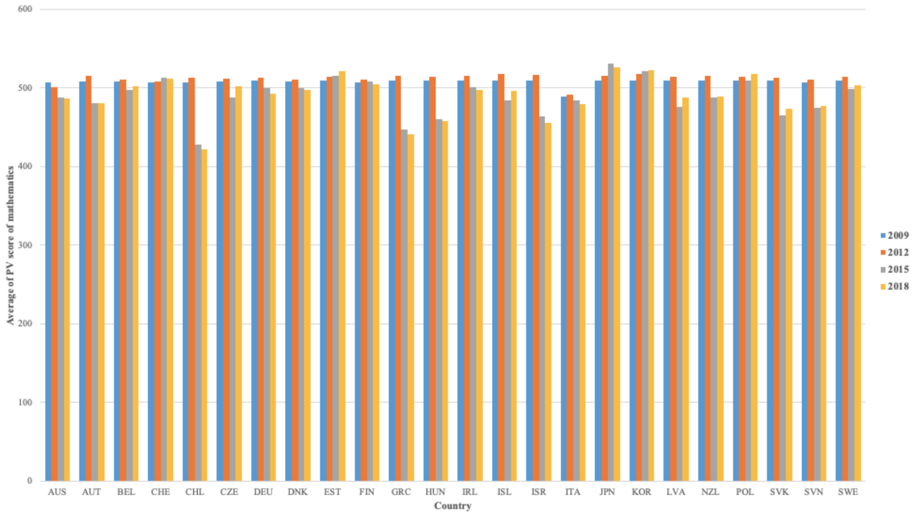
Variables included in this study must appear in all considered waves and are comparable to the previous or subsequent PISA waves. This constraint may reduce the potential explanatory power of the empirical model, though it allows improving the reliability of the panel modeling. The study is conducted at the country level. However, the data is at the school level; therefore, variables that are at the student-level (i.e., from student and ICT familiarity questionnaire) are weighted to obtain school-level variables by using the appropriate weight provided by PISA (in this study, W_FSTUWT, the final student weight provided by PISA is used).

Three model specifications are developed in this study, each with different dependent variables but identical independent variables. The dependent variables reflect the student's proficiency in mathematics (Model 1), science (Model 2), and reading (Model 3), proxied by the weighted plausible values (PVs) provided by PISA—later in this study, it is called the PISA scores.

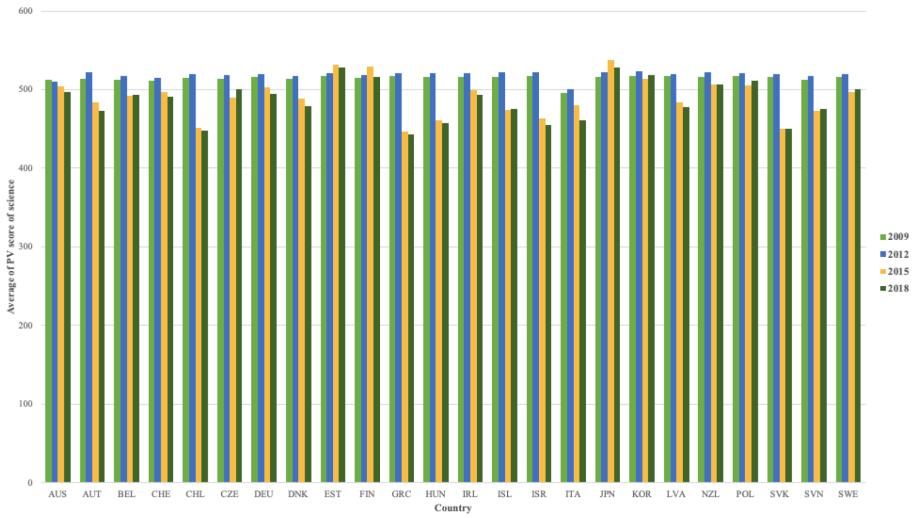
The means of each test score for each country per wave are shown in Figure V-1. Japan has the highest score in mathematics and science (in 2015 and 2018 waves), while Republic of Korea has the highest score in language (in 2009 and 2012 waves). On the other hand, Italy has the lowest score in all domains in 2009 and 2012 waves. Hungary has the lowest standard deviation in mathematics and language, while New Zealand has the lowest score in science. It can be an indication that the education system in those countries could give a more uniform education outcomes measured by the PISA scores; the opposite condition happened in Chile (for mathematics) and Israel (for science and reading).

Readers might notice a relatively large shift of means of the PISA score in 2012 and 2015 for several countries. In PISA 2015 wave onwards, the assessments were administered by using computer-based test in 57 of 72 participating countries—all 24 investigated countries in this study used computer-based test, while in the previous PISA waves, the assessments were administered by using paper-based test. OECD took a great care to ensure that performance would not be significantly affected by the shift from a paper- to a computer-based test. For instance, when developing a fully equivalent computer version for a paper-based task proved challenging because of interface issues, such as students' unfamiliarity with equation editors or drawing tools on computers (OECD, 2016).

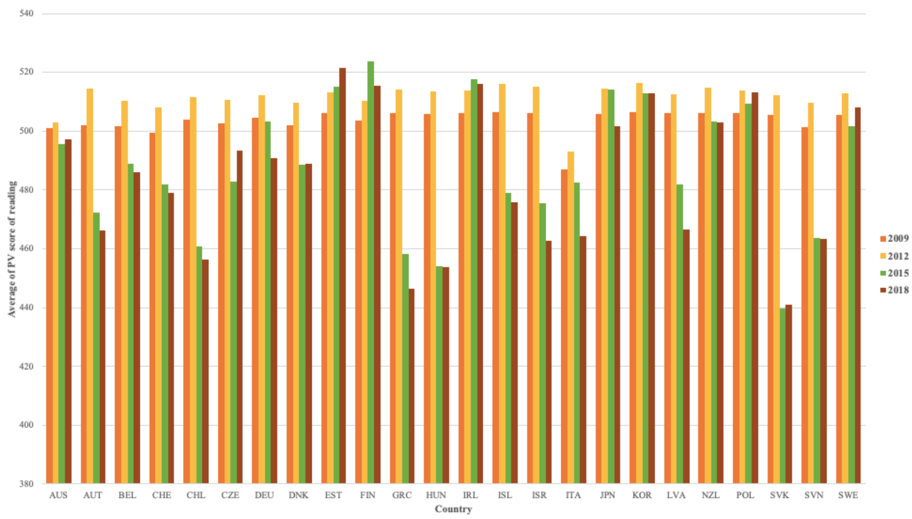
Correlation analyses performed by OECD describe that changes in the mode of delivery are, at best, only a partial explanation for changes in performance between PISA 2012 and PISA 2015, that are observed in countries that conducted PISA 2012 assessment on paper and PISA 2015 assessment on computer. Figure V-2 (a) shows the relationship between a simple indicator of familiarity with ICT (i.e., the share of students in PISA 2012 that are having three or more computers in their homes) with score-point difference between PISA 2012 and PISA 2015 edition. Across all countries, greater exposure to ICT devices in the home explains, at best, only 4% of the variation in the difference between PISA 2012 and 2015 scores (the correlation coefficient is only 0.21). After excluding two countries that show both greater exposure and significant and positive trends (i.e., Denmark and Norway), the correlation between these two measures is only 0.10 across the remaining countries. This means that in Denmark and Norway, students' greater familiarity with ICT (or, perhaps, greater motivation to take a test delivered



(a) Mathematics test score

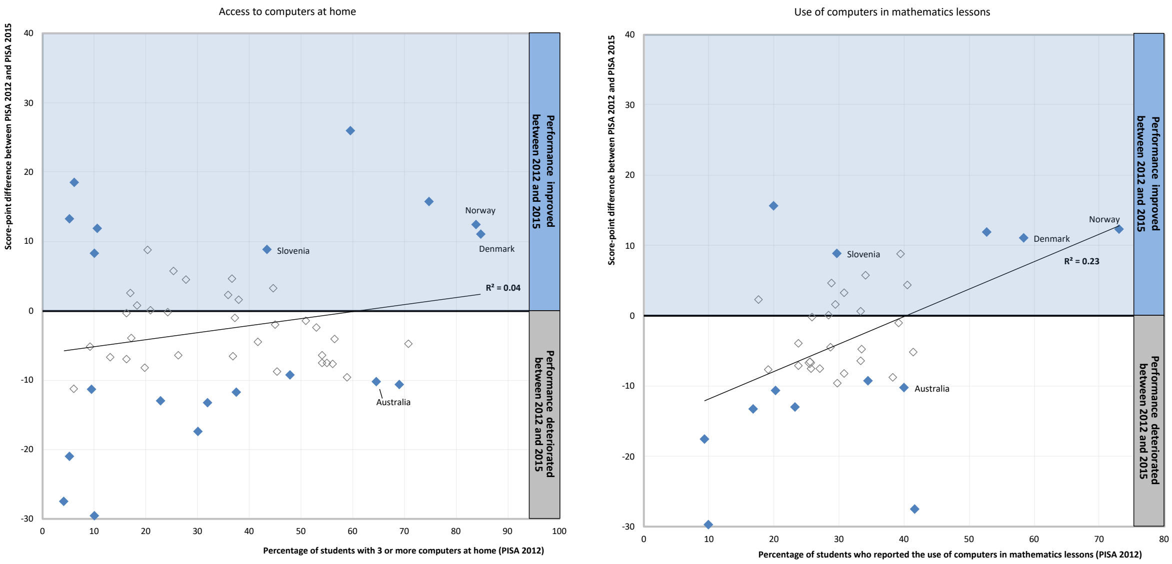


(b) Science test score



(c) Reading test score

Figure V-1 Means of PISA score for each domain per country per wave



(a) Percentage of students with three or more computers at home in PISA 2012 vs. score-point difference between PISA 2012 and PISA 2015

(b) Percentage of students who reported the use of computers in mathematics lessons in PISA 2012 vs. score-point difference between PISA 2012 and PISA 2015

Figure V-2 Relationship between students' familiarity with ICT and change in PISA score

Source: OECD (2016)

on computer rather than one delivered on paper) could be part of the observed improvement (or deterioration) in performance. But in general, countries where students have greater familiarity with ICT tools are almost equally likely to observe positive and negative trends, as are countries where students have less familiarity with ICT. Changes are even less correlated with other indicators of access to computers at home. The correlation coefficient is only 0.17 with the share of students in PISA 2012 who reported having two or more computers at home; and only 0.05 with the share of students in PISA 2012 who reported having one or more computer at home.

For 38 countries, a more specific indicator of familiarity with ICT was used. In PISA 2012, students were asked to report whether they use computers during mathematics lessons for specific tasks, such as drawing the graph of a function or calculating with numbers. The share of students who reported doing at least one of these tasks on computer during mathematics lessons in the month prior to the PISA 2012 assessment correlates positively with the difference in mathematics performance between PISA 2012 and PISA 2015 in these 38 countries (the correlation coefficient is 0.48). But clearly, not all changes in performance can be explained by the use of ICT tools in mathematics lessons ($R^2 = 0.23$), see Figure V-2 (b). For instance, in Australia, a negative trend in performance between PISA 2012 and PISA 2015 was observed despite the fact that students in PISA 2012 assessment reported frequent use of ICT tools in mathematics lessons. Moreover, the correlation of changes in the mean mathematics performance between PISA 2012 and PISA 2015 with differences between the computer-based and the paper-based mathematics performance in 2012 is only 0.18, signaling a weak association.

In sum, it is inconclusive to state that the change in mathematics performance was due to ICT familiarity as a proxy for the changes in the delivery mode (from paper-based test to computer-based test); instead, it may imply that the aspects that are unique to the PISA 2012 computer-based assessment (e.g., the inclusion of items that explicitly measure students' ability to use ICT tools for solving mathematics problems and when the test was conducted) explain a bigger part of the performance differences in PISA 2012 than how the test was delivered. It may also imply that changes in the performance between PISA 2012 and PISA 2015 largely reflect other factors than the mode of delivery, such as changes in student proficiency, or the sampling variability and scaling changes that contribute to the uncertainty associated with trend estimates.

Next, the independent variables are classified in three categories, which reflect the main groups of variables: (a) student's characteristics, i.e., student's age (AGE), index of economic, social, and cultural status (ESCS); (b) school's characteristics, i.e., learning time per week in mathematics (MMINS), science (SMINS), or language (LMINS), school size (SCHSIZE), school type (SCHTYPE), school location (SCHLOC), proportion of female students in a school (PROP_girl), proportion of native students in a school (PROP_nat), student-teacher ratio (STRATIO), proportion of fully certified teacher (PROATCE), student behavior hindering learning (STUBEHA), teacher behavior hindering learning (TEACHBEHA), index of school educational resources (EDUSHORT), index of school staff/teacher resources (STAFFSHORT); (c) ICT-related variables, i.e., the ratio of computers to the total number of students for educational purposes (COMPRATIO), the ratio of computers to the number of these computers that were connected to the internet (WEBCOMP), index of time spent by students in using ICT: outside school for entertainment purposes (ENTUSE), at school (USESCH), and at home for school-related tasks (HOMSCH). List of the independent variables and its description are shown in Table V-1. Two variables are categorized as dummy variables (i.e., SCHTYPE and SCHLOC), while seven variables are derived based on item response theory (IRT) scaling (i.e., STUBEHA, TEACHBEHA, EDUSHORT, STAFFSHORT, ENTUSE, USESCH, and HOMSCH). Note that all ICT-related variables are included as determinants of inefficiency. The selection of inputs has been guided by the existent literature in the field of education economics. Some references to be the justification of input selection are presented in Table V-1.

Several variables have been already discussed in Chapter III; in the following, only variables that are not yet discussed are presented. The first is student's age, which is related to student's maturity. It is argued that age has an impact on educational achievement. This variable has been used in e.g., Dolton et al. (2003). The next is variables derived from the IRT scaling. STUBEHA reflects the school principal's perceptions of student behavior that might influence the provision of instruction at school;

Table V-1 Description of inputs

Variables	Descriptions	References
AGE	Student's age	Dolton et al. (2003); Zoghbi et al. (2013)
ESCS	Index of economic, social, and cultural status	Crespo-Cebada et al. (2014); Ferrera et al. (2011); Perelman and Santín (2011a); Salas-Velasco (2020)
M(S/L)MINS	Mathematics (science/language) learning time per week (in minutes)	Dolton et al. (2003)
SCHSIZE	School size or number of enrolled students	Barnett et al. (2002); Bradley and Taylor (1998); Hanushek and Luque (2003); Mora et al. (2010)
SCHTYPE	Type of school: public, private government-independent, and private government-dependent	Crespo-Cebada et al. (2014); Garcia-Diaz et al. (2016); Kirjavainen (2012)
SCHLOC	School location: located in rural area, in a small town, in a town, in a city, and close to the center of a city with over a million people or elsewhere in a city with over a million people	Kirjavainen (2012); Perelman and Santín (2011a)
PROP_girl	Proportion of female students in a school	Crespo-Cebada et al. (2014); Kirjavainen (2012); Mongan et al. (2011); Perelman and Santín (2011a); Zoghbi et al. (2013)
PROP_nat	Proportion of native students in a school (i.e., students who had at least one parent born in the country	Crespo-Cebada et al. (2014); Perelman and Santín (2011a); Zoghbi et al. (2013)
STRATIO	Student-teacher ratio	Agasisti (2014); Agasisti et al. (2019); Agasisti and Zoido (2019)
PROATCE	Proportion of fully certified teacher	André et al. (2020); Grosskopf et al. (2014)
STUBEHA (IRT)	Student behavior hindering learning	Perelman and Santín (2011a)
TEACHBEHA (IRT)	Teacher behavior hindering learning	Ulkhaq (2021)
EDUSHORT (IRT)	Index of educational material shortage	Crespo-Cebada et al. (2014); Ferrera et al. (2011); Perelman and Santín (2011a); Salas-Velasco (2020)
STAFFSHORT (IRT)	Index of educational staff shortage	Courtney et al. (2022); Lima (2017); Shahini (2021)
COMPRATIO*	Ratio of computers to the total number of students for educational purposes	Perelman & Santín (2011b); Zoghbi et al. (2013)
WEBCOMP*	Ratio of computers at school to the number of these computers that were connected to the internet	Salas-Velasco (2020)
ENTUSE (IRT)*	Index of time spent by students in using ICT outside school for entertainment purposes	
USESCH (IRT)*	Index of time spent by students in using ICT at school	
HOMSCH (IRT)*	Index of time spent by students in using ICT at home for school-related tasks	

*also serve as determinants of inefficiency while TEACHBEHA reflects the school principal's perceptions of teacher behavior that might influence the provision of instruction at school. ENTUSE, USESCH, and HOMSCH belong to the ICT familiarity questionnaire, in which asked students how often digital devices are used outside of school for leisure activities, outside of school for school-work, and for activities in school, respectively. The response for all those questions ranged from "never or hardly ever", "once or twice a month", "once or twice a week", "almost every day", to "every day".

To investigate the influence of ICT on inefficiency, all ICT-related variables are taken into account as determinants of time-varying inefficiency, i.e., ICT infrastructure, including COMPRATIO and WEBCOMP, as well as ICT use, including ENTUSE, USESCH, and HOMSCH. Recall that time-varying inefficiency is defined as a short-run deficit which can be eliminated swiftly without a major structural change. Therefore, it is a common and easier practice for policy makers as an endeavor to reduce this type of inefficiency. One of the common and general trends is the provision of ICT infrastructure to the schools; for instance, as in the case of Spain (the 2.0 School Program), Hungary (The Digital School Plan), Italy (The Intelligent School Program), and Turkey (the FATIH Project). This provision was also extended to the homes of students from low-income families, as in the United Kingdom and Singapore (the Home Access Program) (Gil-Flores et al., 2017). Accordingly, there was a widespread increase in the computer-student ratio between 2009 and 2018; specifically, on average across OECD countries, there was one additional computer available per every four students in 2018 than was available in 2009 (0.26 of an additional computer per student) (OECD, 2019b).

However, the infrastructure policy is not enough since having access to the digital devices does not automatically translate into high rates of use (European Commission, 2013); hence, this policy has to be accompanied by the use of this infrastructure. Perbawaningsih (2013) provided a discussion about how ICT use in higher education could have some bearing on efficiency, measured by saving cost, time, and effort. In this study, the frequency of ICT use by students are measured by three variables. Three under-studied variables are proposed, i.e., USESCH, HOMSCH, and ENTUSE. By including these variables, it is attempted to both extend the literature and provide a more holistic view of the role of ICT in measuring efficiency in education as the previous studies only addressed the ICT infrastructure.

The means (only for numerical data) are shown in Table V-2; while for the ICT-related variables we also show the descriptive statistics, including mean, standard deviation, minimum, and maximum value in Table V-3. Regarding the ICT infrastructure, there are 118 sampled schools which do not have computer available for their pupils for educational purposes (Israel has the most sampled school which do not have computer: 5 schools in 2009 wave, 3, 4, and 5 schools in 2012, 2015, and 2018 wave, respectively). About 109 sampled schools whose available computer do not have access to the internet, and about 21,493 sampled schools across countries and waves whose all-available computers are connected to the internet.

Table V-2 Means of the (numeric) independent variables (without ICT-related variables) across waves

Country	AGE	ESCS	MMINS	SMINS	LMINS	SCHSIZE	PROP_Girl	PROP_immig_nat	STRA-TIO	PRO-ATCE	STU-BEHA	TEACH-BEHA	EDU-SHORT	STAFF-SHORT
AUS	15.778	0.166	230.243	199.737	232.942	943.424	0.491	0.761	13.243	0.964	-0.074	0.094	0.002	-0.038
AUT	15.770	-0.008	184.500	225.826	182.303	446.695	0.481	0.813	10.881	0.857	-0.031	-0.052	-0.009	-0.075
BEL	15.809	0.042	210.216	185.577	211.161	692.092	0.479	0.823	9.025	0.859	0.142	0.206	0.158	0.329
CHE	15.790	0.053	214.603	175.699	217.639	516.852	0.487	0.784	11.979	0.841	0.048	0.014	0.137	-0.141
CHL	15.785	-0.186	323.904	274.963	305.949	886.805	0.492	0.906	19.828	0.237	-0.085	0.075	-0.381	0.096
CZE	15.775	-0.105	200.337	223.104	197.932	433.220	0.490	0.912	12.871	0.906	0.069	-0.122	0.007	-0.184
DEU	15.807	-0.027	208.956	203.969	207.248	701.529	0.479	0.782	14.676	0.908	0.023	0.092	0.092	0.458
DNK	15.777	0.227	244.131	206.449	284.841	513.743	0.486	0.797	12.085	0.910	-0.037	0.028	-0.158	-0.339
EST	15.796	-0.024	216.547	207.853	198.530	497.974	0.497	0.882	11.196	0.874	-0.021	-0.005	-0.006	0.109
FIN	15.750	0.115	199.284	179.424	188.367	430.960	0.487	0.896	10.487	0.890	-0.109	-0.021	-0.089	-0.170
GRC	15.724	-0.118	210.566	204.642	186.648	257.729	0.481	0.854	9.012	0.868	-0.157	-0.277	0.185	-0.006
HUN	15.748	-0.192	185.885	195.406	185.941	486.354	0.490	0.921	10.986	0.942	-0.139	-0.094	0.324	-0.244
IRL	15.735	0.041	205.331	167.889	197.642	597.865	0.506	0.846	13.558	0.980	-0.037	0.153	0.046	-0.041
ISL	15.759	0.229	228.331	167.277	227.110	288.673	0.484	0.902	9.026	0.848	0.012	0.157	-0.189	-0.058
ISR	15.727	0.101	232.201	197.336	205.054	743.619	0.531	0.838	11.488	0.706	0.063	0.037	0.079	0.353
ITA	15.766	-0.092	223.969	162.421	259.699	651.708	0.477	0.867	9.213	0.824	0.010	-0.085	0.051	0.246
JPN	15.771	-0.095	229.127	181.491	213.696	732.208	0.492	0.696	11.671	0.977	0.016	0.082	0.598	0.152
KOR	15.728	-0.049	205.121	182.644	200.038	960.846	0.492	0.938	15.333	0.966	0.031	-0.161	0.249	0.115
LVA	15.775	-0.192	223.318	225.706	182.225	436.793	0.499	0.912	9.750	0.728	0.037	-0.157	-0.122	-0.263
NZL	15.770	0.042	231.956	219.674	230.231	1,012.138	0.504	0.794	14.206	0.912	0.030	0.043	-0.004	-0.046
POL	15.746	-0.146	215.080	183.540	220.110	314.755	0.488	0.932	8.928	0.933	0.006	-0.003	0.040	-0.972
SVK	15.749	-0.174	210.304	179.888	211.931	386.956	0.488	0.924	12.887	0.888	-0.086	-0.261	-0.055	-0.407
SVN	15.743	-0.050	192.451	193.978	198.146	301.079	0.436	0.861	11.448	0.933	-0.155	-0.004	0.109	-0.520
SWE	15.749	0.161	208.869	189.919	199.182	426.649	0.494	0.827	11.941	0.852	-0.003	-0.006	-0.152	0.092

Table V-3 Descriptive statistics of ICT-related variables across waves

Country	COMPRATIO				WEBCOMP				ENTUSE				USESCH				HOMSCH			
	Mean	*Std. dev.*	*Min.*	*Max.*	*Mean*	*Std. dev.*	*Min.*	*Max.*	*Mean*	*Std. dev.*	*Min.*	*Max.*	*Mean*	*Std. dev.*	*Min.*	*Max.*	*Mean*	*Std. dev.*	*Min.*	*Max.*
AUS	1.395	1.332	0	34.696	0.990	0.069	0	1	-0.046	0.307	-3.710	2.787	0.273	0.497	-1.716	3.304	0.050	0.417	-2.297	3.307
AUT	1.240	2.606	0	55.000	0.980	0.111	0	1	-0.077	0.323	-1.125	2.765	-0.202	0.392	-1.408	1.321	-0.179	0.318	-2.297	1.326
BEL	0.833	0.731	0	6.818	0.966	0.121	0	1	-0.069	0.290	-1.902	2.515	-0.214	0.415	-1.246	1.742	-0.171	0.337	-1.590	3.307
CHE	0.662	0.715	0	10.286	0.987	0.074	0	1	-0.066	0.275	-1.791	1.949	-0.143	0.389	-1.716	1.683	-0.192	0.297	-1.956	1.034
CHL	0.624	0.547	0	4.200	0.932	0.196	0	1	-0.149	0.344	-3.594	4.244	-0.142	0.422	-1.668	3.304	-0.109	0.367	-2.483	3.194
CZE	0.921	0.686	0.04	12.500	0.989	0.064	0	1	-0.045	0.327	-1.810	2.061	-0.024	0.464	-1.290	2.020	-0.096	0.362	-1.570	1.757
DEU	0.622	1.007	0	18.000	0.966	0.124	0	1	-0.214	0.243	-1.322	0.965	-0.401	0.257	-1.104	1.107	-0.335	0.253	-2.689	0.543
DNK	0.948	0.818	0	12.500	0.995	0.061	0	1	-0.043	0.270	-0.844	1.833	0.244	0.562	-1.053	2.432	0.037	0.377	-1.766	1.820
EST	0.853	0.741	0.067	13.333	0.990	0.071	0.250	1	-0.092	0.262	-0.742	1.457	-0.280	0.355	-1.668	1.380	-0.029	0.384	-1.235	1.155
FIN	0.583	0.523	0	4.455	0.958	0.160	0	1	-0.124	0.217	-0.895	1.038	-0.122	0.370	-0.960	1.258	-0.307	0.276	-1.432	0.916
GRC	0.251	0.201	0	1.818	0.979	0.089	0.083	1	-0.017	0.420	-3.594	3.265	-0.232	0.451	-1.716	2.023	-0.061	0.438	-2.689	3.307
HUN	0.710	0.590	0	8.750	0.982	0.085	0	1	-0.053	0.405	-2.445	2.212	-0.245	0.458	-1.716	2.216	-0.074	0.471	-2.689	3.604
IRL	0.680	0.479	0	5.649	0.986	0.075	0	1	-0.144	0.228	-1.002	1.559	-0.431	0.259	-1.107	0.677	-0.394	0.224	-2.388	0.485
ISL	1.279	1.071	0	8.333	0.984	0.088	0.229	1	-0.197	0.280	-1.229	1.962	-0.125	0.472	-1.716	1.781	-0.178	0.372	-2.297	2.464
ISR	0.405	0.347	0	2.941	0.866	0.302	0	1	-0.253	0.270	-1.594	2.350	-0.267	0.369	-1.118	1.176	-0.203	0.349	-1.841	2.144
ITA	0.586	0.505	0	8.000	0.958	0.146	0	1	-0.001	0.337	-2.040	3.148	-0.019	0.455	-1.716	3.304	-0.092	0.382	-2.297	3.307
JPN	0.554	0.681	0	6.667	0.949	0.164	0	1	-0.322	0.196	-1.176	0.489	-0.779	0.411	-1.676	0.242	-0.757	0.474	-2.207	0.127
KOR	0.419	0.491	0	3.377	0.974	0.115	0	1	-0.262	0.227	-1.408	1.787	-0.672	0.369	-1.527	1.126	-0.258	0.274	-2.297	1.119
LVA	0.942	0.771	0.013	12.500	0.980	0.091	0.200	1	-0.051	0.336	-1.565	1.549	-0.042	0.467	-0.960	2.371	0.048	0.417	-0.939	2.150
NZL	1.187	0.750	0.15	5.442	0.987	0.088	0	1	-0.139	0.262	-0.716	2.139	-0.028	0.513	-1.668	1.301	-0.090	0.349	-0.788	1.912
POL	0.454	0.338	0.027	2.632	0.987	0.067	0.083	1	-0.080	0.309	-2.404	1.359	-0.234	0.406	-1.240	1.327	-0.044	0.364	-0.822	1.021
SVK	0.881	0.628	0.021	7.533	0.981	0.088	0.056	1	-0.115	0.391	-3.019	2.536	-0.058	0.511	-1.716	2.188	-0.022	0.436	-1.684	2.076
SVN	0.760	0.815	0	7.125	0.988	0.072	0.117	1	-0.137	0.345	-3.594	2.952	-0.097	0.436	-1.699	1.572	-0.063	0.388	-2.689	1.672
SWE	0.811	0.859	0	18.000	0.991	0.057	0.200	1	-0.111	0.268	-1.586	1.259	0.016	0.549	-1.668	1.976	-0.081	0.392	-2.689	1.932

Table V-4 Parameters estimation

Parameter	Model 1		Model 2		Model 3	
	Coef.	SE	Coef.	SE	Coef.	SE
First step						
Constant	-420.42**	48.017	-397.60**	47.271	-331.28**	48.360
AGE	56.942**	3.029	54.273**	2.988	50.636**	3.051
ESCS	78.774**	0.703	78.563**	0.705	81.740**	0.713
M(S/L)MINS	0.010	0.006	0.118**	0.005	-0.065**	0.006
SCHSIZE	0.007**	0.001	0.008**	0.0007	0.008**	0.0007
PROP_Girl	-5.734**	1.332	0.825	1.316	21.430**	1.340
PROP_nat	18.169**	1.823	21.649**	1.805	29.413**	1.838
STRATIO	0.123**	0.060	0.093	0.059	1.367**	0.060
STUBEHA	-2.291**	0.306	-1.969**	0.302	-3.004**	0.308
TEACHBEHA	0.204	0.316	0.498	0.311	1.264**	0.318
EDUSHORT	-0.076	0.276	-0.510*	0.273	-0.312	0.278
STAFFSHORT	0.357	0.282	0.331	0.279	-0.102	0.284
PROATCE	-0.415	1.141	-0.570	1.129	0.530	1.149
SCHTYPE:						
private-indep.	4.566**	1.025	5.232**	1.100	4.533**	1.121
Private gov.-dept.	6.384**	1.112	5.398**	1.014	5.094**	1.034
SCHLOC:						
small town	-2.876**	0.953	-3.359**	0.943	-3.164**	0.961
town	-3.623**	0.931	-3.896**	0.922	-3.234**	0.938
city	-5.605**	1.006	-5.865**	0.996	-4.352**	1.013
large city	-3.708**	1.220	-4.550**	1.208	-1.446	1.229
COMPRATIO	0.755**	0.334	0.923**	0.330	0.299	0.337
WEBCOMP	2.655	2.174	1.618	2.153	0.962	2.191
ENTUSE	-12.762**	1.059	-11.321**	1.051	-11.662**	1.069
USESCH	-25.660**	0.835	-23.669**	0.827	-27.093**	0.841
HOMSCH	10.047**	1.051	3.529**	1.042	10.778**	1.058
Time dummies	Yes		Yes		Yes	
Adjusted R^2	0.534		0.570		0.572	
Second step						
σ_u^2: Constant	4.952**	0.040	4.566**	0.054	3.834**	0.098
σ_v^2: Constant	4.196**	0.028	4.299**	0.026	4.116**	0.028
Third step						
σ_u^2:						
Constant	6.967**	0.180	7.053**	0.169	7.2038**	0.156
COMPRATIO	0.015	0.026	0.026	0.023	0.029	0.022
WEBCOMP	-0.400**	0.174	-0.427**	0.163	-0.367**	0.152
ENTUSE	-0.191**	0.088	-0.090	0.081	-0.239**	0.064
USESCH	0.158**	0.063	0.440**	0.063	0.387**	0.535
HOMSCH	0.310**	0.086	-0.067	0.077	0.101	0.066
σ_v^2 : Constant	6.622**	0.024	6.564**	0.024	6.513**	0.244

*significant at the level of 10%
**significant at the level of 5%

V.4 Results

V.4.1 ICT Influence on education outcomes

As the "four-component stochastic frontier model" involves three steps of estimation, in Step 1, the standard panel fixed-effects regression by including time (wave) effect is employed. The estimation result is shown in Table V-4, see Panel: First step. Note that there are three models with different domains: mathematics (Model 1), science (Model 2), and reading (Model 3).

Since the coefficient of student's age is positive and significant in all model specifications, it implies that maturity has a positive impact on student's proficiency that can be acquired by doing other things before studying or as a consequence of an increased capacity to organize their studies (Dolton et al., 2003). The anticipated significant positive value of the index of economic, social, and cultural status (ESCS) (in all models) indicates as the higher the economic, social, and cultural status of the student (aggregated to school level), the higher the scores will be obtained. This finding confirms the result of other studies, e.g., Perelman and Santín (2011a), Salas-Velasco (2020), and Ulkhaq (2021). The positive sign of the proportion of native students (PROP_nat) in a school implies that as the share of native students increases, the school's test scores tend to increase as well. The highest coefficient is found in the language test score. Schnepf (2008) mentioned this phenomenon by arguing that immigrants with a lack of language skills are likely to achieve more badly for language tasks. The positive sign is also found in school size, but the magnitude is very small (about 0.007~0.008). It means that an additional of 100 students would result in the increasing of 0.7~0.8 of school's test scores.

The negative sign is found in the index of student behavior hindering learning (STUBEHA), meaning that if the behavior of the students in a particular school is improved, the school tends to achieve better test scores. The negative (and statistically significant) coefficients are also found in school location (SCHLOC), indicating that schools located in denser population might have lower test scores for mathematics, science, and language. The share of girls (PROP_girl) is significant and has negative value on mathematics scores; has positive value in on language scores; but not significant on science scores. It suggests that the less female students, the particular school tends to get better result in mathematics proficiency but worse in language proficiency. This finding follows Mancebón et al. (2012) who found that girls perform better at language, but worse for mathematics and science, at which boys achieve better results in PISA 2006 edition.

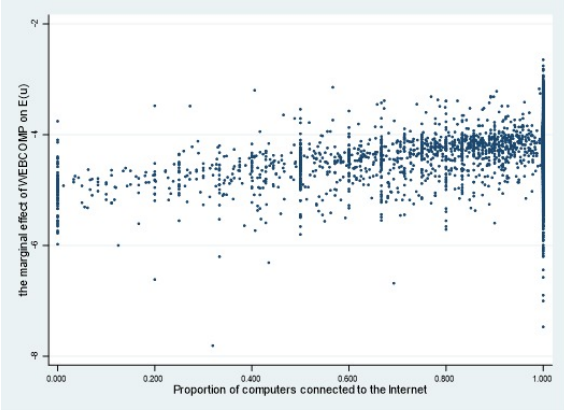
All variables of students' time spent in using ICT are significant. It is no surprise that the sign of students' time spent in using ICT for entertainment purposes (ENTUSE) is negative; meaning that the more students use ICT outside school for entertainment purposes, the lower test scores of the school would be obtained. The positive sign is found in the index of students' time spent in using ICT at home for school-related tasks (HOMSCH), indicating that if students spent their time in using ICT at home for school-related tasks, the school's test scores tend to increase. On the other hand, the slightly unanticipated result is found in the sign of the index of students' time spent in using ICT at school (USESCH), which is negative. One explanation could be that students might use the ICT device for non-educational purposes, e.g., for gaming or browsing for non-educational materials. The ratio of computers to the number of students is found significant to influence mathematics and science scores but not language score. Another ICT-related variable is the ratio of computers connected to the internet which is found to be not significant to influence mathematics, science, and language scores (similar to the finding of Salas-Velasco, 2020). An early observation is perhaps the presence of computer in a school can help students to improve their mathematics and science proficiency, but not about language. In sum, the finding suggests that students' time spent in using ICT does influence the education outcomes (for all domains), whereas the influence of the ratio of computers to the total number of students is only significant to influence mathematics and science scores.

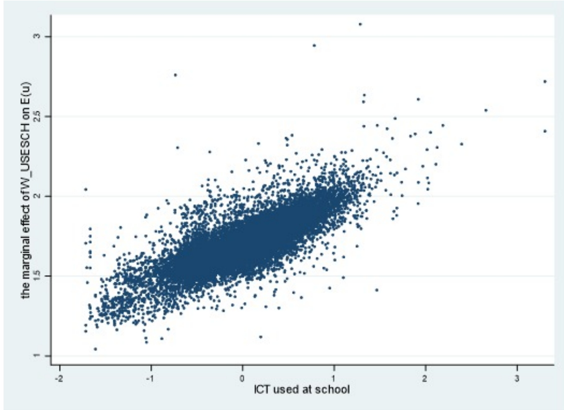
The second step of the analysis encompasses estimating persistent inefficiency. This is achieved by running the standard cross-sectional SFA—as in Equation (V-4)—on the error component $\hat{B}_i$ obtained at the first step. The estimation result is also shown in Table V-4, see Panel: Second step. Since the σ_u^2 coefficient is statistically significant, it indicates that the persistent inefficiency does exist. The persistent inefficiency refers to a long-term or structural inability of an education institution to achieve the potential level of academic outputs. It could be interpreted as a measure of the effect of educational policy to the educational institutions. This type of inefficiency varies across education institutions but not over time. Thus, unless there is a change in something that affects the management of education institutions such as a change in the government policy toward this industry, it is very unlikely that the persistent inefficiency component will be reduced.

V.4.2 ICT Influence on inefficiency

The third step involves estimating time-varying inefficiency. In this step, the influence of ICT on inefficiency, called the determinants of inefficiency, is also estimated. Five ICT-related variables are modeled as the determinants, i.e., the ratio of computers to the total number of students (COMPRATIO), the ratio of computers connected to the internet (WEBCOMP), and the index of time spent by students in using ICT: for entertainment purposes (ENTUSE), at school (USESCH), as well as at home for school-related tasks (HOMSCH). The estimation result is shown in Table V-4, see Panel: Third step.

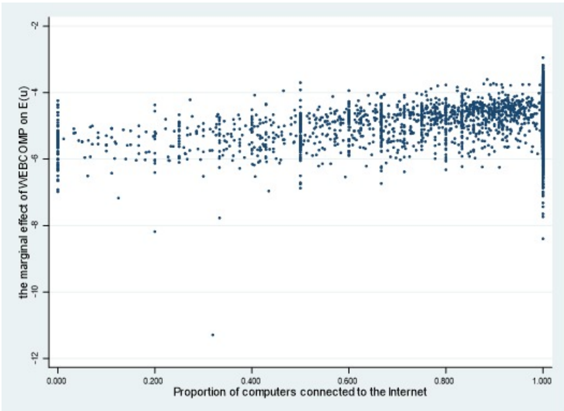
The ICT infrastructure behaves differently to influence time-varying inefficiency. The ratio of computers to the number of students has no influence but the ratio of computers connected to the internet has influence on time-varying inefficiency in all models. However, the infrastructure policy is not enough as this policy has to be accompanied by the use of this infrastructure. In this study, the frequency of ICT use by students are measured by three variables, i.e., the index of student's time spent in using

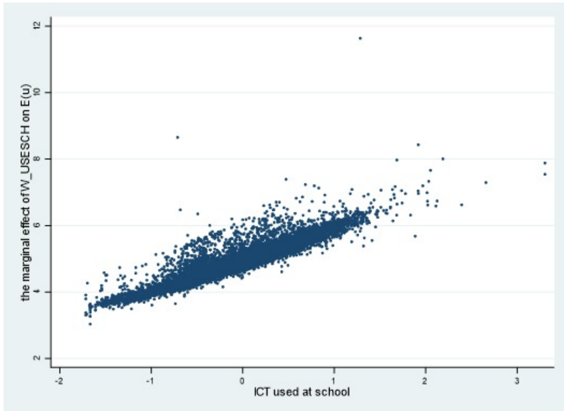




WEBCOMP USESCH

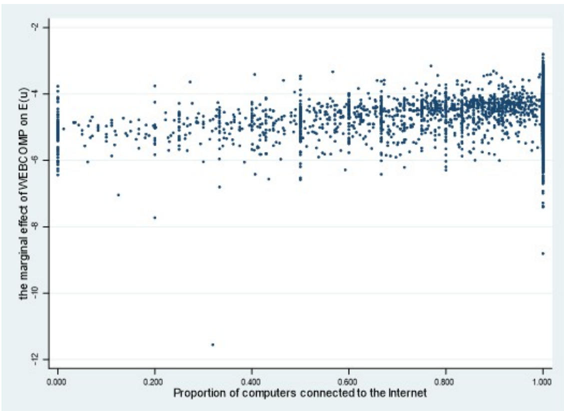
(a) Model 1

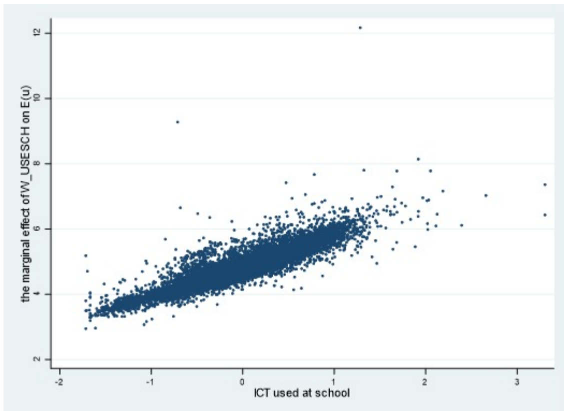




WEBCOMP USESCH

(b) Model 2





WEBCOMP USESCH

(c) Model 3

Figure V-3 The marginal effect of the determinants of inefficiency on $E[u]$

ICT at school (USESCH), at home for school-related tasks (HOMSCH), and outside school for entertainment purposes (ENTUSE). The tendency of students to use ICT at school is found to be a significant factor that can reduce inefficiency. Other variables behave differently, the index of time spent by students in using ICT outside school for entertainment purposes is significant to influence school's inefficiency on mathematics and reading test scores, but the index of time spent by students in using

ICT at home for school-related tasks is only significant to influence school's inefficiency on mathematics test score.

The marginal effect of the determinants of inefficiency on the expected value of inefficiency is shown in Figure V-3. Only the determinants that are significant in all models are shown, i.e., the ratio of computers connected to the internet (WEBCOMP) and the index of student's time spent in using ICT at school (USESCH). In Model 3, for instance, the mean of the marginal effects of WEBCOMP on $E[u]$ is negative; thus, increasing the ratio of computers connected to the internet would decrease, on average, the level of inefficiency. In particular, the level of inefficiency is reduced, on average, by 4.482 for every one value increase in the ratio of computers connected to the internet. The tendency of students to use ICT at school is found to be a significant factor as the mean of the marginal effects of USESCH on $E[u]$ is positive. Following the finding from Step 1, it can be observed that the more students spent their time using digital devices at school, on average, the test scores would decrease but inefficiency would increase.

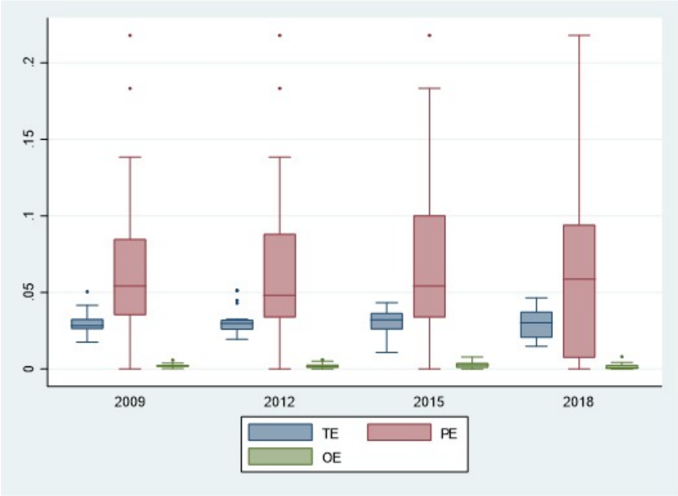
V.4.3 Efficiency estimation

Inefficiency is composed into persistent and time-varying inefficiency. Persistent efficiency is calculated using BC estimator after applying the standard cross-sectional SFA as in Equation (V-4), while time-varying efficiency is also obtained by using BC estimator after applying the standard cross-sectional SFA as in Equation (V-5). Finally, the overall efficiency is calculated as in Equation (V-6). The result is displayed in Figure V-4.

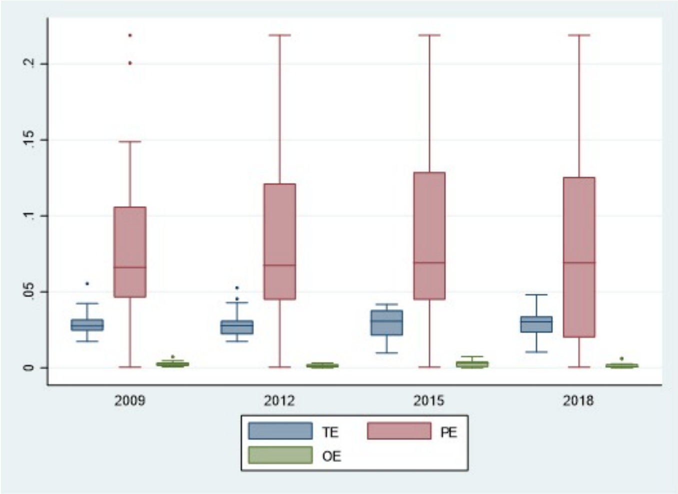
The values of persistent efficiency are higher and more dispersed (in all waves) than the values of time-varying efficiency. It could imply that the education system uniquely in each country plays major role in constructing efficiency in education. The nature of persistent inefficiency is time-invariant so that it can only be changed in the long-run through some restructuring of the school. On the other hand, time-varying inefficiency can be changed in the short-run. In this study, ICT-related variables are modeled as determinants of time-varying inefficiency. Policy makers could (or should) decrease this time-varying inefficiency by manipulating those ICT-related variables. For example, policy makers should increase the number of computers connected to internet at schools. As has been described earlier, this might decrease school's inefficiency (or increase efficiency). Another strategy would limit time spending of students in using digital devices at school as this might increase school's efficiency while at the same time improve their mathematics, science, and language proficiency.

V.4.4 Robustness checking

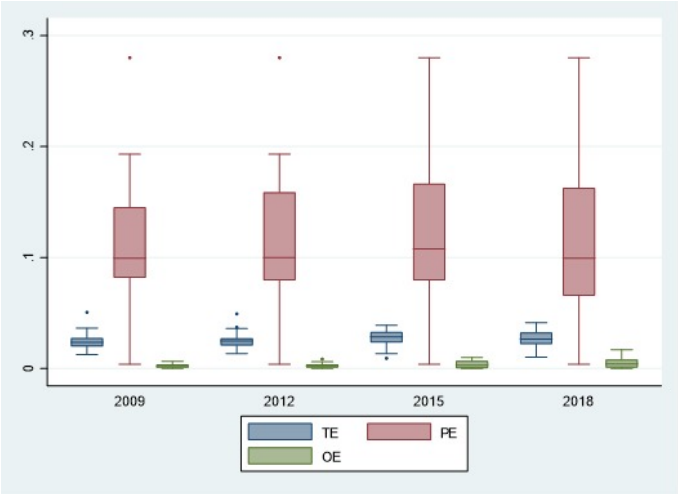
To examine the robustness of the findings, three tests are performed. The first test examines whether the sign and significance of ICT-related variables (as inputs and determinants of inefficiency)



(a) Model 1



(b) Model 2



(c) Model 3

Figure V-4 Box plots of the persistent (PE), time-varying (TE), and overall efficiency (OE)

differ when other test scores are used. Second, by excluding non-ICT-related variables which are not significant in all models at the first step, the *new* model is re-estimated, the efficiency scores are re-calculated, and then the result is compared to the previous result. Lastly, the efficiency scores are re-calculated, relaxing the hypothesis of existing (unobserved) heterogeneity across education institutions.

In the first test, three other corresponding PISA scores are used as dependent variable alternatives. Result of the first robustness analysis is shown in Table V-5. In all models, the sign and significancy of the ratio of computers to the total number of students (COMPRATIO)'s coefficients are not changed. As the input, the coefficients are still significant with positive value in Model 1 and 2; but not significant in Model 3. As the determinant of inefficiency, the coefficient values are not significant in all models. The ratio of computers connected to the internet is still not significant as input but significant as determinant of inefficiency. The coefficients of variables related to students' time spent in using ICT (ENTUSE, USESCH, and HOMSCH) as inputs are all significant in all models; while the influences are marginal as determinants of inefficiency. The values of the coefficients, if one observes, are slightly similar; the difference is trivial. In sum, we could say that the model is robust.

In the second test, non-ICT-related variables which are not significant at the level of 5% in all models at the first step are excluded, i.e., proportion of fully certified teacher (PROATCE), the index of educational material shortage (EDUSHORT), and the index of educational staff shortage (STAFFSHORT). The sign and significancy of the remaining variables are again investigated. The result is displayed in Table V-6. In Model 1, at the first step, the ratio of computers connected to the internet becomes significant at the level of 10% as in the baseline model this variable is not significant. In Model 3, at the third step, index of time spent in using ICT at home becomes significant at the level of 5% as in the baseline model this variable is not significant. Note that the changes in the magnitude are trivial. The only change in the sign is the index of time spent in using ICT at home as determinant of inefficiency in Model 2, in which in the baseline model is negative but in the robustness check model is positive. However, this finding can be neglected since this variable is not significant.

The changes in efficiency scores are also inspected. It is expected that by removing the non-significant variables, the efficiency would increase. The result is shown in Table V-7. One can observe that as the insignificant non-ICT-related variables are removed, the efficiency score increases.

The last test is performed to test whether the heterogeneity assumption in the model specification influences the results. The present model accounts for education institution heterogeneity, time-varying and persistent inefficiency, as well as statistical noise. The question raises whether part or all of the structural differences between education institution (that are supposed to be picked up by the persistent inefficiency), are accidentally eliminated in the estimation. To handle these challenges, with the same data, we estimate the model specification of Kumbhakar and Heshmati (1995) (KH) and then compare the results we obtained to the baseline model. The difference between two specifications only lies within

Table V-5 Robustness analysis – parameters estimation using different PISA scores

Parameter	Model 1				Model 2				Model 3			
	Baseline	PV2	PV3	PV4	Baseline	PV2	PV3	PV4	Baseline	PV2	PV3	PV4
As inputs												
COMPRATIO	0.755**	0.874**	0.708**	0.863**	0.923**	0.772**	0.852**	0.909**	0.299	0.099	0.259	0.389
WEBCOMP	2.655	2.742	3.053	3.354	1.618	1.393	0.394	1.597	0.962	0.912	0.111	1.358
ENTUSE	-12.762**	-11.933**	-11.829**	-12.512**	-11.321**	-12.856**	-10.809**	-10.713**	-11.662**	-11.267**	-10.973**	-11.363**
USESCH	-25.660**	-25.795**	-25.912**	-25.705**	-23.669**	-23.361**	-23.677**	-23.779**	-27.093**	-28.009**	-27.492**	-27.141**
HOMSCH	10.047**	9.777**	9.655**	9.399**	3.529**	5.640**	3.062**	3.426**	10.778**	11.042**	10.229**	10.150**
As determinants of inefficiency												
COMPRATIO	0.015	0.008	-0.007	0.009	0.026	0.029	0.021	0.016	0.029	0.030	0.030	0.035
WEBCOMP	-0.400**	-0.408**	-0.331*	-0.397**	-0.427**	-0.370**	-0.418**	-0.399**	-0.367**	-0.340**	-0.364**	-0.322**
ENTUSE	-0.191**	-0.153*	-0.150*	-0.220**	-0.090	-0.074	-0.143*	-0.097	-0.239**	-0.205**	-0.262**	-0209**
USESCH	0.158**	0.173**	0.176**	0.145**	0.440**	0.415**	0.425**	0.407**	0.387**	0.418**	0.404**	0.396**
HOMSCH	0.310**	0.326**	0.317**	0.391**	-0.067	-0.027	-0.005	-0.004	0.101	0.059	0.091	0.079

*significant at the level of 10%;
**significant at the level of 5%

Table V-6 Robustness analysis – parameters estimation excluding non-significant-ICT-related variables

Parameter	Model 1		Model 2		Model 3	
	Baseline	Robustness	Baseline	Robustness	Baseline	Robustness
First step						
Constant	-420.42**	-427.97**	-397.60**	-391.43**	-331.28**	-319.89**
COMPRATIO	0.755**	0.848**	0.923**	0.972**	0.299	0.464
WEBCOMP	2.655	3.557*	1.618	2.596	0.962	2.100
ENTUSE	-12.762**	-11.101**	-11.321**	-9.260**	-11.662**	-10.377**
USESCH	-25.660**	-25.635**	-23.669**	-25.320**	-27.093**	-27.431**
HOMSCH	10.047**	7.955**	3.529**	2.396**	10.778**	8.438**
AGE	56.942**	57.305**	54.273**	53.825**	50.636**	49.734**
ESCS	78.774**	77.340**	78.563**	77.418**	81.740**	80.170**
M(S/L)MINS	0.010	0.007	0.118**	0.105**	-0.065**	-0.056**
SCHSIZE	0.007**	0.008**	0.008**	0.009**	0.008**	0.008**
PROP_Girl	-5.734**	-4.929**	0.825	1.320	21.430**	22.491**
PROP_immig_nat	18.169**	19.638**	21.649**	24.098**	29.413**	29.168**
STRATIO	0.123**	0.080	0.093	0.022	1.367**	0.061
STUBEHA	-2.291**	-2.501**	-1.969**	-2.431**	-3.004**	-3.281**
TEACHBEHA	0.204	0.601**	0.498	1.033**	1.264**	1.641**
EDUSHORT	-0.076		-0.510		-0.312	
STAFFSHORT	0.357		0.331		-0.102	
PROATCE	-0.415		-0.570		0.530	
SCHTYPE:						
private-indep.	4.566**	5.117**	5.232**	5.493**	4.533**	4.187**
Private gov.-dept.	6.384**	6.575**	5.398**	5.640**	5.094**	4.648**
SCHLOC:						
small town	-2.876**	-3.220**	-3.359**	-3.365**	-3.164**	-3.245**
town	-3.623**	-3.944**	-3.896**	-3.930**	-3.234**	-3.187**
city	-5.605**	-5.718**	-5.865**	-5.665**	-4.352**	-4.188**
large city	-3.708**	-4.067**	-4.550**	-4.108**	-1.446	-1.551
Second step						
σ_u^2: Constant	4.952**	4.298**	4.566**	-3.520	3.834**	-6.087
σ_v^2: Constant	4.196**	4.348**	4.299**	4.586**	4.116**	4.335**
Third step						
σ_u^2:						
Constant	6.967**	6.618**	7.053**	6.570**	7.2038**	7.341**
COMPRATIO	0.015	0.011	0.026	0.021	0.029	0.029
WEBCOMP	-0.400**	-0.469**	-0.427**	-0.456**	-0.367**	-0.436**
ENTUSE	-0.191**	-0.232**	-0.090	-0.105	-0.239**	-0.250**
USESCH	0.158**	0.096*	0.440**	0.354**	0.387**	0.317**
HOMSCH	0.310**	0.350**	-0.067	0.012	0.101	0.153**
σ_v^2: Constant	6.622**	6.618**	7.120**	6.570**	6.513**	6.497**

*significant at the level of 10%
**significant at the level of 5%

the assumption regarding heterogeneity. The KH model contains only three components: persistent and time-varying inefficiency and statistical noise. It only contains one time-invariant parameter (i.e., persistent inefficiency). Heterogeneity is compounded in the inefficiency term, presumably leading to overall relatively lower efficiency value. If our assumption is correct and heterogeneity is only accounted for at the individual level in our model, one would expect the estimated values of efficiency in KH model to be lower (because unchangeable factors are still included in the inefficiency term).

Table V-7 Robustness analysis – efficiency scores excluding non-significant-ICT-related variables

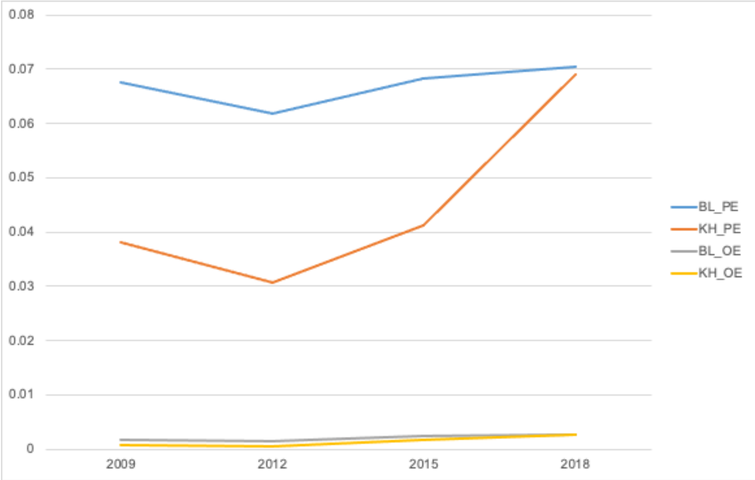
Baseline Model									
Wave	Model 1			Model 2			Model 3		
	PE	TE	OE	PE	TE	OE	PE	TE	OE
2009	0.06763	0.02939	0.00174	0.08098	0.02848	0.00207	0.11482	0.02489	0.00262
2012	0.06173	0.02984	0.00158	0.07732	0.02836	0.00190	0.11356	0.02517	0.00258
2015	0.06827	0.03119	0.00246	0.08161	0.03066	0.00286	0.11651	0.02819	0.00360
2018	0.07031	0.03147	0.00259	0.08294	0.03071	0.00292	0.11537	0.02847	0.00362
Robustness Check									
Wave	Model 1			Model 2			Model 3		
	PE	TE	OE	PE	TE	OE	PE	TE	OE
2009	0.09326	0.02895	0.00245	0.087631	0.02778	0.02434	0.096307	0.02366	0.02278
2012	0.08781	0.02938	0.00231	0.087626	0.02763	0.02420	0.096307	0.02389	0.02301
2015	0.09219	0.03079	0.00314	0.087630	0.02994	0.02624	0.096307	0.02712	0.02612
2018	0.09498	0.03091	0.00328	0.087633	0.02981	0.02613	0.096307	0.02729	0.02628

Figure V-5 shows the comparison of KH model and the baseline model. Note that we just show the values of persistent (PE) and overall efficiency (OE) since the values of time-varying efficiency are identical in both models due to similar specification. The persistent inefficiency of KH model is lower than the baseline model, showing that heterogeneity is an important factor in the education institution. This result shows that the present model is robust even when different assumptions about the role of education institution's heterogeneity are applied.

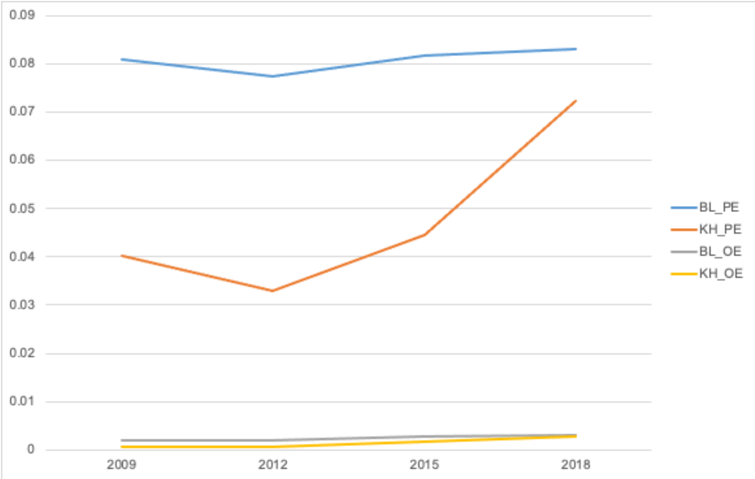
V.5 Discussion and Concluding Remarks

This study aims to investigate the influence of ICT on education outcomes and inefficiency. The analysis is performed using the "four-component stochastic frontier model" that separates education institution effects, persistent and time-varying inefficiency, as well as statistical noise. This study uses PISA data from 2009 to 2018 waves of 24 OECD countries. The ICT-related variables used are the ratio of computer at school to the total number of students for educational purposes, the ratio of computers connected to the internet, and the indices of time spent by students in using ICT: (i) outside school for entertainment purposes, (ii) at school, and (iii) at home for school-related tasks. We use three model specifications as each dependent variable reflecting test scores of mathematics (Model 1), science (Model 2), and reading (Model 3).

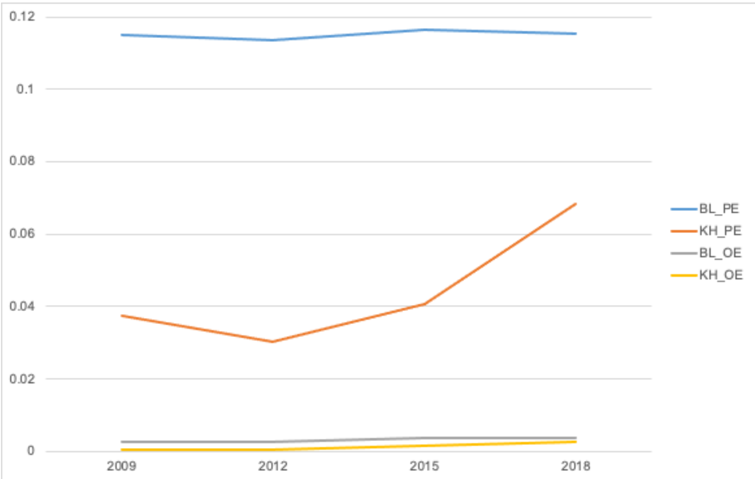
The results suggest that not all ICT-related variables significantly influence education outcomes and inefficiency. The ratio of computers to the total number of students influences positively and significantly school's test score of mathematics and science, but not language test score. This finding is partially similar to the finding of Perelman and Santín (2011b) who found that the ratio of computer to the total number of students did not significantly influence mathematics and language scores; and Zoghbi et al. (2013) who argued that the ratio was only significant in three (out of six) models they proposed. As the determinants of inefficiency, this variable does not have any influence on inefficiency.



(a) Model 1



(b) Model 2



(c) Model 3

Figure V-5 Comparison between the baseline model (BL) and Kumbhakar & Heshmati (1995)'s model (KH)

Another ICT-related variable is the ratio of computers connected to the internet which is found to be not significant to influence mathematics, science, and language scores (similar to the finding of Salas-Velasco, 2020). However, manipulating this ratio is argued to improve efficiency. Time spent by students in using ICT does significantly influence the education outcomes, measured by the PISA score of mathematics, science, and language. However, as the determinants of inefficiency, the influence of these variables have only a marginal effect on inefficiency. The influence of time spent in using ICT at school is significant influencing inefficiency in terms of mathematics, science, and language proficiencies. On the other side, the influence of time spent in using ICT for entertainment purposes is only significant influencing inefficiency in terms of mathematics and language, while the influence of time spent in using ICT at home for school-related tasks is only significant influencing inefficiency in terms of mathematics proficiency.

The practical implications suggested by the findings of this study are the following. The influence of variables related to the ICT infrastructure on education outcomes and inefficiency has the opposite result, meaning that increasing the number of computers and computer connected to the internet might increase students' proficiency, but on the other side might lower school's efficiency. It is of interest to highlight the classic argument of Hanushek (1996, 2003) that putting more money into schools does not guarantee per se better education outcomes—in this context the money can be considered as an investment in ICT by installing more computers and connecting them to the internet. On the other side, it can be observed that the more students spent their time in using ICT at school, the literacy and school's efficiency would decrease. When they spent the time more for non-school related works, their proficiency would decrease but school's efficiency might increase.

One policy lesson that might be derived from this study is discussed as the following. As the provision of ICT infrastructure is considered a key element for schools to be able to exploit the many benefits that digital technologies bring to teaching and learning, however, having access to digital technologies does not automatically translate into high rates of use. A survey from the European Commission reported that around 50% of students at grades 8 and 11 in general education use a desktop or a laptop during lessons at school at least weekly, but around 20% of the students at the same grades never or almost never use a computer during lessons (European Commission, 2013). As this study found that the frequency of ICT use by students does significantly influence the education outcomes, hence, it suggests that infrastructure-related policies should be accompanied by complementary measures in other areas, such as the use of this infrastructure.

CHAPTER VI. CONCLUSION

This chapter provides a summary of the dissertation, limitations of the research, and future research directions.

VI.1 Summary

This dissertation consists of four studies discussing the role of ICT in a relation to the efficiency analysis. The first study presented in Chapter II provides a systematic literature review on the role of ICT in the literature of efficiency analysis. Using the Scopus database, there are 41 articles that satisfy the predefined criteria to be further analyzed. First, scientometrics analyses are used to visualize the bibliometric clusters, i.e., co-authorship analysis, co-citation analysis, co-occurrence analysis, and bibliographic coupling analysis. According to the results of the scientometric analyses, this study revealed which authors produced the most articles; which authors and journals cited the most; which keywords appeared the most; and which countries contributed the most in this research domain.

Next, a qualitative approach is used to classify the extracted articles according to the level of analysis, data source, method used, as well as the ICT-related variables used—whether as inputs, outputs, or determinants of (in)efficiency. The findings suggest that non-parametric method, i.e., data envelopment analysis (DEA) has been preferred to other methods, such as stochastic frontier analysis (SFA). The majority of the studies used the international database such as PISA, TIMSS, and PIRLS that contains very useful information about education outcomes and other background information related to students and schools. It is not surprising that most of ICT-related variables are derived from these databases. ICT-related variables are mainly used as inputs in the model. The output is usually associated with the achievements of the students, therefore only two papers dealing with ICT-related variables as outputs are analyzed in this review. Quality of school resources related to ICT is considered as an input in most of the papers being reviewed. By considering ICT-related variables as inputs, this study shows that there are contrasting findings on the impact of these variables on education outcomes. For instance, the quality of school resources was found to have positive and significant influence on the education outcomes (Salas-Velasco, 2020), but has negative influence in Ferrera et al. (2011) and Perelman and Santìn (2011b). As the determinants of inefficiency, only one article included ICT into the model (i.e., internet users), and the result showed that ICT has no significant influence. Three articles included ICT as determinants of efficiency into their two-stage DEA models: quality of school resources was found to be significant in Agasisti and Zoido (2019), proportion of students who have regular access to the internet was found to be significant in Agasisti (2014), while ICT possession at home was found to only significant in smaller samples (Deutsch et al., 2013). According to this review, it is hard to conclude the influence of ICT-related variables on the education outcomes and (in)efficiency.

Chapter II also provides three critical discussions. The first critical remark is that there are limited studies about international comparisons of the efficiency of education analysis. The second remark is about the definition of ICT as previous studies only dealt with the physical infrastructure of ICT. The last remark discusses the method used in the efficiency measurement. It should be noted that articles reviewed in Chapter II did not apply the recent method in analyzing efficiency. The next three studies depicted in Chapter III to Chapter V are built upon these remarks.

Chapter III aims to measure efficiency of schools in six countries in South-East Asia. This study uses the recent OECD PISA 2018 data, which is well-regarded as an authoritative source of comparison for educational achievement across the world; thus, it deals with the first remark. The SFA allowing for heteroscedasticity is used. The ICT infrastructure variables, i.e., ratio of computers at school to the total number of students (COMPRATIO) and ratio of computers connected to the internet (WEBCOMP) are modeled as inputs and determinants of inefficiency. The result reveals that Singapore has the (relatively) best performance among other countries in terms of efficiency in education proxied by the PISA scores of mathematics, science, and reading. Among thirteen inputs used, seven inputs are statistically significant (at the level of 5%) to influence the education outcomes. COMPRATIO is found to be not significant influencing education outcomes while WEBCOMP does influence. As the determinant of inefficiency, WEBCOMP has no influence, while COMPRATIO affects school's efficiency in terms of mathematics and science.

Chapter IV also analyses the efficiency of schools in South-East Asian countries. However, different to Chapter III, study in Chapter IV uses the two-stage super-efficiency approach. ICT infrastructure is used as inputs in the first stage and determinants of efficiency in the second stage. The super-efficiency model has the ability to differentiate among the efficient schools. The bootstrapped quantile regression is used in the second stage to investigate the influence of determinants of efficiency. The ratio of computers to the total number of students is not significant in all observed quantiles but is significant in 25% percentile. On the other hand, ratio of computers connected to the internet is found to be significant. The results suggest a number of policy implications for South-East Asian schools, indicating different courses of action for schools with higher and lower efficiency levels.

Chapter V aims to investigate the influence of ICT infrastructure (COMPRATIO and WEBCOMP) as well as ICT use, including index of time spent by students in using ICT at school (USESCH), outside school for entertainment purposes (ENTUSE), and at home for school-related tasks (HOMSCH), on both education outcomes and time-varying inefficiency. Apart from ICT infrastructure, study in this chapter adds more ICT-related variables; thus, it deals also with the second remark of Chapter II. Using the OECD Pisa data from 2009 to 2018 wave from 24 OECD countries, this study extends the application of the "four-component stochastic frontier model"; and thus, deals with the first and third remarks of Chapter II. Results show that WEBCOMP is not statistically significant to influence education outcomes, while COMPRATIO does have influence education outcomes in terms of mathematics and science. On the other hand, all three variables belong to ICT use is influencing education outcomes. As the determinants of time-varying inefficiency, COMPRATIO is not significant while WEBCOMP and USESCH are. HOMSCH is only significant in terms of mathematics whereas ENTUSE is significant in terms of mathematics and reading. This study is then expected to provide a more holistic view of the role of ICT in the efficiency of education measurement as the previous studies only addressed the ICT infrastructure.

VI.2 Limitations and Future Research Directions

As the OECD PISA data is used, prior achievements cannot be considered; and therefore, the robustness of the finding only relies upon the ability of ESCS to capture prior academic history of the students. Next, the ICT-related variables incorporated in the model only include the physical infrastructure and the intensity of use. It does not account for, say, the expenditure related to ICT. This information is considerable important in the efficiency analysis (se e.g., Johnes, 2006, 2008, 2013). These two limitations are due to data limitations (as a matter of fact, this information is not available in the PISA data); and hopefully, future research will relax these constraints and stimulate further advancements in the knowledge of the field.

The study presented in Chapter III can be extended to be performed using the previous OECD PISA waves. It aims to examine whether inefficiency has been persistent over time or time varying. This extension can be compared to the study in Chapter V to investigate the influence of ICT in more heterogenous educational systems. Such variations in the efficiency levels could be addressed by different policies applied in different countries.

REFERENCES

Abu-Bakar, M., Nozaki, Y., & Luke, A. (2006). Recent educational reforms in Singapore: Towards a new perspective for educational policy and practice. *The Journal of Educational Sociology, 78*, 469-485.

Agasisti, T. (2011). How competition affects schools' performances: Does specification matter?. *Economics Letters, 110*(3), 259-261.

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.

Agasisti, T. (2014). The efficiency of public spending on education: An empirical comparison of EU countries. *European Journal of Education, 49*(4), 543-557.

Agasisti, T., & Gralka, S. (2019). The transient and persistent efficiency of Italian and German universities: A stochastic frontier analysis. Applied Economics, 51(46), 5012-5030.

Agasisti, T., & Zoido, P. (2018). Comparing the efficiency of schools through international benchmarking: Results from an empirical analysis of OECD PISA 2012 data. *Educational Researcher, 47*(6), 352-362.

Agasisti, T., & Zoido, P. (2019). The efficiency of schools in developing countries, analysed through PISA 2012 data. *Socio-Economic Planning Sciences, 68*, 100711.

Agasisti, T., & Zoido, P. (2019). The efficiency of schools in developing countries, analysed through PISA 2012 data. *Socio-Economic Planning Sciences, 68*, 100711.

Agha, S. R., Kuhail, I., Abdul Nabi, N., Salem, M., & Ghanim, A. (2011). Assessment of academic departments efficiency using data envelopment analysis. *Journal of Industrial Engineering and Management*, *4*(2), 301-325.

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*, 21–37.

Al-araibi, A. A. M., Naz'ri bin Mahrin, M., Yusoff, R. C. M., & Chuprat, S. B. (2019). A model for technological aspect of e-learning readiness in higher education. *Education and Information Technologies, 24*(2), 1395-1431.

Alves, P. J. H., & Araújo, J. M. D. (2018). A study on the educational results obtained by municipalities of Paraíba in the years 2011, 2013 and 20151, 2. *Ensaio: Avaliação e Políticas Públicas em Educação, 26*, 1038-1057.

Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science, 39*(10), 1261-1264.

André, C., Pareliussen, J., & Hwang, H. (2020). Swedish school results, student background, competition and efficiency. *Educational Studies Moscow, 2020*(3), 8–36.

Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning, *The Economic Journal*, *112*, 735-765.

Angrist, J., Chernozhukov, V., & Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the US wage structure. *Econometrica, 74*(2), 539-563.

Aparicio, J., Cordero, J. M., & Ortiz, L. (2019). Measuring efficiency in education: The influence of imprecision and variability in data on DEA estimates. *Socio-Economic Planning Sciences, 68*, 100698.

Aristovnik, A. (2012). The impact of ICT on educational performance and its efficiency in selected EU and OECD countries: A non-parametric analysis. *Turkish Online Journal of Educational Technology 11*(3), 144-152.

Aristovnik, A. (2013), ICT expenditures and education outputs/outcomes in selected developed countries: An assessment of relative efficiency. *Campus-Wide Information Systems*, *30*(3), 222-230.

Aristovnik, A. (2014). Development of the information society and its impact on the education sector in the EU: Efficiency at the regional (NUTS 2) level. *Turkish Online Journal of Educational Technology*, *13*(2), 54-60.

Arshad, M., Amjath-Babu, T. S., Aravindakshan, S., Krupnik, T. J., Toussaint, V., Kächele, H., & Müller, K. (2018). Climatic variability and thermal stress in Pakistan's rice and wheat systems: A stochastic frontier and quantile regression analysis of economic efficiency. *Ecological indicators, 89*, 496-506.

Asadullah, M. N., Perera, L. D. H., & Xiao, S. (2020). Vietnam's extraordinary performance in the PISA assessment: A cultural explanation of an education paradox. *Journal of Policy Modeling, 42*(5), 913-932.

Asian Development Bank (2011). *Asian Development Outlook 2011*. Manila: Asian Development Bank.

Avvisati, F. (2020). The measure of socio-economic status in PISA: A review and some suggested improvements. *Large-scale Assessments in Education 8*(8).

Badunenko, O., Mazrekaj, D., Kumbhakar, S. C., & de Witte, K. (2021). Persistent and transient inefficiency in adult education. *Empirical Economics*, *60*(6), 2925-2942.

Ballestamon, S. U., Narvasa, B. L, Cabasal, M. P, Gonda, B. A., & Prado, E. G. (2000). The Filipino's commitment to quality education. *Journal of Southeast Asian Education*, *1*(1), 163-184.

Banker, R. D., & Natarajan, R. (2008). Evaluating contextual variables affecting productivity using data envelopment analysis. *Operations Research, 56*(1), 48-58.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Models for the estimation of technical and scale inefficiencies in Data Envelopment Analysis. *Management Science, 30*, 1078-1092.

Barber, M., & M. Mourshed (2007). *How the World's Best-Performing School Systems Come Out on Top*. London: McKinsey & Company.

Barnett, R., Glass, J., Snowdon, R., & Stringer, K. (2002). Size, performance and effectiveness: cost-constrained measures of best-practice performance and secondary school size‖. *Education Economics, 10*(3), 291-311.

Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of eEconometrics, 38*(3), 387-399.

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics, 20*(2), 325-332.

Biagi, F., & Loi, M. (2012). *ICT and Learning: Results from PISA 2009*. Luxembourg: Publications Office of the European Union.

Biagi, F., & Loi, M. (2013). Measuring ICT use and learning outcomes: Evidence from recent econometric studies. *European Journal of Education, 48*(1), 28-42.

Block, F., & Keller, M. R. (2009). Where do innovations come from? Transformations in the US economy, 1970–2006. *Socio-Economic Review, 7*(3), 459–483.

Bogetoft, P. (2012). *Performance Benchmarking: Measuring and Managing Performance*. Springer.

Bradley, S., & Taylor, J. (1998). The effect of school size on exam performance in secondary schools. *Oxford Bulletin of Economics and Statistics, 60*(3), 291-324.

Cancino, C., Merigó, J. M., Coronado, F., Dessouky, Y., & Dessouky, M. (2017). Forty years of Computers & Industrial Engineering: A bibliometric analysis. *Computers & Industrial Engineering, 113*, 614-629.

Carlota D.-G., & Ignacio C. (2021). A DEA-inspired model to evaluate the efficiency of education in OECD countries. *Revista de Métodos Cuantitativos para la Economía y la Empresa, 31*, 329-346.

Carlsson, B., & Fridh, A. C. (2002). Technology transfer in United States universities. *Journal of Evolutionary Economics, 12*(1), 199-232.

Carroll, J. (1963). A model of school learning. *Teachers college record, 64*(8), 723-723.

Castano, M. C. N., & Cabanda, E. C. (2007). Performance evaluation of the efficiency of Philippine Private Higher Educational Institutions: Application of frontier approaches. *International Transactions in Operational Research, 14*(5), 431-444.

Chakraborty, K., Biswas, B., & Lewis, W. C. (2001). Measurement of technical efficiency in public education: A stochastic and nonstochastic production function approach. *Southern Economic Journal, 67*(4), 889-905.

Charnes, A., Cooper, W. W., Golany, B., Seiford, L. M., & Stutz, J. (1985). Foundation of data envelopment analysis and Pareto-Koopmans empirical production functions. *Journal of Econometrics*, *30*, 91-107.

Charnes, A., & Cooper, W. W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly, 15*, 333-334.
Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research, 2*(6), 429-444.
Chen, X., & Shu, X. (2021). The Scientific and Technological Innovation Performance of Chinese World-Class Universities and its Influencing Factors. *IEEE Access, 9*, 84639-84650.
Chen, Z., Yang, Z., & Yang, L. (2020). How to optimize the allocation of research resources? An empirical study based on output and substitution elasticities of universities in Chinese provincial level. *Socio-Economic Planning Sciences, 69*, 100707.
Cheng, Y. C. (1999). Recent education developments in South East Asia: An introduction. *School Effectiveness and School Improvement, 10*(1), 3-9.
Chiang, T. (2021a). A fuzzy-based hybrid approach for estimating interdisciplinary learning efficiency. *IEEE Access, 9*, 143275-143283.
Chiang, T. (2021b). Estimating the artificial intelligence learning efficiency for civil engineer education: A case study in Taiwan*. Sustainability, 13*(21), 11910.
Ciroma, Z. I. (2014). ICT and education: Issues and challenges. *Mediterranean Journal of Social Sciences, 5*(26), 98-100.
Coelli, T. (1995). Estimators and hypothesis tests for a stochastic frontier function: A monte carlo analysis. *Journal of Productivity Analysis, 6*, 247–268.
Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, J. E. (2005). *An Introduction to Efficiency and Productivity Analysis* (2nd ed.). Springer.
Coleman, J. S., Campbell, E., Hobson, C. J., McPartland, J., Mood, A., Weinfeld, F. D., & York, R. (1966). *Equality of Educational Opportunity* (US Government Printing Office, Washington, DC).
Colombi, R., Kumbhakar, S. C., Martini, G., & Vittadini, G. (2014). Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis, 42*(2), 123–136.
Comi, S. L., Argentin, G., Gui, M., Origo, F., & Pagani, L. (2017). Is it the way they use it? Teachers, ICT and student achievement. *Economics of Education Review, 56*, 24-39.
Cooper, W. W., Huang, Z., Lelas, V., Li, S. X., & Olesen, O. B. (1998). Chance constrained programming formulations for stochastic characterizations of efficiency and dominance in DEA. *Journal of productivity analysis, 9*(1), 53-79.
Cooper, W. W., Seiford, L. M., & Tone, K. (2006). *Introduction to Data Envelopment Analysis and Its Uses: With DEA- solver Software and References*. Springer.
Cordero, J. M., Santín, D., & Simancas, R. (2017). Assessing European primary school performance through a conditional nonparametric model. *Journal of the Operational Research Society, 68*(4), 364-376.
Cortes, K. E. (2006). The effects of age at arrival and enclave schools on the academic performance of immigrant children. *Economics of Education Review, 25*(2), 121-132.
Courtney, M., Karakus, M., Ersozlu, Z., & Nurumov, K. (2022). The influence of ICT use and related attitudes on students' math and science performance: multilevel analyses of the last decade's PISA surveys. *Large-scale Assessments in Education, 10*(1), 1-26.
Crespo-Cebada, E., Pedraja-Chaparro, F., & Santín, D. (2014). Does school ownership matter? An unbiased efficiency comparison for regions of Spain. *Journal of Productivity Analysis, 41*(1), 153-172.
Curi, C., Daraio, C., & Llerena, P. (2012). University technology transfer: how (in) efficient are French universities?. *Cambridge journal of economics, 36*(3), 629-654.
Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis, 24*(1), 93-121.
Daraio, C., Simar, L., & Wilson, P. W. (2018). Central limit theorems for conditional efficiency measures and tests of the 'separability'condition in non-parametric, two-stage models of production. *The Econometrics Journal, 21*(2), 170-191.
Daraio, C., Kerstens, K., Nepomuceno, T., & Sickles, R. C. (2020). Empirical surveys of frontier applications: a meta-review. *International Transactions in Operational Research*, *27*(2), 709-738.

Darling-Hammond, L. (2010). *The Flat World and Education: How America's Commitment to Equity will Determine Our Future*. New York, NY: Teachers College Press.

De Witte, K., & López-Torres, L. (2017). Efficiency in education: A review of literature and a way forward. *Journal of the Operational Research Society*, *68*(4), 339-363.

De Witte, K., & N. Rogge (2014). Does ICT matter for effectiveness and efficiency in mathematics education? *Computers & Education*, *75*, 173-184.

Deng, Z., & Gopinathan, S. (2016). PISA and high-performing education systems: Explaining Singapore's education success. *Comparative Education, 52*(4), 449-472.

Deutsch, J., Dumas, A., & Silber, J. (2013). Estimating an educational production function for five countries of Latin America on the basis of the PISA data. *Economics of Education Review, 36*, 245–262.

Deutsch, J., Dumas, A., & Silber, J. (2019). School externalities and scholastic performance: an efficiency analysis. *International Journal of Manpower, 40*(1), 102-119.

Dolton, P., Marcenaro, O. D., & Navarro, L. (2003). The effective use of student time: a stochastic frontier production function case study. *Economics of Education Review, 22*(6), 547-560.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*(1), 54–75.

European Commission (2013). *Survey of Schools: ICT in Education. Benchmarking Access, Use and Attitudes to Technology in Europe's Schools.* Luxembourg, Publications Office of the European Union.

Falck, O., Mang, C., & Woessmann, L. (2018). Virtually no effect? Different uses of classroom computers and their effect on student achievement. *Oxford Bulletin of Economics and Statistics, 80*(1), 1-38.

Făt, S., & Labăr, A. V. (2009). The Efficiency of ICT Usage for Education. *Evaluative Research Report*.

Felipe, J. (1998). On the interpretation of coefficients in multiplicative- logarithmic functions: A reconsideration. *Applied Economics Letters, 5*, 397–400.

Fernández-Gutiérrez, M., Gimenez, G., & Calero, J. (2020). Is the use of ICT in education leading to higher student outcomes? Analysis from the Spanish Autonomous Communities. *Computers & Education, 157*, 103969.

Ferrera, J. M. C., Cebada, E. C., Chaparro, F. P., & Santín, D. (2010). Factors affecting educational attainment: Evidence from Spanish PISA 2006 Results. *Regional and Sectoral Economic Studies, 10*(3), 55-76.

Ferrera, J. M. C., Cebada, E. C., Chaparro, F. P., & Santín, D. (2011). Exploring educational efficiency divergences across Spanish regions in PISA 2006. *Revista de economía aplicada, 19*(57), 117-145.

Filippini, M., & Greene, W. H. (2016). Persistent and transient productive inefficiency: A maximum simulated likelihood approach. *Journal of Productivity Analysis, 45*, 187–196.

Franta, M., & Konecny, T. (2009). Stochastic frontier analysis of the efficiency of czech grammar schools. *Czech Sociological Review, 45*(6), 1265-1282.

Fried, H., Lovell, C. A. K., & Schmidt, S. (Eds.) (2008). *The Measurement of Productive Efficiency and Productivity Growth*. University Press, UK: Oxford.

Frýd, L., & Sokol, O. (2021). Relationships between technical efficiency and subsidies for Czech farms: A two-stage robust approach. *Socio-Economic Planning Sciences, 78,* 101059.

Fu, J. (2013). Complexity of ICT in education: A critical literature review and its implications. *International Journal of Education and Development using Information and Communication Technology*, *9*(1), 112-125.

Gajewski, B. J., Lee, R., Bott, M., Piamjariyakul, U., & Taunton, R. L. (2009). On estimating the distribution of data envelopment analysis efficiency scores: an application to nursing homes' care planning process. Journal of Applied Statistics, 36(9), 933-944.

Garcia-Diaz, R., del Castillo, E., & Cabral, R. (2016). School competition and efficiency in elementary schools in Mexico. *International Journal of Educational Development, 46*, 23-34.

Gaviria-Marin, M., Merigo, J. M., & Popa, S. (2018). Twenty years of the Journal of Knowledge Management: A bibliometric analysis. *Journal of Knowledge Management*, *22*, 1655-1687.

Gerami, J., Sivandzadeh, F., & Manzari, H. (2014). Performance assessment and ranking of Shiraz high schools using DEA. *Advances in Environmental Biology, 8*(12), 138-146.
Gil-Flores, J., Rodríguez-Santero, J., & Torres-Gordillo, J. J. (2017). Factors that explain the use of ICT in secondary-education classrooms: The role of teacher characteristics and school infrastructure. *Computers in Human Behavior, 68*, 441-449.
Gimenez, G., & Vargas-Montoya, L. (2021). ICT use and successful learning: The role of the stock of human capital. *Mathematics, 9*(14), 1648.
Giménez, V., Prior, D., & Thieme, C. (2007). Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison. *Journal of the Operational Research Society, 58*(8), 996-1007.
Glewwe, P. (2016). What explains Vietnam's exceptional performance relative to other countries, and what explains gaps within Vietnam, on the 2012 PISA assessment?. *VNU Journal of Science: Social Sciences and Humanities, 32*, 138-148.
Gómez-Fernández, N., & Mediavilla, M. (2021). Exploring the relationship between Information and Communication Technologies (ICT) and academic performance: A multilevel analysis for Spain. *Socio-Economic Planning Sciences, 77*, 101009.
Gralka, S. (2018). Persistent inefficiency in the higher education sector: evidence from Germany, *Education Economics, 26*(4), 373-392.
Greene, W. H. (1980). On the estimation of a flexible frontier production model. *Journal of Econometrics, 13(*1), 101-115.
Greene, W. H. (2003). Simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis, 19*(2), 179-190.
Grimes, D. & Warschauer, M. (2008). Learning with laptops: a multi-method case study. *Journal of Educational Computing Research*, *38*(3), 305-32.
Gronewold, N. (2019). *Booming Southeast Asia's Dirty Secret: Coal*. E&E News. Available from: https://subscriber.politicopro.com/article/eenews/1061593609.
Grosskopf, S., Hayes, K. J., & Taylor, L. L. (2014). Efficiency in education: Research and implications. *Applied Economic Perspectives and Policy, 36*(2), 175-210.
Guan W. (2003). From the help desk: Bootstrapped standard errors. *Stata Journal*, *3*(1), 71–80.
Guarini, G., Laureti, T., & Garofalo, G. (2020). Socio-institutional determinants of educational resource efficiency according to the capability approach: An endogenous stochastic frontier analysis. *Socio-Economic Planning Sciences, 71*, 100835.
Haller, E. J. (1992). High school size and student indiscipline: Another aspect of the school consolidation issue? *Educational Evaluation and Policy Analysis*, *14*, 145–156.
Hamid, R. (2000). Education in Brunei Darussalam. *Journal of Southeast Asian Education*, *1*(1), 21-51.
Hanushek, E. A. (1979). Conceptual and empirical issues in the estimation of educational production functions. *Journal of Human Resources*, *14*(3), 351-388.
Hanushek, E. A. (2003). The failure of input-based schooling policies. *The economic journal, 113*(485), F64-F98.
Hanushek, E. A. & Luque, J. (2003). Efficiency and equity in schools around the world. *Economics of Education Review, 22*, 481-502.
Hanushek, E.A., Rivkin, S.G., & Taylor, L.L. (1996). Aggregation and the estimated effects of school resources. *The Review of Economics and Statistics*, *78*(4), 611-627.
Hu, W., Dong, J., Hwang, B., Ren, R., & Chen, Z. (2019). A scientometrics review on city logistics literature: Research trends, advanced theory and practice. *Sustainability*, *11,* 2724.
Huang, T., Ken, Y., Wang, W. C., Wu, C. H., & Shiu, S. H. (2011). Assessing the relative performance of US university technology transfer: non-parametric evidence. *Wseas Transactions on Business and Economics, 8*(3), 79-109.
Ibrahim, M. D., Alola, A. A., & Ferreira, D. C. (2021). A two-stage data envelopment analysis of efficiency of social-ecological systems: Inference from the sub-Saharan African countries. *Ecological Indicators, 123*, 107381.

Iliyasu, A., & Mohamed, Z. A. (2016). Evaluating contextual factors affecting the technical efficiency of freshwater pond culture systems in Peninsular Malaysia: A two-stage DEA approach. *Aquaculture Reports, 3,* 12-17.

Jin, R., Yuan, H., & Chen, Q. (2019). Science mapping approach to assisting the review of construction and demolition waste management research published between 2009 and 2018. *Resources, Conservation and Recycling*, *140*, 175-188.

Johnes, G., & Johnes, J. (2009). Higher education institutions' costs and efficiency: Taking the decomposition a further step. *Economics of Education Review, 28*(1), 107–113.

Johnes, G., & Virmani, S. (2020). The efficiency of private and public schools in urban and rural areas: Moving beyond the development goals. *International Transactions in Operational Research, 27*(4), 1869-1885.

Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of education review, 25*(3), 273-288.

Johnes, J. (2008). Efficiency and productivity change in the English higher education sector from 1996/97 to 2004/5. *The Manchester School, 76*(6), 653-674.

Johnes, J. (2013). Efficiency and mergers in English higher education 1996/97 to 2008/9: Parametric and non-parametric estimation of the multi-input multi-output distance function. *The Manchester School, 82*(4), 465-487.

Jondrow, J., Lovell, C. K., Materov, I. S., & Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics, 19*(2-3), 233-238.

Kim, Y. (2013). The ivory tower approach to entrepreneurial linkage: productivity changes in university technology transfer. *The Journal of Technology Transfer, 38*(2), 180-197.

Kirjavainen, T. (2012). Efficiency of Finnish general upper secondary schools: An application of stochastic frontier analysis with panel data. *Education Economics, 20*(4), 343–364.

Kodde, D. A., & Palm, F. C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, *54*, 1243–1248.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica, 46*(1), 33-50.

Kozma R. B. (2008). Comparative analysis of policies for ICT in education. In J. Voogt, & G. Knezek (Eds.), *International Handbook on Information Technology in Primary and Secondary Education*. Springer.

Krueger, A. B. (1999). Measuring labor's share. *American Economic Review, 89*(2), 45-51.

Kumbhakar, S. C., & Heshmati, A. (1995). Efficiency measurement in Swedish dairy farms: an application of rotating panel data, 1976–88. *American Journal of Agricultural Economics, 77*(3), 660-674.

Kumbhakar, S. C., & Lovell, C. K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press.

Kumbhakar, S. C., Lien, G., & Hardaker, J. B. (2014). Technical efficiency in competing panel data models: A study of Norwegian grain farming. *Journal of Productivity Analysis, 41*(2), 321–337.

Kumbhakar, S. C., Wang, H., & Horncastle, A. P. (2015). *A Practitioner's Guide to Stochastic Frontier Analysis using Stata*. Cambridge University Press.

Lai, H. -P., & Kumbhakar, S. C. (2018). Panel data stochastic frontier model with determinants of persistent and transient inefficiency. *European Journal of Operational Research, 271*, 746-755.

Lavado, R. F., & Cabanda, E. C. (2009). The efficiency of health and education expenditures in the Philippines. *Central European Journal of Operations Research, 17*(3), 275-291.

Le, M. H., Afsharian, M., & Ahn, H. (2021). Inverse frontier-based benchmarking for investigating the efficiency and achieving the targets in the Vietnamese education system. *Omega, 103*, 102427.

Le, T. D., Ngo, T., Ho, T. H., & Nguyen, D. T. (2022). ICT as a Key Determinant of Efficiency: A Bootstrap-Censored Quantile Regression (BCQR) Analysis for Vietnamese Banks. *International Journal of Financial Studies, 10*(2), 44.

Lee, K. (2009). Do charter schools spur improved efficiency in traditional public schools in Michigan?. *KEDI Journal of Educational Policy, 6*(1), 41-59.

Lee, Y.-J. (2014). Science Education in a Straightjacket: The Interplay of People, Policies, and Place in an East Asian Developmental State. In *Inquiry into the Singapore Science Classroom*, A.-L. Tan, C.-L. Poon, & S. L. Lim (Eds.), pp. 165–189. Springer: Singapore.

Lei, J., & Zhao, Y. (2008). One-to-one computing: what does it bring to schools? *Journal of Educational Computing Research*, *39*(2), 97-122.

Lim, C. P., Ra, S., Chin, B., & Wang, T. (2020). Leveraging information and communication technologies (ICT) to enhance education equity, quality, and efficiency: case studies of Bangladesh and Nepal. *Educational Media International, 57*(2), 87–111.

Lima, G. D. S. (2017). *Efficiency in school education: A semi-parametric study of school efficiency in OECD countries* [Master's thesis, ISCTE Business School].

Link, A. N., & Siegel, D. S. (2005). Generating science-based growth: an econometric analysis of the impact of organizational incentives on university–industry technology transfer. *European Journal of Finance, 11*(3), 169-181.

Luke, A., P. Freebody, S. Lau, & S. Gopinathan (2005). Towards research-based innovation and reform: Singapore schooling in transition. *Asia Pacific Journal of Education, 25*(1), 5–28.

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.

Martínez-López, F. J., Merigó, J. M., Valenzuela-Fernández, L., & Nicolás, C. (2018). Fifty years of the European Journal of Marketing: a bibliometric analysis. *European Journal of Marketing*, *52*(1-2), 439-468.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, *18*, 435–444.

Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, *36*(3), 363–377.

Ministry of Education Malaysia (2008). *Education in Malaysia: A Journey to Excellence*. Educational Planning and Research Division, Ministry of Education Malaysia.

Mizala, A., Romaguera, P., & Farren, D. (2002). The technical efficiency of schools in Chile. *Applied Economics, 34*(12), 1533-1552.

Mobi, I. M., Onyenanu, I. U., & Ikwueto, O. C. (2015). A study of the negative influences of ICT on secondary school students in Nigeria. *American Academic & Scholarly Research Journal, 7*(5), 136-142.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PloS Medicine*, *6*(7), e1000097.

Mongan, C. J., Santín, D., & Valino, A. (2011). Towards the equality of educational opportunity in the province of Buenos Aires. *Journal of Policy Modelling, 33*(4), 583–596.

Mora, T., Escardíbul, J. O., & Di Pietro, G. (2018). Computers and students' achievement: An analysis of the One Laptop per Child program in Catalonia. *International Journal of Educational Research*, *92*, 145-157.

Mora, T., Escardíbul, J. O., & Espasa, M. (2010). The effects of regional educational policies on school failure in Spain. *Revista de Economía Aplicada, 18*(54), 79-106.

Moutinho, V., Madaleno, M., & Robaina, M. (2017). The economic and environmental efficiency assessment in EU cross-country: Evidence from DEA and quantile regression approach. *Ecological Indicators, 78*, 85-97.

Nwaogbe, O. R., Wanke, P., Ogwude, I. C., Barros, C. P., & Azad, A. K. (2018). Efficiency driver in Nigerian airports: A bootstrap DEA–censored quantile regression approach. *Journal of Aviation Technology and Engineering, 7*(2), 2.

OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*, PISA, OECD Publishing, Paris.

OECD (2019a), *PISA 2018: Insights and Interpretations*, OECD Publishing.

OECD (2019b), *PISA 2018 Results (Volume I): What Students Know and Can Do*, OECD Publishing.

Office of the Education Council (2004). *Education in Thailand 2004*. Bangkok: Amarin Printing and Publishing.

Olesen, O. B., & Petersen, N. C. (2016). Stochastic data envelopment analysis—A review. *European Journal of Operational Research*, *251*, 2-21.

Pastor, J. T., Ruiz, J. L., & Sirvent, I. (2002). A statistical test for nested radial DEA models. *Operations Research, 50*(4), 728–735.

Perbawaningsih, Y. (2013). Plus minus of ICT usage in higher education students. *Procedia-Social and Behavioral Sciences, 103*, 717-724.

Perelman, S., & Santín, D. (2011a). Measuring educational efficiency at student level with parametric stochastic distance functions: An application to Spanish PISA results. *Education Economics, 19*(1), 29–49.

Perelman, S., & Santín, D. (2011b). Imposing monotonicity on outputs in parametric distance function estimations. *Applied Economics, 43*(30), 4651–4661.

Portela, M. C., & Camanho, A. S. (2007). Performance assessment of Portuguese secondary schools. *Working Paper in Economics, No. 07/2007*.

Primont, D. F., & Domazlicky, B. (2006). Student achievement and efficiency in Missouri schools and the No Child Left Behind Act. *Economics of Education Review, 25*(1), 77-90.

Purwadi, A., & and Muljoatmodjo, S. (2000). Education in Indonesia: Coping with the challenges in the third millennium. *Journal of Southeast Asian Education*, *1*(1), 79–112.

Qi, S., Peng, H., Zhang, X., & Tan, X. (2019). Is energy efficiency of Belt and Road Initiative countries catching up or falling behind? Evidence from a panel quantile regression approach. *Applied Energy*, *253*, 113581.

Rhaiem, M. (2017). Measurement and determinants of academic research efficiency: a systematic review of the evidence. *Scientometrics, 110*(2), 581–615.

Rodgers, T., & Ghosh, D. (2001). Measuring the determinants of quality in UK higher education: a multinomial logit approach. *Quality Assurance in Education, 9*(3), 121-126.

Rudd, E. (1984). A comparison between the results achieved by women and men studying for first degrees in British universities. *Studies in Higher Education, 9*(1), 47-57.

Ruth, A., Shizhu, L., Caesar, A. E., & Moping, F. (2019). Measuring efficiency and effectiveness of workers' performance and information flow system in Technical Universities in Ghana, *American International Journal of Business Management*, *2*(6), 29-38.

Salas-Velasco, M. (2020). Assessing the performance of Spanish secondary education institutions: Distinguishing between transient and persistent inefficiency, separated from heterogeneity. *The Manchester School*, *88*(4), 531-555.

Salcedo, R. E. (2020). Performance efficiency of the teacher education programs of a state university in the Philippines: A data envelopment analysis study. *Journal of Critical Review, 7*(7), 96-103.

Sangrà, A., & González-Sanmamed, M. (2010). The role of information and communication technologies in improving teaching and learning processes in primary and secondary schools. *Research in Learning Technology*, *18*(3), 207-220.

Santín, D., & Sicilia, G. (2017). Dealing with endogeneity in data envelopment analysis applications. *Expert Systems with Applications, 68*, 173-184.

Santín, D., & Sicilia, G. (2018). Using DEA for measuring teachers' performance and the impact on students' outcomes: evidence for Spain. *Journal of Productivity Analysis, 49*(1), 1-15.

Schnepf, S. V. (2008). Inequality of learning amongst immigrant children in industrialised countries. *HWWI Research Paper*, No. 1-12.

Scippacercola, S., & D'Ambra, L. (2014). Estimating the relative efficiency of secondary schools by stochastic frontier analysis. *Procedia Economics and Finance, 17*, 79-88.

SEAMEO Secretariat (2001). *Workshop on SEAMEO's Role in the 21st Century*, Mallaca, Malaysia.

Shahini, A. (2021). Inequalities in Albanian education: Evidence from large-scale assessment studies. *Kultura i Edukacja, 4*(134), 40-70.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics, 136*(1), 31-64.

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417–453.

Smith, J. & Naylor, R. (2001). Determinants of degree performance in UK universities: a statistical analysis of the 1993 student cohort. *Oxford Bulletin of Economics and Statistics, 63*(1), 29-60.

Sosin, K., Lecha, B. J., Agarwal, R., Bartlett, R. L., & Daniel, J. I. (2004). Efficiency in the use of technology in economic education: Some preliminary results. *American Economic Review, 94*(2), 253-258.

Sowlati, T., & Paradi, J. C. (2004). Establishing the "practical frontier" in data envelopment analysis. *Omega, 32*(4), 261-272.

Spiezia, V. (2010). Does computer use increase educational achievements? Student-level evidence from PISA. *OECD Journal: Economic Studies, 2010*(1), 1-22.

Srijamdee, K., & Pholphirul, P. (2020). Does ICT familiarity always help promote educational outcomes? Empirical evidence from PISA-Thailand. *Education and Information Technologies, 25*(4), 2933-2970.

Stevenson, R. E. (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics, 13*(1), 57-66.

Sultan, W. I., & Crispim, J. (2018). Measuring the efficiency of Palestinian public hospitals during 2010–2015: An application of a two-stage DEA method. *BMC Health Services Research, 18*(1), 1-17.

Sutopo, W., Astuti, R. W., & Suryandari, R. T. (2019). Accelerating a technology commercialization; with a discussion on the relation between technology transfer efficiency and open innovation. *Journal of Open Innovation: Technology, Market, and Complexity, 5*(4), 95.

Symaco, L. P., & Chao, R. Y. (2019). Comparative and International Education in East and South East Asia. In C. C. Wolhuter & A. W. Wiseman (Eds.), *Comparative and International Education: Survey of an Infinite Field* (Vol. 36, pp. 213-228). Emerald Publishing Limited.

Thieme, C., Giménez, V., & Prior, D. (2012). A comparative analysis of the efficiency of national education systems. *Asia Pacific Education Review, 13*(1), 1-15.

Tinio, V. L. (2003). *ICT in Education: An E-primer*. United Nations Development Programme.

Titus, M. A. (2020). Examining degree production and financial context at public master's colleges and universities in the United States: A distance function approach. *Tertiary Education and Management, 26*(2), 215-231.

Titus, M. A., Vamosiu, A., & McClure, K. R. (2017). Are public master's institutions cost efficient? A stochastic frontier and spatial analysis. *Research in Higher Education, 58*(5), 469-496.

Tone, K. (1993). An ε-free DEA and a new measure of efficiency. *Journal of the Operations Research Society of Japan*, *36*, 167-174.

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research, 130*, 498–509.

Tone, K. (2002). A slacks-based measure of super-efficiency in data envelopment analysis, *European Journal of Operational Research, 143*, 32-41.

Tsionas, E. G., & Kumbhakar, S. C. (2014). Firm heterogeneity, persistent and tran- sient technical inefficiency: A generalized true random-effects model. *Journal of Applied Econometrics, 29*, 110–132.

Türkan, S., & Özel, G. (2017). Efficiency of state universities in Turkey during the 2014–2015 academic year and determination of factors affecting efficiency. *Education and Science, 42*(191), 307-322.

Ulkhaq, M. M. (2021). Efficiency analysis of Indonesian schools: A stochastic frontier analysis using OECD PISA 2018 data. in *Proceedings of the Second Asia Pacific International Conference on Industrial Engineering and Operations Management*, Surakarta, Indonesia, September 14-16.

van Eck, N. J., & Waltman, L. (2007). Bibliometric mapping of the computational intelligence field. International Journal of Uncertainty, *Fuzziness and Knowledge-Based Systems*, *15*(5), 625–645.

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523-538.

van Eck, N. J., Waltman, L., Van den Berg, J., & Kaymak, U. (2006). Visualizing the computational intelligence field. *IEEE Computational Intelligence Magazine*, *1*(4), 6–10.

Venable, J.R., Pries-Heje, J., Bunker, D., & Russo, N.L. (2011). Design and diffusion of systems for human benefit: toward more humanistic realization of information systems in society. *Information Technology & People*, *24*(3), 208-16.

Việt Nam News (2019, December 6). VN gets high scores but not named in PISA 2018 ranking.

Vinayak H. V., Thompson, F., & Tonby O. (2014). *Understanding ASEAN: Seven Things you Need to Know*. McKinsey & Company. Available from: https://www.mckinsey.com/~/media/McKinsey/Industries/Public%20Sector/Our%20Insights/Understanding%20ASEAN%20Seven%20things%20you%20need%20to%20know/Understanding%20ASEAN%20Seven%20things%20you%20need%20to%20know.pdf

Wang, H. J. (2002). Heteroscedasticity and non-monotonic efficiency effects of a stochastic frontier model. *Journal of Productivity Analysis, 18*(3), 241-253.

Wang, H. J., & Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis, 18*(2), 129-144.

Wheeler, S., Waite, S. J., & Bromfield, C. (2002). Promoting creative thinking through the use of ICT. *Journal of Computer Assisted Learning, 18*(3), 367-378.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for information Science, 32*(3), 163-171.

Willms, J. D., & Tramonte, L. (2019). The measurement and use of socioeconomic status in educational research. In L. E. Suter, B. Denman, & E. Smith (Eds.), *The SAGE Handbook of Comparative Studies in Education*. London: Sage.

Woetzel, J., Madgavkar, A., Seong, J., Manyika, J., Sneader, K., Tonby, O., Cadena, A., Gupta R., Leke, A., Kim, H., Gupta, S. (2018). *Outperformers: High-Growth Emerging Economies and the Companies that Propel them*. McKinsey Global Institute.

Wohlrabe, K., de Moya Anegon, F., & Bornmann, L. (2019). How efficiently do elite US universities produce highly cited papers?. *Publications, 7*(1), 4.

World Bank (2010). *Emerging Stronger from the Crisis, World Bank East Asia and Pacific Economic Update*, *Vol. 1*. Washington, DC: The International Bank for Reconstruction and Development/The World Bank.

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics, 9*(3), 245–268.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation, 31*(2-3), 114-128.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039.

Zhang, D., & Liu, L. (2016). How does ICT use influence students' achievements in math and science over time? Evidence from PISA 2000 to 2012. *Eurasia Journal of Mathematics, Science and Technology Education, 12*(9), 2431-2449.

Zhang, S., & Wang, X. (2022). Does innovative city construction improve the industry–university–research knowledge flow in urban China?. *Technological Forecasting and Social Change, 174*, 121200.

Zhang, S., Wang, X., & Zhang, B. (2022). Innovation ability of universities and the efficiency of university–industry knowledge flow: the moderating effect of provincial innovative agglomeration. *Chinese Management Studies, 16*(2), 446-465.

Zoghbi, A. C., Rocha, F., & Mattos, E. (2013). Education production efficiency: Evidence from Brazilian universities. *Economic Modelling, 31*, 94-103.

Zou, L., & Zhu, Y. W. (2021). Universities' Scientific and Technological Transformation in China: Its Efficiency and Influencing Factors in the Yangtze River Economic Belt. *Plos one, 16*(12), e0261343.

APPENDIX

JUNIOR RESEARCHERS' SECTION - Statistica & Applicazioni - Vol. XVIII, n. 2, 2020, pp. 197-220

CLUSTERING COUNTRIES ACCORDING TO THE WORLD HAPPINESS REPORT

M. Mujiya Ulkhaq*

SUMMARY

The World Happiness Report (WHR) has drawn international attention since the first initiative in 2012 as it can help the policy makers to evaluate their policy options. There are six factors to describe the variation of the happiness across the countries, i.e., gross domestic product per capita, social support, healthy life expectancy, freedom to make life choices, perception of corruption, and generosity. This study aims to cluster the countries according to the WHR 2020. Nine clustering algorithms (k-means, k-means++, k-medoids, clustering large applications, affinity propagation, spectral clustering, density-based spatial clustering of applications with noise, agglomerative nesting, and divisive analysis) are presented and three internal validation indices (silhouette index, Dunn's index, and Calinski-Harabasz' index) are utilized to compare the algorithms. This study is expected to give an insight about how to implement clustering algorithms into the real world (not artificial) data set and how to interpret the result.

Keywords: Clustering Algorithm, Cluster Validation, World Happiness Report.
DOI: 10.26350/999999_000036
ISSN: 18246672 (print) 2283-6659 (digital)

1. INTRODUCTION

The term *happiness*, referring to the experience of joy, contentment, or positive well-being, combined with a sense that one's life is good, meaningful, and worthwhile (Lyubomirsky, 2008), is progressively a common subject in cross-national study. It regards as an appropriate measurement for social growth and public policy goal. The first comparative investigation on happiness was arguably conducted by Cantril (1965), in which fourteen countries as representative samples have been participated in the study. Since then, happiness has been embraced in several international survey programs, for instance, the World Value Survey, the Euro-barometer, the European Welfare Survey, and the World Happiness Report (WHR).

The WHR is a landmark survey of the state of global happiness that ranks countries around the globe by how happy their citizens perceive themselves to be. The report was written by a group of independent experts acting in their personal capacities. The initiative began in 2012 as the motivation is to pursue policies to increase

* Dipartimento di Economia e Management - Università di Brescia - Via San Faustino 74/B - 25122 BRESCIA (e-mail: m.ulkhaq@unibs.it).

the public's happiness as much as it does to raise the public's national income. Since then, a series of the reports continues to obtain global recognition as governments, organizations, and civil society increasingly use happiness indicators to inform their policy-making decisions.

To explain happiness, the WHR 2020 (Helliwell, Layard, Sachs and De Neve, 2020) presented six factors, i.e., gross domestic product (GDP) per capita, social support, healthy life expectancy (HLE), freedom to make life choices, perception of corruption, and generosity (see Section 2.1 for more elaboration). The report revealed that Nordic countries are among the top happiest countries. Specifically, Finland is on the top list with the total score of 7.809; followed by Denmark, Iceland, and Norway, in the second, fourth, and fifth places, respectively, while Sweden is the seventh. Contrarily, the happiness score in the top 10 countries is more than twice as high as in the bottom 10. The later mostly suffered some combination of economic, political, and social stresses, such as South Sudan, Yemen, Afghanistan, and Central African Republic.

This research aims to cluster countries according to the WHR 2020. Clustering is a process of classifying objects, observations, or data which have feature(s) into groups (or clusters). Clustering has been addressed in many contexts and by researchers in many disciplines, such as in biology (e.g., Kapourani and Sanguinetti, 2019; Wang, Li, Deng and Pan, 2010), marketing (e.g., Minako, Ulkhaq, 'Sa Nu, Pratiwi and Akshinta, 2019; Ulkhaq, Fidiyanti, Adyatama, Maulani and Nugroho, 2019; Utami, Ginanjar, Fadlia, Lubis and Ulkhaq, 2019), psychology (e.g., Brusco, Steinley, Stevens and Cradit, 2019; van Lettow, Vermunt, de Vries, Burdorf and van Empelen, 2013), image processing (e.g., Cai, He, Li, Ma and Wen, 2004; John and Ramesh, 2017), and pattern recognition (e.g., Unglert, Radiæ and Jellinek, 2016). Many different types of clustering algorithms have been proposed in the literature. In this research, nine clustering algorithms are presented and then compared to look for ''the best'' way to partition the countries. The WHR 2020 will be the basis information to perform clustering.

This research employed R, a programming language for statistical computing and graphics. It is motivated by the recognition of R in the field of statistics, data mining, and machine learning; and also, by the aid of its well-established clustering packages. This study is also intended to assist researchers who have programming skills in R language but have little experience in clustering data.

The paper is structured as follows. In the following section, the data used in this research is presented as well as the procedure of data cleansing and imputation. A brief overview of the clustering process is described in Section 3. Next, the results of each clustering algorithm and the performance evaluation to compare the algorithms are demonstrated. Finally, conclusion and future research direction are presented in the last section.

2. DATA

2.1 *Variables Used*

The data set used in this research is adopted from the online data collection reported on the website of the WHR (https://worldhappiness.report). The data set contains two descriptors, one response variable, six predictors of the response variables, and several additional variables that were either calculated or gathered from external sources. The descriptors are: country (`country`) as the name of the surveyed country and year (`year`) as the year of data collection (from 2005 to 2019). The response variable is happiness score or subjective well-being (SWB) (`life_ladder`). The information was collected from the February 28, 2020 release of the survey of Gallup World Poll (GWP) covering years from 2005 to 2019. Unless stated otherwise, it is the national average response of the following question (called the Cantril ladder question): ‘‘Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?’’

There are six predictors of the response variable. The first is GDP per capita in purchasing power parity (PPP) at constant 2011 international dollar prices, normalized by taking its natural logarithm (`log_gdp`). The data were from the November 28, 2019 update of the World Development Indicators (WDI) released by the World Bank. The second is HLE at birth (`hle`). The data were extracted from the World Health Organization’s (WHO) Global Health Observatory data repository. The third predictor is social support (`social_support`). The data which is also from GWP survey is the national average of the binary responses (either 0 or 1) to the question ‘‘If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?’’ The next predictor is freedom to make life choices (`freedom`). It is the national average of responses to the GWP question ‘‘Are you satisfied or dissatisfied with your freedom to choose what you do with your life?’’ The fifth is generosity (`generosity`). It is the residual of regressing national average of response to the GWP question ‘‘Have you donated money to a charity in the past month?’’ on GDP per capita. The last determinant is corruption perception (`corruption`). It is the national average of the responses to two GWP questions ‘‘Is corruption widespread throughout the government or not’’ and ‘‘is corruption widespread within businesses or not?’’ The overall perception is the average of the two 0-or-1 responses.

The other variables are omitted since only the six predictors are used to describe the happiness.

2.2 *Data Cleansing and Imputation*

Since the WHR 2020 stated that only the average value of `life_ladder` from 2017 to 2019 were used, so that only this particular range of years was included in this research. However, several territories/countries have no information in one or more of the predictor variables over that survey period, but the information is available in earlier years; for example, they may have GDP statistics in 2015 but not in the period from 2017 to 2019. In this case, the most recent information was used as if they are the 2017-2019 information. Three years limit for how far back in looking for the missing values was applied.

A few territories/countries do not have data on HLE, for instance, for Hong Kong, this information was obtained by calculating the health life-to-life expectancy ratio using estimates reported in Law and Yip (2003). The same ratio information for Swaziland in the period 2005-2010 can be found in Salomon, Wang, Freeman, Vos, Flaxman, Lopez and Murray (2012). The ratios for Hong Kong and Swaziland, respectively, were multiplied with their life expectancy time series in the WDI to get HLE up to 2017; the time series was then extrapolated to 2019. Salomon *et al.* (2012) also provided information for Taiwan and the Palestinian Territories, but the WDI does not provide HLE data for these two regions. In this case, their 2010 HLE data as if they are the 2017-2019 value were used. For Kosovo, its time series of life expectancy (available in the WDI) was adjusted to a time series of HLE by assuming that its health life-to-life expectancy ratio equals to the world average.

Finally, the statistics of Cyprus were used as information for Northern Cyprus' missing values of GDP per capita and HLE.

3. CLUSTERING: AN OVERVIEW

Clustering is regarded as one of the most useful methods in machine learning and data mining for finding the existence of groups (called clusters) as well as investigating interesting patterns in the data set. It is about dividing, or separating, or partitioning the data set into clusters. The general objective is that the objects or data points or observations in a cluster are closer to (or more similar) each other than to other data points in different partition(s) or cluster(s) (Manly and Alberto, 2017). Clustering analysis can be applied in many disciplines, such as (but not limited to) psychology, life sciences, marketing, engineering, and medical sciences. It might be found under different terms, for instance, typology (in social sciences), numerical taxonomy (in biology, ecology), unsupervised learning (in pattern recognition), and partition (in graph theory) (Theodoridis and Koutroumbas, 2008). There are no predefined groups or classes (called the ground truth label) in the clustering analysis that show what kind of associations or relations among the data. For this reason, clustering analysis is also called as unsupervised learning. Classification is a counterpart of clustering analysis as the predefined classes or categories are available (it is also called supervised learning).

The basic steps in clustering can be summarized as follows (see also Figure 1):

1. *Data cleansing and imputation*
   Real-world databases often contain errors (trivial or non-trivial, syntactic or semantic) and missing values. Data pre-processing might be necessary to ensure the information is consistent, accurate, and high-quality prior to their utilization in clustering analysis. Refer to the previous section to recall the process of data cleansing and imputation in this research.

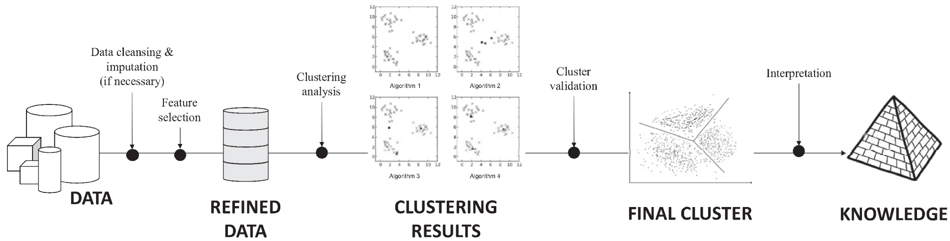
2. *Feature selection*
This step aims to choose proper features on which clustering analysis is to be conducted. Some works of literature relate this step with dimensionality reduction when it deals with high-dimensional data. Principle component analysis is typically used. It deals with constructing a linear combination of a set of vectors which could explain the data variance. However, since there are only plenty of features used in this research, this approach is not applied. Moreover, the result of the clustering analysis could be different with and without the dimensionality reduction (Manly and Alberto, 2017); also, computational time is not a vital issue in this research. Rather, in this research, this step was performed by utilizing the multiple linear regression analysis (see Section 4.2). The predictors which significantly explain the response variable are then used as features for the next step.

3. *Clustering analysis*
It refers to the choice of clustering algorithms. Several clustering algorithms have been proposed by scholars. Obviously, it is not possible to present and review all the algorithms; instead, in the following section, only algorithms used in this study will be presented.

4. *Cluster validation*
Once clusters have been obtained by performing a clustering algorithm, such question could arise: ''How well does the obtained clusters fit the data set?'' The question is essential since several different clustering algorithms (or different configurations of similar clustering algorithm) could generate different clusters (Pal and Biswas, 1997); thus, one could analyse different clustering algorithms and choose the algorithm that best fits the data.

5. *Interpretation*
In several cases, experts and professionals in the field of application somehow have to integrate the result obtained from clustering algorithm with other analysis or experimental evidence to draw correct conclusion as well as gain insightful knowledge.

FIGURE 1. - *Steps of Clustering Process*



3.1 *Clustering Algorithms*

Many different types of clustering algorithms have been proposed in the literature. In this subsection, nine clustering algorithms used in this study are presented. The algorithms were selected according to the type of the data used, the objective of the algorithms, as well as to represent each type of the clustering algorithm. Following is a brief explanation for each algorithm used in this research.

1. *k-means*
   k-means (MacQueen, 1967) is arguably the most broadly clustering algorithm used in literature due to the computational speed and its simplicity. It requires distance matrix and number of clusters *k*. Initially, each object or observation is connected with one cluster according to its distance to centroid or cluster centre. The objective of this algorithm is minimizing the average squared distance between observations in the same cluster. The predefined number of clusters is one of the main limitations of this algorithm since the final clusters depend on the choice of the number of clusters. Moreover, *k*-means is considered as sensitive to the initial seed selection (Jain, Murty and Flynn, 1999). The algorithm of *k*-means works as follows:
   (i) Select *k* cluster centres (or centroids) to coincide with *k* arbitrarily defined observations.
   (ii) Assign each observation to the closest centroids.
   (iii) Recompute the centroids using the current cluster memberships.
   (iv) When a convergence criterion cannot be fulfilled, go back to step (ii). The convergence criterion is minimal (or even no) reassignment of observations to the latest centroids, or minimal reduction in squared error.

2. *k-means++*
   As previously stated, one of the drawbacks of *k*-means is that the algorithm is sensitive to the initialization of the centroids. In sum, a poor initialization could result in a poor clustering. To overcome the drawback, *k*-means++ was proposed by Arthur and Vassilvitskii (2007). This algorithm guarantees smarter initialization and improves the clustering quality. Apart from the initialization, the rest of the algorithm is similar to the standard k-means. The algorithm of k-means++ works as follows:
   (i) Choose the first cluster centre randomly.
   (ii) For each observation, calculate the distance from the nearest cluster centre.
   (iii) Choose the next cluster centre from the observations such that the probability of choosing a point as cluster centre is proportionate to the distance from the closest, formerly chosen cluster centre (the point that has maximum distance from the closest cluster centre is expected to be chosen as the next cluster centre).
   (iv) Reiterate steps (ii) and (iii) until *k* cluster centres have been sampled.

3. *k-medoids*
   This algorithm is also called as Partitioning Around Medoid (PAM). It was

proposed by Kaufman and Rousseeuw (1990). A medoid can be defined as the representative of the objects in the cluster, whose dissimilarities with all the other objects in the cluster is minimum. It is considered as a less sensitive (or robust) alternative to *k*-means algorithm since *k*-medoids uses medoids as centroids as an alternative of means which is used in *k*-means. The algorithm of *k*-medoids works as follows:

(i) Choose *k* points to be the medoids.
(ii) Compute dissimilarity matrix (see Kaufman and Rousseeuw, 1990, for more explanation).
(iii) Assign each data point to the closest medoid.
(iv) For each partition, investigate if any of the point of the partition reduces the average dissimilarity coefficient. If the reduction occurs, choose the point as the medoid for this cluster. This point should reduce the coefficient the most.
(v) If at least one medoid has changed, go back to step (iii); otherwise, stop the process.

4. *Clustering large applications* (CLARA)
The algorithm by Kaufman and Rousseeuw (1986) is an extension to *k*-medoids which deals with huge data (having more than several thousand objects or data points). This extension aims to reduce storage problems and computational time. Instead of identifying all medoids for all data set, the algorithm considers only a small sample of the data with fixed size. Consequently, *k*-medoids algorithm is applied to look for an optimal number of medoids for the predefined sample. CLARA repeats the sampling and clustering processes a pre-specified number of times to minimize sampling bias. The algorithm of CLARA works as follows:
(i) From the data set, create several subsets (or samples) randomly with fixed size.
(ii) Perform *k*-medoids algorithm on each sample, then select k representative medoids. Assign each object to the nearest medoid.
(iii) Compute the mean of the dissimilarities of the data points to their nearest medoids.
(iv) If the mean is smaller than the mean obtained from the previous step, then these *k*-medoids are kept as the best *k*-medoids.

5. *Affinity propagation*
One of the main limitations of *k*-means algorithm and also other similar algorithms is that the number of clusters and the initial set of points have to be preliminary defined. Affinity propagation proposed by Frey and Dueck (2007), on the other hand, takes similarity between pairs of observations as input parameters, and considers all observations as potential ''exemplars''. Real-valued messages are swapped between observations. These messages would be updated in response to the values from other pairs. Iteratively, this updating would run until convergence, at which point the final exemplars are

selected, and thus, the final clusters are obtained. Each iteration in this algorithm contains two message-passing steps:

(i) Compute responsibilities $r(i,k)$. It reflects the accumulated evidence for how appropriate point k is to be the exemplar for point i, considering other exemplar candidates for point i. The responsibility is sent from point i to point k as the candidate exemplar.

(ii) Compute availability $a(i,k)$. It reflects the accumulated evidence for how proper point i to select point k as its exemplar, considering the support from other points that point k might be an exemplar. The availability is sent from point k as the candidate exemplar to point i.

The main drawback of this algorithm is its complexity, which makes this algorithm most suitable for small to medium sized data set.

6. *Spectral clustering*

Traditional clustering algorithms like k-means and k-means++ use an elliptical or spherical metric to group observations; thus, they would not perform well if the partitions are non-convex. Spectral clustering, on the other hand, can be considered as a generalization of traditional clustering algorithm which is intended for this kind of situation. The algorithm works as follows:

(i) Represent the data points as a similarity graph. Compute pairwise similarities s'_{ii} between each data point. The data points are then represented in an undirected similarity graph $G = \langle V, E, \rangle$ where the vertices V are the data points and the edges E are weighted by s'_{ii}.

(ii) Cluster the data points by partitioning G by its connected components. All symmetric nearest neighbours are connected with edges weighted with *sii*‘ and points that are not nearest neighbours are not connected.

(iii) Compute the graph Laplacian L.

(iv) Calculate the eigen-decomposition of L. Find the m eigenvectors $\mathbf{Z}_{N\times m}$ corresponding to the m smallest eigenvalues of L, ignoring the trivial constant eigenvector.

(v) Use a standard clustering algorithm to cluster the rows of $\mathbf{Z}$ (see Hastie, Tibshirani and Friedman, 2017, for more elaboration).

7. *Density-based spatial clustering of applications with noise* (DBSCAN)

This algorithm, which was proposed by Ester, Krieger, Sander and Xu (1996), views partitions as high-density areas separated by low density areas. It is one of the most well-known density-based clustering algorithms. The essential component of this algorithm is the concept of core samples, which are samples in high density areas. The algorithm works according to the idea of clusters and noise. For each point in a cluster, the neighbourhood of a given area must contain at least a minimum number of points. The algorithm works as follows:

(i) «Find all the neighbour points within ‘‘eps’’ (the neighbourhood around a data point) and identify the core points or visited with more than ‘‘MinPts’’ (minimum number of data points within ‘‘eps’’ radius) neighbours.

(ii) For each core point if it is not already assigned to a cluster, create a new cluster.
(iii) Find recursively all its density connected points and assign them to the same cluster as the core point. A point *a* and *b* are said to be density connected if there exist a point *c* which has a sufficient number of points in its neighbours and both the points *a* and *b* are within the ''eps'' distance. This is a chaining process; so, if *b* is neighbour of *c*, *c* is neighbour of *d*, *d* is neighbour of *e*, which in turn is neighbour of *a* implies that *b* is neighbour of *a*.
(iv) Iterate through the remaining unvisited points in the data set. Those points that do not belong to any cluster are noise.

8. *Agglomerative nesting* (AGNES)
Previously mentioned four algorithms belong to the class of *partitioning* clustering. It means that observations are classified into *k* clusters, in which each cluster has at least one observation and each observation belongs to exactly one cluster. In addition, two different clusters cannot have same observation(s) and the *k* clusters contain all the objects in the data set (Kaufman and Rousseeuw, 1990). AGNES (and also the following algorithm, DIANA) belongs to the class of *hierarchical* clustering which does not generate partitions or clusters. AGNES, in particular, starts by considering one observation as one partition or cluster. Pairs of partitions are combined sequentially until all partitions were merged into one big cluster that contains the entire set of observations. The result of this algorithm is a dendrogram; it is a tree-based representation of the hierarchical agglomerative process. The algorithm uses (dis)similarity between each pair of observations in the entire data set. Then, it uses linkage function to merge observations which are in close proximity to form the dendrogram. If one would create a partition, the cut-off point of the hierarchical tree should be determined.

9. *Divisive analysis* (DIANA)
DIANA is also the type of hierarchical clustering which is the inverse of AGNES. It starts by including all observations in one big cluster. Iteratively, the most heterogeneous pairs of observations would be separated into two subsets. This step is repeated until all observations are located at their own clusters. This algorithm poses computational problems: the first step involved considering all possible partitions into two subsets; this might be infeasible because of a huge number of combinations. Consequently, some scholars have restricted their attention to AGNES. (In the literature, DIANA has been largely ignored; as a matter of fact, when people discuss the hierarchical clustering, they often mean AGNES.)

3.2 *Clustering Performance Evaluation*

Clustering algorithms deal with several parameters; frequently they have to deal with noisy, incomplete and sampled data, as well as run in high dimensional spaces; hence, their performance could differ for different types of data in different applications. The method for evaluating the performance of clustering algorithms is called *cluster validation*; it regards as one of the most central concerns in clustering analysis (Halkidi, Batistakis and Vazirgiannis, 2001).

There are two criteria proposed for cluster validation, i.e., compactness (or cohesion) and separation. The former means that the member(s) of each partition should be as close as possible to each other; and the later implies that the partitions should be widely spaced. Validity measures used for assessing the performance of the algorithms with respect to those previous two criteria can be classified into relative, internal and external validation.

Relative validation assesses the clustering by changing different parameter values for the same algorithm (for instance, changing the number of clusters k). It is commonly used for investigating the optimal number of clusters. Internal validation is according to the information inherent to the data set and assesses the quality of the cluster algorithm without any external information. Conversely, the external validation measures the similarity between the clustering algorithm's result and the ''correct'' partitioning (or the ground truth label) of the data set. Since the ground truth label is unavailable (this study used the real data set, not artificial data set), only relative and internal validations were used here

In this study, the elbow method was used as a relative validation. It is performed by running the particular algorithm several times with a rising number of cluster k. Its sum of squared errors is then calculated and plotted against the number of clusters k. If the plot seems like an arm, then the ''elbow'' of the arm corresponds to the optimal number of clusters.

There are several internal validation indices in the literature, yet in this study, only three indices are used as follows, see Arbelaitz, Gurrutxaga, Muguerza, Pérez and Perona (2013) for more comprehensive discussion. Before, let us denote some notations used. First, let us define a data set X as a set of N observations characterized as vectors in an F-dimensional space: $X = \{x_1, x_2, \ldots, x_N\} \subseteq \mathcal{R}F$. A partition in X is a set of disjoint partitions (or clusters) that divides X into k clusters: $C = \{c_1, c_2, \ldots, c_k\}$. The centroid of cluster i c_i is the mean vector $\overline{c_i}$; and the centroid of the data set is the mean vector $\bar{X}$ of the whole data set. The Euclidean distance between objects x_l and x_p is denoted as $\mathrm{d_E}(x_l, x_p)$. the three validation indices used in this study are as follows.

1. *Silhouette index* (SI)
   SI is a normalized summation-type index which can be calculated as (Rousseeuw, 1986):

$$SI(C) = \frac{1}{N} \Sigma_{c_i \in C} \Sigma_{x_l \in c_i} \frac{b(x_l, c_i) - a(x_l, c_i)}{max\{a(x_l, c_i), b(x_l, c_i)\}},$$

   where $b(x_l, c_i) = \min\limits_{c_j \in C \setminus c_i} \{1/|c_i| \Sigma_{x_p \in c_i} \mathrm{d_E}(x_l, x_p)\}$ and

$a(x_l, c_i) = 1/|c_i|\Sigma_{x_p \in c_i} \mathrm{d_E}(x_l, x_p)$. Note than $|c_i|$ is the cardinality of the set or number of objects belonging to set c_i. The compactness is assessed according to the distance between all the observations in the same partition, while the separation is according to the closest neighbour distance. The value of *SI* ranges from -1 to 1. If the value of *SI* for one single observation is near to -1, it implies that the observation is closer to other cluster than its own cluster; otherwise, if the value is near to 1, it implies that the average distance to the cluster to which it belongs is smaller than to any other cluster. The value around zero indicates overlapping clusters. The higher the *SI* value, the more separated and compact are the clusters.

2. *Dunn's index* (DI)
   DI can be defined as the ratio between the minimum distance between two partitions and the size of the largest partition. The index can be calculated as (Dunn, 1973):

$$DI(C) = \frac{\min_{c_i \in C} \left\{ \min_{c_j \in C \setminus c_i} \delta(c_i, c_j) \right\}}{\max_{c_i \in C} \Delta(c_i)},$$

   where $\delta(c_i, c_j) = \min_{x_l \in c_i} \min_{x_p \in c_j} d_E(x_l, x_p)$ and $(c_i) = \max_{x_l, x_p \in c_i} \mathrm{d_E}(x_l, x_p)$. The cohesion is estimated by the nearest neighbour distance while the separation is estimated the by the maximum cluster diameter. A high value of this index indicates a compact and well-separated cluster.

3. *Calinski-Harabasz' index* (CHI)
   The cohesion is estimated according to the distances of the observation in a cluster from the centroid; while the separation is based on the distance of the centroid from the global centroid. The higher the index the better. The index can be defined as (Caliñski and Harabasz, 1974):

$$CHI(C) = \frac{N-k}{k-1} \cdot \frac{\sum_{c_i \in C} |c_i| d_E(\overline{c_k}, \bar{X})}{\sum_{c_i \in C} \sum_{x_l \in c_i} d_E(x_l, \overline{c_i})}.$$

4. RESULT AND DISCUSSION

4.1 *Descriptive Statistics*

The data used in this study consists of 153 objects (the countries), one response variable (`life_ladder`), and six predictors (`log_gdp, hle, social_support, freedom, generosity,` and `corruption`). Table 1 shows the summary of the data. The distribution of Cantril ladder question's answers could give a portray to compare happiness levels as well as inequality across the countries. The global average is 5.473 (out of 10) and the standard deviation is 1.112. Afghanistan became the country which has the lowest score of the Cantril ladder question (2.567) while Finland has the highest score (7.809). The scores fluctuated significantly among population-weighted regions, where the highest score is North America, Australia and New Zealand (NAAZ) region (7.174), followed by Western Europe (6.899), Latin America and the Caribbean (5.982), Central and Eastern Europe (5.884), East Asia (5.715), Southeast Asia (5.383), the Commonwealth of Independent States (CIS) (5.358), the Middle East and North Africa (MENA) (5.227), South Asia (4.475), and Sub-Saharan Africa (4.383). The happiness inequality can be evaluated by the standard deviation of the distributions of individual happiness scores. The lowest scores are found in East Asia, Western Europe, and NAAZ; while largest scores are found in Sub-Saharan Africa, Latin America and Caribbean, and MENA.

The country with the highest value of GDP is Luxembourg while Burundi has the lowest one. NAAZ is the region which the highest average GDP and Sub-Saharan African region has the lowest value. The inequality in terms of HLE is very big as shown by its standard deviation: the maximum value is 78.505 (Singapore) and the minimum value is 45.2 (Central African Republic). The region with the highest value of HLE is Western Europe (72.864). Central African Republic has the lowest score of the GWP question about social support while Icelanders are among people who confidently answered that they do have relatives or friends whenever they are in trouble. Afghanistan has the lowest value of freedom to make life choices, meaning that people there are not satisfied with their freedom to choose what they do with their lives. Another interesting fact is that Myanmar and Indonesian people are among the most generous people compared to other citizens, making Southeast Asia region is placed in the second position (after NAAZ region) for the highest generosity aspect. Singaporeans do believe that the corruption is not widespread throughout both the government and within the business (the lower score is the better), while Bulgarians are less confident that their government (and in the business as well) is not being corrupted.

TABLE 1. - *Summary of the Data*

Variables	Mean	Standard Deviation	Max.	Min.
life_ladder	5.473	1.112	7.809	2.567
log_gdp	9.296	1.202	11.451	6.493
hle	64.446	7.058	76.805	45.200
social_support	0.809	0.121	0.975	0.319
freedom	0.783	0.118	0.975	0.397
generosity	-0.015	0.152	0.561	-0.301
corruption	0.733	0.175	0.936	0.110

4.2 *Feature Selection*

The multiple linear regression model was used to describe how the six predictors (`log_gdp`, `hle`, `social_support`, `freedom`, `generosity`, and `corruption`) explain the variation of `life_ladder` across countries. The estimates of the regression parameter model are shown in Table 2 (see the second column). The third column denotes the *p*-value for the hypotheses test of the population regression parameters. Note that the coefficients of `log_gdp`, `hle`, `social_support`, `freedom`, and `corruption` are statistically significant at the level of 5%, while the coefficient of generosity is not statistically significant. Altogether, the six predictors explain more than 73% of the variation in happiness among the countries being investigated. Specifically, the sample coefficient of determination R^2 is 74.8%, while the adjusted R^2 is 73.8%. (Theil, 1978, suggested to use adjusted R^2 than R^2 since R^2 is likely to yield an overly optimistic image of the fitted value, especially when the number of predictors is not too small compared to the number of observations.) The value tells the proportion of variation in the response variable described by the predictor variables. In this case, the value of adjusted R^2 equals to 73.8% means that 73.8% of the variability of SWB can be described by the previously mentioned regression model. This value is a sign of a good model.

TABLE 2. - *Regression Result*

Variables	Original Model		Refined Model	
	Estimates	***p*-value**	**Estimates**	***p*-value**
constant	-2.059	0.002*	-1.939	0.003*
log_gdp	0.229	0.006*	0.214	0.009*
hle	0.035	0.007*	0.035	0.008*
social_support	2.723	0.000*	2.742	0.000*
freedom	1.777	0.000*	1.922	0.000*
generosity	0.411	0.225	N/A	N/A
corruption	-0.628	0.048*	-0.728	0.018*

* Coefficient is significant at the 0.05 level (2-tailed)

Since the `generosity` is not statistically significant, this variable is discarded for the next analysis. The new linear regression model which only consists of five predictors have all statistically significant coefficients (see the last column of Table 2) while the adjusted R^2 is 73.7%. The overall model (by employing the analysis of variance) is also statistically significant (*p*-value = 0.000).

4.3 *Clustering Result*

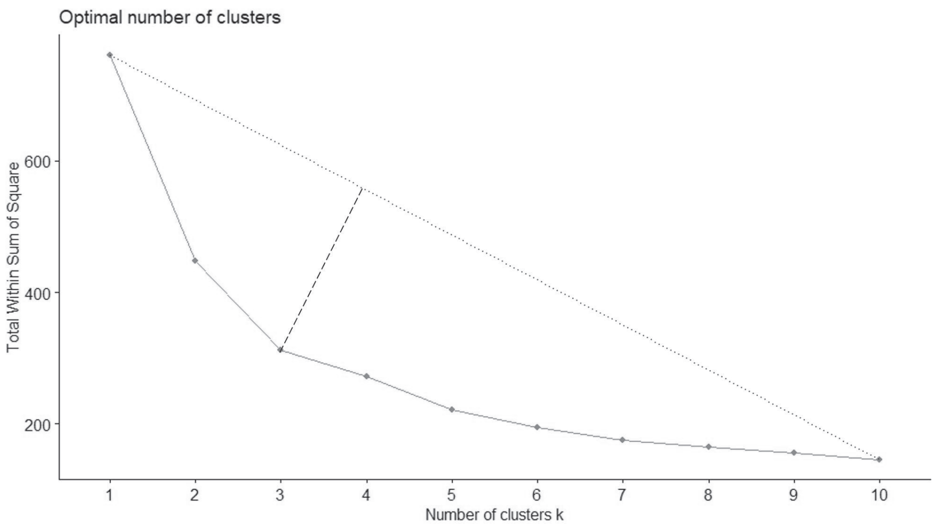
Before performing clustering analysis with several clustering algorithms, it is necessary to standardized data, obtaining 0 means and standard deviation equals to 1. The standardized value can be defined by:

$$Z_{ij} = \frac{X_{ij} - \overline{X_j}}{s_j},$$

where Z_{ij} is the z-score for object i and variable j, X_{ij} is the original data value, $\overline{X_j}$ is the mean or average of variable j, and s_j is the standard deviation of variable j. This standardized value has an advantage of unitless (the numerator and the denominator are in the same units). In addition, it is beneficial for the next analysis.

The first algorithm used is k-means. kmeans function in R is used (in `stats package`). In R, the format is `kmeans(x,centers)`, where `x` is the observations and `centers` is the predefined number of clusters. In this study, the elbow method was used to investigate the optimal number of clusters (see Section 3.2). The elbow graph is depicted in Figure 2. Note that the curve is plotted in solid line, while the dotted line connects the start and end points of the curve, and the dashed line is orthogonal to the dotted line that crosses the curve, maximizing the distance between the dashed line and the blue curve. It gives the optimal number of clusters = 3. The first cluster has 25 members (countries), the second cluster has 88 members, and cluster number three has 40 members. The centroid of each cluster is shown in Table 3. (Due to space limitation, the cluster membership for each algorithm is not shown but will be provided by the author upon request.)

FIGURE 2 - *The Elbow Graph for k-means Algorithm*



The second algorithm is *k*-means++. `kmeanspp` function is used (in `licors` package). The format is `kmeanspp(x,centers)`. The centroid of each cluster is shown in Table 3. Note that the result is identical to *k*-means. Next, *k*-medoids (PAM) is performed by means of the function `pam` in `cluster` package. The format is `pam(x,centers,metric)`, where `metric` specifies the distance metrics to be used (`metric=eucledian` was used, meaning that we used the Euclidean distance). By also employing the elbow graph, the optimal number of clusters is found, i.e., 3. The medoid of each cluster is presented in Table 3. Note that the cluster membership is almost similar to *k*-means (or *k*-means++), yet only five countries, i.e., Iceland, Belgium, Uruguay, Estonia, and Japan, have different memberships.

The function `clara` in `cluster` package is used for identifying cluster membership in CLARA algorithm. The format is `clara(x,centers,metric, samples)`, where the Euclidean distance was also used and samples means the number of samples drawn from the data set (`samples=50` was chosen). By also employing the elbow graph, the optimal number of clusters is identified, i.e., 3. It is worth nothing that the cluster membership is identical to *k*-medoids. The result is unsurprising since the data used is not considered as large enough, so that the algorithm behaves like PAM.

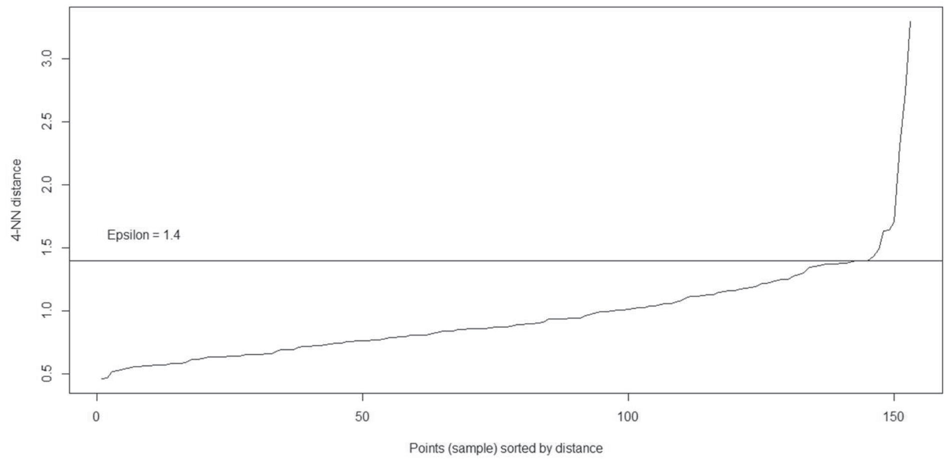
Affinity propagation algorithm can be executed by using `apcluster` function in `apcluster` package. The format is `apcluster(s,x)`, where `s` is the similarity matrix of the objects or similarity function. The `negDistMat` with `r=2` from `Matrix` package was used as similarity function. It created a square matrix of mutual pairwise similarities of vectors as negative distances (`r=2` is applied to obtain negative squared distances as what Frey and Dueck, 2007, demonstrated). The algorithm results 13 clusters. To do spectral clustering algorithm in R, `specc` function from `kernlab` package is used. The format is `specc(x,centers)`, where `centers=3` was used. The centroids are presented in Table 3.

The only density-typed clustering used in this study, i.e., DBSCAN, is run by utilizing function `dbscan` in `dbscan` package. The format is `dbscan(x,eps, MinPts)`, where `eps` is the size of the epsilon neighbourhood and `MinPts` is the number of minimum points in the epsilon region. To determine the epsilon value, the `kNNdist` function was used: `kNNdist(x,k=4)`. The idea is to calculate the average of the distances of every object to its *k* closest neighbours. Next, these *k*-distances would be plotted in an ascending order. The aim is to define the *knee*, which corresponds to the optimal epsilon. A knee is defined as a threshold where a sharp change occurs along the *k*-distance curve. The curve is depicted in Figure 3. One can observe that the optimal epsilon is around 1.4. Note that in DBSCAN, there are no cluster centres, and clusters are produced by linking adjacent points to one another. The algorithm resulted only one cluster, while four countries (Central African Republic, Rwanda, Swaziland, and Uzbekistan) are called *noises* that do not belong to any cluster.

TABLE 3. - *The Cluster Centres of Each Algorithm's Result*

Clusters (number of members)	log_gdp	hle	social_ support	freedom	corruption
k-means:					
Cluster 1 (40)	10.691	72.348	0.925	0.901	0.432
Cluster 2 (88)	9.611	66.496	0.846	0.787	0.803
Cluster 3 (25)	7.730	54.995	0.654	0.701	0.767
k-means++:					
Cluster 1 (40)	10.691	72.348	0.925	0.901	0.432
Cluster 2 (88)	9.611	66.496	0.846	0.787	0.803
Cluster 3 (25)	7.730	54.995	0.654	0.701	0.767
k-medoids:					
Cluster 1 (43)	10.813	72.301	0.939	0.909	0.365
Cluster 2 (90)	9.566	66.480	0.897	0.800	0.771
Cluster 3 (20)	7.751	54.468	0.638	0.707	0.762
CLARA:					
Cluster 1 (43)	10.813	72.301	0.939	0.909	0.365
Cluster 2 (90)	9.566	66.480	0.897	0.800	0.771
Cluster 3 (20)	7.751	54.468	0.638	0.707	0.762
Affinity propagation:					
Cluster 1 (16)	10.9	73.3	0.929	0.907	0.326
Cluster 2 (16)	8.6	63.9	0.814	0.891	0.732
Cluster 3 (15)	10.2	69.7	0.894	0.743	0.852
Cluster 4 (18)	10.5	70.0	0.889	0.873	0.677
Cluster 5 (22)	9.67	67.0	0.878	0.850	0.852
Cluster 6 (12)	9.69	67.4	0.857	0.614	0.817
Cluster 7 (11)...	7.47	54.1	0.621	0.720	0.760
Cluster 8 (15)	9.33	62.9	0.749	0.725	0.791
Cluster 9 (9)	8.19	56.1	0.696	0.806	0.779
Cluster 10 (7)	7.34	53.9	0.576	0.529	0.770
Cluster 11 (1)	6.63	45.2	0.319	0.641	0.892
Cluster 12 (1)	7.60	61.1	0.541	0.901	0.184
Cluster 13 (10)	7.88	55.0	0.760	0.686	0.802
Spectral clustering:					
Cluster 1 (46)	7.85	56.0	0.665	0.713	0.761
Cluster 2 (90)	9.74	67.1	0.860	0.797	0.793
Cluster 3 (17)	10.90	73.3	0.929	0.902	0.341

FIGURE 3. - *k-distance Curve*



The last two algorithms, AGNES and DIANA, which belong to hierarchical clustering will be analysed differently. Those two can be executed by using `agnes` and `diana` functions in `cluster` package. The format is `agnes`—or `diana(x,diss,-method,metric,stand)`, where `diss` is a logical flag: if `T` (or true), then `x` is treated as it is a dissimilarity matrix, otherwise, `x` is assumed to be a matrix of observations by variables. Note that `diss=F` was used in this study. This research used `ward` (from Ward's method) in the argument `method`, which minimizes the total within-cluster variance. It implies that at each iteration, the pair of clusters with minimum between-cluster distance will be combined. The Euclidean distance also used to fill the argument `metric`. The argument stand is also a logical flag: if `T`, then `x` will be standardized first before the dissimilarities are computed; since `x` is already standardized, so `stand=F` was used.

The results of those two algorithms are dendrograms. (It is hard to see the hierarchy since there are many countries; and, due to space limitation, the full dendrogram is not displayed.) The dendrogram is interpreted as follows. As we move up the dendrogram (or the hierarchical tree), similar countries or objects are merged into a twig. Again, similar combined objects would be merged into a bigger twig (or branch). The process is repeated until we have one completed tree combining from bigger branches. The height of the fusion of the similar (combined) objects which is displayed on the vertical axis, shows the distance or (dis)similarity between two objects (or combined objects). The higher the height, the less similar the objects (or combined objects) are.

The main limitation of the hierarchical clustering is that this clustering does not provide the number of clusters. It is because the objective does not try to form partitions, rather, it tries to describe the data to be structured like an evolutionary tree (Kaufman and Rousseeuw, 1990). However, one could cut the dendrogram at a certain value of height to divide the objects into clusters. The number of clusters equals to three is chosen arbitrarily, just to show how the dendrograms are cut, see Figure 4 and Figure 5.

FIGURE 4. - *''Cut'' Dendrograms of AGNES Algorithm*

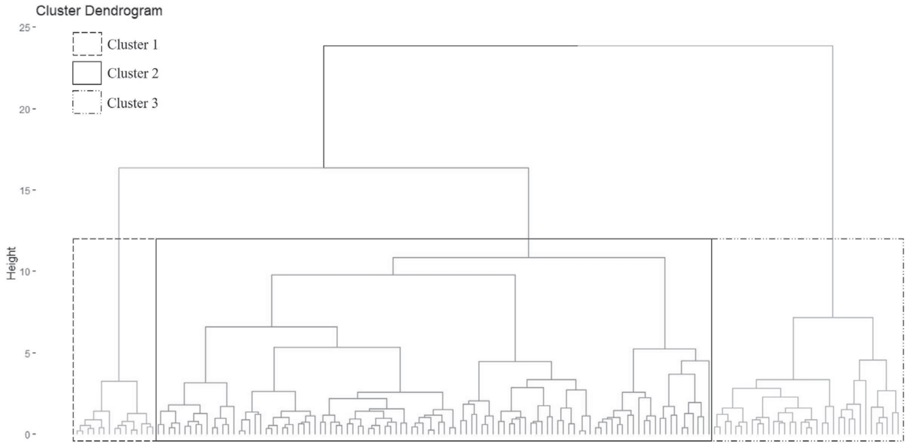
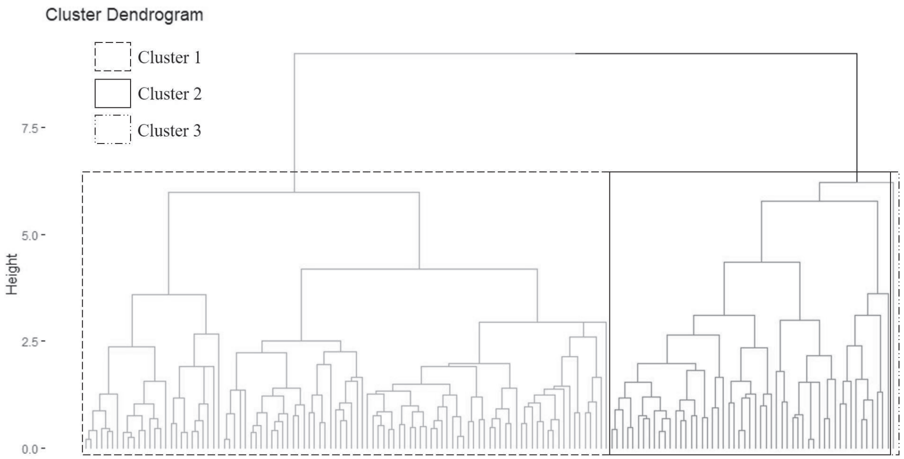


FIGURE 5. - *''Cut'' Dendrograms of DIANA Algorithm*



4.4 *Cluster Validation*

Previous subsection has demonstrated the algorithms as well as the results generated by each algorithm. This subsection would describe how to compare the algorithms based on the cluster validation techniques. The relative cluster validation using the elbow method to choose the optimal number of clusters. The external cluster validation cannot be used in this study since obviously, it is only able to be used in a controlled test environment. This study used real data set so that the structure of the data is unknown and hence, the correct partition (or the ground truth) is unavailable (Arbelaitz *et al*., 2013). Therefore, only the internal validation will be discussed here.

Table 4 shows the algorithms, along with three internal validation indices, i.e., SI, DI, and CHI, as well as the average within error sum of squares (SSE) and the average between SEE. Note that the hierarchical clustering (AGNES and DIANA) is not included here, and also DBSCAN since it only contains one cluster so that it is not possible to calculate the distance between clusters. Affinity propagation algorithm has both the minimal average within SSE and average between SSE, while *k*-medoids and CLARA both have the maximum values. Among three internal validation indices, the algorithms being consideration are overlapping each other. According to SI, spectral clustering is considered as ‘‘the best’’ algorithm, while based on DI, affinity propagation is ‘‘the best’’; however, *k*-means and *k*-means++ are ‘‘the best’’ according to CHI.

TABLE 4. - *Comparing Clustering Algorithms*

Algorithms	Number of clusters	Average within SSE	Average between SSE	SI	DI	CHI
k-means	3	1.8797	3.5692	0.3569	0.0977	107.6658
k-means++	3	1.8797	3.5692	0.3569	0.0977	107.6658
k-medoids	3	1.8866	3.6105	0.3654	0.0628	105.6677
CLARA	3	1.8866	3.6105	0.3654	0.0628	105.6677
Affinity propagation	13	1.2512	3.0222	0.2171	0.1435	59.4024
Spectral clustering	3	1.8840	3.6146	0.3700	0.0628	104.1439

The result is not unanticipated since the differences of the previous internal validation indices make it hard to compare in the same environment. Some scholars showed that there is no single internal validation index which surpasses other indices (Arbelaitz *et al*., 2013; Dimitriadou, Dolñicar and Weingessel, 2002; Maulik and Bandyopadhyay, 2002; Milligan and Cooper, 1985). Therefore, it is not recommended to proclaim ‘‘the best’’ algorithm when comparing clustering algorithms (Xu and Wunsch II, 2005). For the next discussion, the algorithm which has the highest CHI value will be analysed. There are two algorithms in this case, and *k*-means algorithm is selected arbitrarily.

4.5 *Interpretation of Results of k-means Algorithm*

Providing users with meaningful insights from the original data could be considered as the ultimate goal of clustering analysis. It allows users to effectively solve the problems they face. This subsection would discuss how to interpret the algorithm result as we can gain some insights and knowledge. Previously, the *k*-means algorithm was chosen to be analysed. Note that it does not make *k*-means the best algorithm among others since the decision is rather arbitrary. Map of cluster membership is shown in Figure 6. Each cluster's characteristic will be discussed as follows.

FIGURE 6. - *Map of Cluster Membership According to the k-means Algorithm*

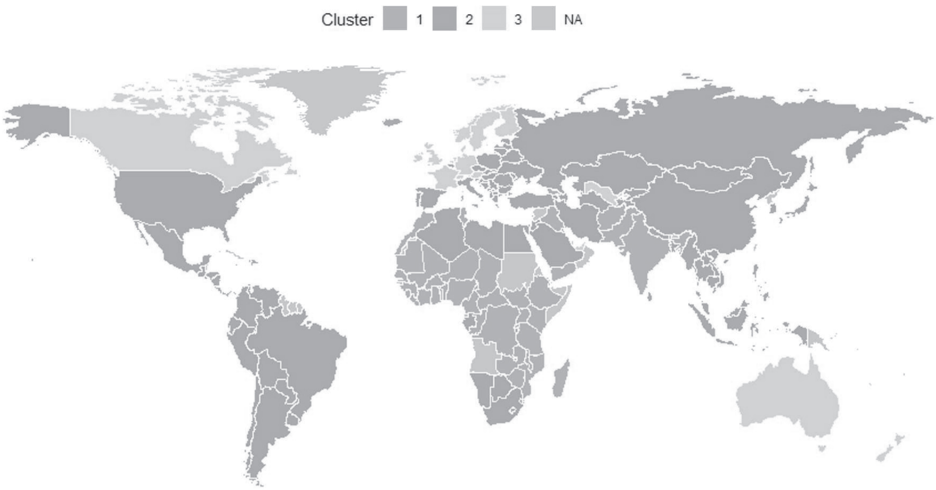
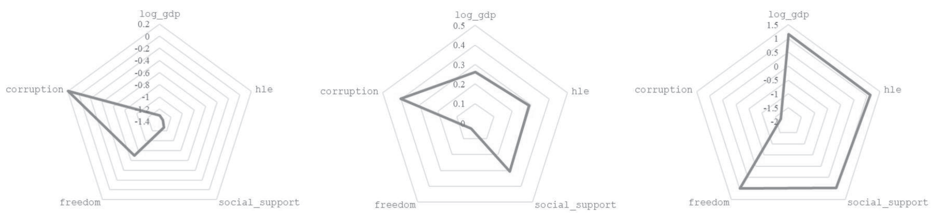


FIGURE 7. - *The Radar Charts of Each Cluster from k-means Algorithm*



(a) First Cluster (b) Second Cluster (c) Third Cluster

The first cluster consists of 40 countries, from four regions: one Latin America and Caribbean country (Haiti), two from MENA (Morocco and Yemen), three from South Asia (Afghanistan, India, Pakistan), and the rest are from Sub-Saharan Africa. The radar chart is depicted in Figure 7, where the values are the (standardized) means of each predictor variable, making it easier to make comparisons. This cluster has the lowest average value of GDP (the lowest GDP among all countries is in this cluster). In term of HLE, this cluster is also the worst, with the average value of 55 years. The condition also happens in other two predictors, i.e., social support and freedom to make life choices. The average values of social support (0.654) and freedom (0.701) are the worst among other clusters. An interesting fact is that the citizens of the countries belonging to this cluster feel better in term of perception of corruption than the second cluster. Comparing with the data from World Economic Situation and Prospect Report (United Nations, 2020), majority countries belong to low-income economies, also the member of the first cluster. It is arguably to say that this cluster has the least happy citizens.

There are 88 countries in the second cluster, making it the most widely spread cluster since the members are coming from all regions. This cluster has the average value for all predictor variables. The members of the cluster have the average log GDP value of 9.611; HLE at birth of 67 years; social support of 0.846; freedom to make life choices of 0.787. The government and the business operated in the countries of this cluster are perceived bad (the worst among all clusters) by the citizens in term of corruption. Countries in the Eastern Europe that have a bad reputation in this aspect belong to this cluster; the bottom three of this feature, i.e., Bulgaria, Romania, and Bosnia and Herzegovina are members of this cluster. Another interesting fact is that this cluster has the United States of America, the only one country in the region of NAAZ. This country suffers of the bad perception of corruption and health facilities.

The third cluster has all features that make it the happiest cluster. Its average values of all features are the best among all clusters. It contains only 25 countries; one from CIS (Uzbekistan), Central and Eastern Europe (Estonia), Southeast Asia (Singapore), Latin America and Caribbean (Uruguay), MENA (United Arab Emirates), two from East Asia (Hong Kong and Japan), three from NAAZ (Canada, Australia, and New Zealand), and the rests are from Western Europe. The lowest value in social support of this cluster (i.e., Hong Kong with 0.846) is still better than the country which has the highest social support value in the first cluster (i.e., Yemen with 0.818). Apart from the absence of the United States in this cluster (also other high-income countries, such as South Korea and Italy), another surprising information is that Uzbekistan, a lower-middle-income country according to United Nations (2020), is the member of this cluster. This country has the highest value of freedom to make life choices, compensating its low GDP and HLE.

5. CONCLUSION AND FUTURE RESEARCH DIRECTION

This research has demonstrated several clustering algorithms to partition countries according to the WHR 2020. The features used as basis for clustering are `log_gdp`, `hle`, `social_support`, `freedom`, and `corruption`. Nine clustering algorithms were selected according to the type of the data used (all the numerical information is interval-type of data), the objective of the algorithms, and to represent each type of the clustering algorithm (i.e., partitioning clustering: *k*-means, *k*-means++, *k*-medoids; affinity propagation; spectral clustering; density-type clustering: DBSCAN; and hierarchical clustering: AGNES and DIANA). Note that there is no the best clustering algorithm. *k*-means was selected arbitrarily as a representative of the algorithm to show the interpretation of its result. The (selected) final clustering contains three clusters whose characteristics are described in Section 4.5. It can be arguably inferred that the first cluster has the least happy citizens, while the third cluster is the happiest one.

One possible future research is that since this study only takes the *hard* clustering algorithms into account, it is of interest to also apply the *fuzzy* clustering algorithm. Lastly, this study only somehow connects the finding with the economic state of the country (i.e., the GNI). There are several recent contributions in the literature that connect happiness to sustainability, e.g., Carlsen (2018), Cloutier, Jambeck and Scott (2014), Cloutier and Pfeiffer (2015), therefore, linking country's cluster membership to its state of sustainability is an interesting area to be pursued.

ACKNOWLEDGEMENT

The author would like to thank Paola Zuccolotto for her helpful advice in preparing and writing this paper and the anonimous reviewer for her/his useful suggestions which improved the final version of the paper.

REFERENCES

Arbelaitz O., Gurrutxaga I., Muguerza J., Pérez J.M., Perona I. (2013). An extensive comparative study of cluster validity indices. *Pattern Recognition*, **46**, 243-256.

Arthur A., Vassilvitskii S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027-1035.

Brusco M.J., Steinley D., Stevens J., Cradit J.D. (2019). Affinity propagation: An exemplar-based tool for clustering in psychological research. *British Journal of Mathematical and Statistical Psychology,* **72**, 155-182.

Cai D., He X., Li Z., Ma W.Y., Wen J.R. (2004). Hierarchical clustering of WWW image search results using visual, textual and link information. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, 952-959.

Caliñski T., Harabasz J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, **3**, 1-27.

Cantril H. (1965). *The Pattern of Human Concern*. Rutdgers University Press, New Brunswick.

Carlsen L. (2018) Happiness as a sustainability factor. The world happiness index: A posetic-based data analysis. *Sustainability Science*, **13**, 549-571.

Cloutier S., Pfeiffer D. (2015). Sustainability through happiness: A framework for sustainable development. *Sustainable Development*, **23**, 317-327.

Cloutier S., Jambeck J., Scott N. (2014). The Sustainable Neighborhoods for Happiness Index (SNHI): A metric for assessing a community's sustainability and potential influence on happiness. *Ecological Indicators*, **40**, 147-152.

Dimitriadou E., Dolñicar S., Weingessel A. (2002). An examination of indexes for determining the number of clusters in binary data sets. *Psychometrika*, **67**, 137-159.

Dunn J.C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, **3**, 32-57.

Ester M., Kriegel H.P., Sander J., Xu X. (1996). Density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, 226-231.

Frey B.J., Dueck D. (2007). Clustering by passing messages between data points. *Science*, **315**, 972-976.

Halkidi M., Batistakis Y., Vazirgiannis M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, **17**, 107-145.

Hastie T., Tibshirani R., Friedman J. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition. Springer.

Helliwell J.F., Layard R., Sachs J., De Neve J.-E. (2020). *World Happiness Report 2020*. Sustainable Development Solutions Network, New York.

Jain A.K., Murty M.N., Flynn P.J. (1999). Data clustering: A review. *ACM Computing Surveys*, **31**, 264-323.

John R., Ramesh H. (2017). Colour based segmentation of a landsat image using k-means clustering algorithm. *Journal of Image Processing & Pattern Recognition Progress*, **4**, 31-38.

Kapourani C.A., Sanguinetti G. (2019). Melissa: Bayesian clustering and imputation of single-cell methylomes. *Genome Biology*, **20**, 61.

Kaufman L., Rousseeuw P.J. (1986). Clustering large data sets (with discussion). In E.S. Gelsema and L.N. Kanal (Eds.), *Pattern Recognition in Practice II* (pp. 425-437). Elsevier, Amsterdam.

Kaufman L., Rousseeuw P.J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis.* John Wiley & Sons, Hoboken.

Law C.K., Yip P.S.F. (2003). Healthy life expectancy in Hong Kong Special Administrative Region of China. *Bulletin of the World Health Organization*, **81**(1), 43-47.

Lyubomirsky S. (2008). *The How of Happiness: A Scientific Approach to Getting the Life You Want*. Penguin Press, New York.

MacQueen J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 281-297.

Manly B.F.J., Alberto J.A.N. (2017). *Multivariate Statistical Methods: A Primer*. 4th Edition. CRC Press, Boca Raton.

Maulik U., Bandyopadhyay S. (2002). Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 1650-1654.

Milligan G.W., Cooper M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, **50**, 159-179.

Minako F.S., Ulkhaq M.M., 'Sa Nu D., Pratiwi A.R.A., Akshinta P.Y. (2019). Clustering internet shoppers: An empirical finding from Indonesia. In *Proceedings of the 5th International Conference on E-business and Mobile Commerce*, 35-39.

Pal N.R., Biswas J. (1997). Cluster validation using graph theoretic concepts. *Pattern Recognition*, **30**, 847-857.

Rousseeuw P.J. (1986). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, **20**, 53-65.

Salomon J.A., Wang H., Freeman M.K., Vos T., Flaxman A.D., Lopez A.D., Murray C.J. (2012). Healthy life expectancy for 187 countries, 1990-2010: A systematic analysis for the Global Burden Disease Study 2010. *The Lancet*, **380**, 2144-2162.

Theil H. (1978). *Introduction to Econometrics*. Prentice Hall, Englewood Cliffs.

Theodoridis S., Koutroumbas K. (2008). *Pattern Recognition*. 2nd Edition. Academic Press, San Diego.

Ulkhaq M.M., Fidiyanti F., Adyatama A., Maulani Z.A., Nugroho A.S. (2019). Segmentation of cinema audiences: An empirical finding from Indonesia. In *Proceedings of the 2nd International Conference on Data Storage and Data Engineering*, 3-8.

Unglert K., Radiæ V., Jellinek A.M. (2016). Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra. *Journal of Volcanology and Geothermal Research*, **320**, 58-74.

United Nations. (2020). *World Economic Situation and Prospects 2020*. United Nations, New York.

Utami A.A., Ginanjar A.R., Fadlia N., Lubis I.A., Ulkhaq M.M. (2019). Using shopping and time attitudes to cluster food shoppers: An empirical finding from Indonesia. *Journal of Physics: Conference Series*, **1284**, 012005.

van Lettow B., Vermunt J.K., de Vries H., Burdorf A., van Empelen P. (2013). Clustering of drinker prototype characteristics: What characterizes the typical drinker? *British Journal of Psychology*, **104**, 382-399.

Wang J., Li M., Deng Y., Pan Y. (2010). Recent advances in clustering methods for protein interaction networks. *BMC Genomics*, **11**(S3), S10.

Xu R., Wunsch II D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks,* **16**, 645-678.

Assessing the tendency of judging bias in student competition: a data mining approach

M. Mujiya Ulkhaq
Department of Industrial Engineering, Diponegoro University, Semarang, Indonesia and
Department of Economics and Management, University of Brescia, Brescia, Italy
Susatyo N.W. Pramono
Department of Industrial Engineering, Diponegoro University, Semarang, Indonesia, and
Arga Adyatama
PT Kawan Lama Sejahtera, Jakarta, Indonesia

Received 12 February 2022
Revised 28 May 2022
28 September 2022
Accepted 8 October 2022

Abstract

Purpose – Judging bias is ironically an inherent risk in every competition, which might threaten the fairness and legitimacy of the competition. The patriotism effect represents one source of judging bias as the judge favors contestants who share the same sentiments, such as the nationalistic, racial, or cultural aspects. This study attempts to explore this type of judging bias in a university student competition. In addition, this study tries to expand the literature on judging bias by proposing the term universitarian bias as the judge is suspected to give a higher score to contestants from the same university.

Design/methodology/approach – The association rule of data mining is used to accomplish the objective of the study. To demonstrate the applicability of the method, the data set from the annual national university student competition in Indonesia is exploited.

Findings – The result strongly discovers that the universitarian bias is likely to be present. Some recommendations are also provided in order to minimize the bias that might happen again in the future.

Practical implications – As the implication of the presence of the universitarian bias, the committee should remove all the university features attributed to the participants. This endeavor is expected to minimize the universitarian bias that might happen.

Originality/value – This research is claimed to be the first attempt in implementing the data mining technique in the field of judging bias.

Keywords Association rule, Data mining, Judging bias, Universitarian bias, Student competition

Paper type Research paper

Introduction

The spirit of every competition is that the result is neither determined in advance (or prior to the competition), nor affected by any event outside of what is going on in the competition. However, no competition is immune to dispute regarding the judging decision as in many competitions the winners and the losers are decided by the decisions of the judges. As a consequence, even a small amount of bias in the judges' decision would determine between winning and losing. Therefore, biased judges must be avoided since this can highly damage the reputation of competitive and fair competition. Truthful contestants (or athletes in a sport competition), coaches (or teachers in student competition), fans (or supporters), as well as officials and sponsors extremely wish for the objective, fair and unbiased judges.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

2050-7003
DOI 10.1108/JARHE-02-2022-0053

A concern about this potential judging bias in competition leads to two issues: efficiency and fairness (Page and Page, 2010). The efficiency issue arises when the best contestant may eventually not be selected as the winner. The second issue raises a question about the fairness of the competition: did a contestant suffer from some disadvantages relative to other contestants for irrelevant motives and causes?

In the literature, there are six judging biases that have been empirically proposed (Auweele *et al.*, 2004). The first bias is due to the patriotism effect as the judge favors contestants who share the same sentiments, such as the nationalistic, racial, or cultural aspects. One of the most discussed subjects of this bias is nationalistic bias, which refers to the tendency of judges to prefer contestants from their own country (Ansorge and Scheer, 1988; Callahan *et al.*, 2016; Lyngstad *et al.*, 2020; Zitzewitz, 2006). The halo effect as the second bias refers to the tendency of judges to generalize their scores on one dimension to others, instead of carefully distinguishing between the different performance dimensions (Anderlucci *et al.*, 2021; Borman, 1975). The third is the order bias, which is defined as the tendency of judges to expect a bad or good performance as a function of the rank order, in which the performance takes place (Ansorge *et al.*, 1978; Page and Page, 2010; Plessner, 1999). Next, the memory bias refers to the tendency of judges to be affected by the prior performance (Ste-Marie, 2003; Ste-Marie and Lee, 1991; Ste-Marie and Valiquette, 1996; Ste-Marie *et al.*, 2001). The fifth is the reputation bias, which refers to the tendency of judges to base their decisions on the reputation of the contestant (Findlay and Ste-Marie, 2004). The last is the conformity effect, which is defined as the tendency of judges to adapt their decisions to their colleagues' (other judges) decisions (Scheer *et al.*, 1983; Wanderer, 1987).

This paper aims to determine whether or not a pattern of bias was shown by judges in a (university) student competition by employing the association rule of data mining. The observed competition was suspected to suffer from judging bias, where the tendency of participants to get a medal is higher when they were assessed by a judge from the same university. Therefore, the patriotism effect bias—we call it a "universitarian bias" since a judge coming from a particular university tends to give a higher score to a participant coming from the same university—is raised.

This paper contributes to the literature on judging bias in several ways. First, we propose the term "universitarian bias" as a variant of the patriotism effect bias. Second, we extend the scope and application of the study on the judging bias to student competition as most of the literature on judging bias is applied to sports competitions. Finally, the use of the association rule of data mining to identify the tendency of bias is considered novel in the literature on judging bias. Note that previously, scholars used the sign test (Ansorge and Scheer, 1988; Campbell and Galbraith, 1996; Popovic, 2000; Whissell *et al.*, 1993), the permutation test (Emerson and Meredith, 2011) and the linear regression (Callahan *et al.*, 2016; Emerson *et al.*, 2009; Leskošek *et al.*, 2012; Lyngstad *et al.*, 2020; Sampaio, 2012; Sandberg, 2018; Zitzewitz, 2006) to investigate the bias.

The remaining of the paper is described as follows. The next section illustrates the data set and the method used in this research. The result of the study is presented in the third section. The discussion section after the result section argues in more detail about the method used, the result of the study, the limitations of the study, and also the possible future research. Finally, the last section concludes and discourses the managerial implication.

Judging bias: theoretical foundation

Judgment and decision-making play a major role in every competition, with the adequacy of the processes being directly related to the success or failure in the competition. The study of judgment and decision-making can be traced back to the late 1940s, evidenced mainly by three major, quite independent approaches with the implicit and/or explicit purpose of improving their outcome: the decision- and game-theoretical, the psychological, and the social-psychological/sociological approaches (see Bar-Eli *et al.*, 2011, for more detailed review).

In a widely accepted operational way, judgment can be defined as the differentiation between different objects or identification of single objects in terms of certain qualitative or quantitative features (Eiser, 1990). In this basic sense, judgments are distinct psychological phenomena that do not need to be (but often are) connected with decisions. The empirical study of human judgment can be traced back to at least the middle of the nineteenth century when scholars tried to identify relationships between the objective (i.e. physically measurable) magnitude or intensity of a stimulus and the subjective magnitude or intensity that people experience. Since then, several different routes have been taken to reveal and understand the processes that underlie human judgment. This has led to the development of theories with various degrees of specificity, e.g. psychophysics, social judgment theory and social cognition (Fiske and Taylor, 2008; Kunda, 1999). On the other side, the number of theories in decision-making naturally depends on the broadness of the definition of what it means to make a decision. In competition, decision-making theories can be classified according to (1) their nature (deterministic, probabilistic, or deterministic/probabilistic), and (2) their timeline: static (i.e. all options compared at one time), dynamic (a sample of options is considered in sequential sampling) or static/dynamic (Bar-Eli *et al.*, 2011).

In every competition, there is a notion: "May be the best man win". In order to increase the chances that the best contestant(s), indeed wins a competition, judges (sometimes called referees, umpires, officials, or linesmen) are installed in almost all competitions to ensure the course of the competition in accordance with the rules. Judges must assess the performance of the team or individual live, without comprehensive technical assistance, surrounded by cheering spectators, and according to instructions specified in scoring regulations. Two components of judging are accuracy and fairness (Heiniger and Mercier, 2021). The first component relates to how the judges enforce the laws of the game (Plessner and Betsch, 2002). If the rule is complex, it is quite difficult to evaluate every single aspect; this leads to an inevitable element of subjectivity and randomness in the marks given by each judge. Novice judges often consult their scoring sheet much more often than experienced judges, thus missing execution errors. Furthermore, experienced judges have superior perceptual anticipation, and are more capable of detecting errors.

The second component relates to impartiality and lack of favoritism. Judges are susceptible to biases. As a human, judges are hardly making a rational (or ideal) decision. The Nobel laureate Herbert Simon proposed the notion of "bounded rationality", meaning that human rationality—when compared to any *ideal* and/or normatively *rational* models—is bounded by limited cognitive information-processing ability, by factors such as imperfect information and time constraints, and, last but not least, by emotions. Therefore, it is unfortunate that judging bias is an inherent risk in every competition as simply the fact that judges are human beings.

The judging bias in competition leads to efficiency and fairness issues (Page and Page, 2010). The first issue happens when *the best* contestant(s) may eventually not be selected as the champion. The second issue happens when the contestant(s) may suffer from some disadvantages relative to others due to the bias. In the literature, six judging biases have been empirically established (Auweele *et al.*, 2004). The first bias is called the patriotism effect (some literature called nationalistic bias), which is considered the most prevalent bias. It comes in two flavors: judges can favor contestant(s) who share(s) the same sentiments, such as nationalistic, racial, or cultural aspects; and at the same time penalize their competitors. One of the most frequently encountered biases of this type of bias is nationalistic bias, which refers to the tendency of judges to prefer contestant(s) from their own country (see for example, Ansorge and Scheer, 1988; Callahan *et al.*, 2016; Lyngstad *et al.*, 2020; Zitzewitz, 2006).

The second bias is called the halo effect (Anderlucci *et al.*, 2021). Judges succumbing to the halo effect assign scores to the contestant(s) by attending to a global impression of each contestant rather than by carefully distinguishing among levels of performance that individual contestant exhibit on different performance dimensions. These judges may "justify" their overall evaluations of each contestant by providing consistently high (or low, or average) scores across all performance dimensions, when in fact, many contestants exhibit significant relative strengths and weaknesses on different performance dimensions (Borman, 1975).

The third is the order bias. It refers to the tendency of judges to evaluate the performance of the contestant(s) as a function of the rank order (Ansorge *et al.*, 1978; Page and Page, 2010; Plessner, 1999). The psychological literature suggests that sequential presentation of information may influence the way each piece of information is processed and recorded Mussweiler (2003). Studies in economics (Neilson, 1998) and marketing (Novemsky and Dhar, 2005) have also found that a choice among situations of sequential choices may be dependent on the history of the sequence. This issue is of special importance in the competition. If there is any effect of the order, in which contestant(s) are assessed, it means that the evaluation process is biased. There are two main reasons why biases may result from sequential ordering. The first is that judges may not remember equally well the different performances in the sequence, and second, the criteria/benchmark of the evaluation may change over time (Page and Page, 2010).

The fourth is the memory bias. It refers to the tendency of judges to be affected by the prior performance (Ste-Marie, 2003; Ste-Marie and Lee, 1991; Ste-Marie and Valiquette, 1996; Ste-Marie *et al.*, 2001). Memory effects on perception and recognition judgments have led to the suggestion that perception of specific events is greatly influenced by a single prior processing episode (e.g. Eich, 1984; Witherspoon and Allan, 1985). In a typical gymnastics competition, for example, an athlete is allowed a brief warm-up just before the competitive performance, and the judges watched both the competitive performance and the warm-up. If the memory for the specific prior episode (in this case, the warm-up) influences the perception of the later performance (the competitive performance), then this influence could be either detrimental or beneficial to the athlete. For instance, if the athlete made an error during the warm-up but not during the competition, the judges' memory, for that error might bias the perception of the better performance during the competition (Ste-Marie and Lee, 1991).

The fifth is the reputation bias. It refers to the tendency of judges to base their decisions on the reputation of the contestant (Findlay and Ste-Marie, 2004). It is difficult to enter any evaluative social situation that is not prefaced in some way by expectations or assumptions about the various persons and characteristics involved in the situation. For example, it has been shown in workplace settings that raters who were highly familiar with the worker tended to give more positive overall ratings than when they were not familiar with that individual (Kingstrom and Mainstone, 1985). In evaluations of academic teaching, it has been demonstrated that students who are expected to do well receive higher evaluations than those for whom the teacher does not have high expectations (Murphy *et al.*, 1985; Rosenthal and Jacobson, 1968). Findlay and Ste-Marie (2004) found that a reputation bias does exist when judging figure skating, and that it is present during the evaluation phase of sports performance appraisal.

The last is the conformity effect. It refers to the tendency of judges to adapt their decisions to their colleagues' (other judges) decisions (Scheer *et al.*, 1983; Wanderer, 1987). In gymnastics competitions, the officials have frequently been evaluated in terms of the degree of agreement among judges. A series of studies over the last 30 years evaluated the judging at selected championship meets solely by intercorrelating the scores among judges, the implication being that an objectively judged meet would be one, in which there was high agreement among judges while poor objectivity would result from the low agreement (Faulkner and Loken, 1962; Johnson, 1971). Equating objectivity with high agreement among judges undoubtedly introduces some pressure on officials to conform. The pressure to conform would result in lowered objectivity for each judge who succumbs to the pressure, but increased agreement among judges.

Data and method

Data

The data used in this study are extracted from the annual national (university) student competition held in Indonesia. The competition aims to look for (the best) students' innovation through the creative research-based program. There are two types of competitions, i.e. oral presentation and poster. Due to the confidentiality issue, the original data set cannot be presented here; instead, a mock-up data set—to give the readers such an illustration—is shown in Table 1.

Method

To assess the tendency of judging bias, the association rule of data mining is used. Briefly, the association rule is the task of discovering patterns that describe association relationships between variables in a large data set (Zhang and Zhang, 2002). It is often illustrated in the rule form expressing attribute conditions that occur frequently together in a given data set. For instance, an association rule in the form of $\{a, b\} \Rightarrow \{c\}$ indicates that if a set of $\{a, b\}$ appears together in the given data set, it is likely that $\{c\}$ also appears. Useful applications of the association rule include finding groups of genes that have related functionality, identifying Web pages that are accessed together, understanding the relationships between different elements of Earth's climate system, or discovering products that are frequently brought together by customers. In this study, we extend the use of this method in the area of judging bias.

Formally, the association rule can be defined as follows. Let $I = \{i_1, i_2, \ldots, i_k\}$ be a set of k distinct (binary) attributes called "item sets" in the data set $D = \{r_1, r_2, \ldots, r_n\}$, where r_i is called "record". Each record in D has a unique ID and contains a subset of the item sets in I. Let X and Y are subsets of I. An association rule is defined as an implication of the form: $X \Rightarrow Y$,

Participants (or teams)	Judges	Gold medal	Silver medal	Bronze medal
Class: A; Room: 1				
University A	University A	University A	University C	University B
University B	University K			
University C	University H			
University D				
University E				
Class: A; Room: 2				
University D	University D	University D	University H	University I
University F	University G			
University H	University K			
University I				
University J				
Class: B; Room: 1				
University A	University D	University D	University A	University K
University D	University F			
University E	University L			
University G				
University K				
Class: C; Room: 1				
University B	University A	University B	University I	University F
University C	University E			
University F	University G			
University H				
University I				

Table 1.
Mock-up data set

where $X, Y \subseteq I$ and both are disjoint, i.e. $X \cap Y = \varnothing$. The rule suggests a (strong) relationship exists between set X and set Y. Note that X is named antecedent—some references called left-hand-side (*LHS*), and Y is called consequent or right-hand-side (*RHS*).

For a given k distinct items, there are 2^k possible candidate item sets. The number of possible association rules R is defined as (Tan *et al.*, 2016):

$$R = \sum_{i=1}^{k-1}\left[\binom{k}{i} \times \sum_{j=1}^{k-i}\binom{k-i}{j}\right]. \tag{1}$$

Selecting interesting rules from the set of all possible rules is constrained by several measures. In this study, three important measures are presented, i.e. support, confidence, and lift. The support of X, *supp* (X), refers to the proportion of records that has X in the data set. For instance, when *supp* (X) equals 0.3, it means that the item set X appears in 30% of all the records in the data set. This measure is characterized as "the higher the better", meaning that for a given association rule, higher support is preferred.

The confidence of the rule, *conf* ($X \Rightarrow Y$), determines how frequent item sets in Y appear in the records that contain X. It can be written as,

$$conf\,(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}. \tag{2}$$

Let say, for a given data set *conf* ($X \Rightarrow Y$) = 0.75, it means that for 75% of the records containing X, the rule is correct. The confidence of the rule can be interpreted as an estimate of $P(Y|X)$, i.e. the probability of finding Y in the records under the condition that these records also contain X (Hipp *et al.*, 2000). This measure is also the higher the better.

The lift of the association rule, *lift* ($X \Rightarrow Y$), is defined as the ratio of the observed support to that expected if X and Y were independent (Brijs *et al.*, 2000; Brin *et al.*, 1997). It can be written mathematically as,

$$lift(X \Rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)supp(Y)} = \frac{conf\,(X \Rightarrow Y)}{supp(Y)}. \tag{3}$$

If the value of the lift is 1 (one), it implies that X and Y are independent of each other since X and Y appear together under the conditional independence assumption. However, if the value is more than 1, it means that the degree, to which X and Y appear together is dependent on one another. If the lift is less than one, it means that X and Y are substitutes for each other, meaning that the occurrence of X has a negative effect on the occurrence of Y, and vice versa. Greater lift values indicate stronger associations.

Equation (1) implies that it is computationally expensive if one would like to identify all possible rules. There are abundant algorithms trying to solve this issue (see Goethals and Zaki, 2004, who compared several algorithms). Among those, this research implemented the *apriori* algorithm, developed by Agrawal and Srikant (1994), which is a level-wise, breadth-first algorithm that counts records. The algorithm is the first association rule algorithm that initiated the implementation of support-based pruning to systematically restrict the candidate item sets from growing exponentially. It means that this algorithm could help to reduce the number of candidate item sets explored during frequent item set generation. For technical issues, refer to Tan *et al.* (2016).

Result

The rule is defined as follows: *LHS* $\Rightarrow$ *RHS*. The antecedent or *LHS* is that at least one of the participants and one of the judges in the particular room are coming from the same university. The consequent (or *RHS*) varies, namely, (1) the participants got a gold medal ($LHS \Rightarrow$ *gold medal*), (2) got silver a medal ($LHS \Rightarrow$ *silver medal*), (3) got bronze a medal ($LHS \Rightarrow$ *bronze medal*), (4) got any medal, either gold, silver, or bronze ($LHS \Rightarrow$ *any medal*) and (5) got nothing ($LHS \Rightarrow$ *no medal*). The rule then can be interpreted as follows. $LHS \Rightarrow$ *gold medal* means when a participant from say, University A is assessed by a judge from also University A, it is likely that the participant will get the gold medal.

The 2020 data (the recent event of the competition) is used in this research. Recall there are two types of competitions, i.e. oral presentation and poster. The data set is categorized into four classes: (1) full data set containing records from both oral presentation and poster session, (2) sub data set I: containing records from oral presentation only, (3) sub data set II: records from poster session only and (4) sub data set III: records only from universities that got the medals.

The *apriori* algorithm is used to run the association rule of data mining. The algorithm is based on the *apriori* principle, stating that if an item set is frequent, then all of its subsets must also be frequent (Tan *et al.*, 2016). To execute the algorithm, we employed *R*, a programming language for statistical computing and graphics. It is motivated by the recognition of *R* in the field of statistics, data mining and machine learning. apriori command from arules package is used (Hahsler *et al.*, 2005). We use 0 (zero) as the minimum support and confidence measures, implying that this mechanism would never discard the rule when *supp* ($LHS \Rightarrow RHS$) and *conf* ($LHS \Rightarrow RHS$) are more than 0. We choose this threshold since we conduct the *backward procedure*, i.e. the rule is determined prior to running the algorithm. The result consisting of triple measures of the association rule, namely, support, confidence and lift, are shown in Table 2 (see the second to the fourth columns).

Consequent (RHS)	Support	Confidence	Lift	*p*-value
(i) Full data set				
(a) Gold medal	0.0105	0.3037	3.2461	0.0000*
(b) Silver medal	0.0061	0.1778	1.9159	0.0005*
(c) Bronze medal	0.0074	0.2148	2.2231	0.0000*
(d) Any medal	0.0240	0.6963	2.4606	0.0000*
(e) No medal	0.0105	0.3037	0.4236	0.0000*
(ii) Sub data set I				
(a) Gold medal	0.0103	0.3077	3.2787	0.0000*
(b) Silver medal	0.0051	0.1538	1.4706	0.1872
(c) Bronze medal	0.0087	0.2615	2.7419	0.0000*
(d) Any medal	0.0241	0.7231	2.4607	0.0000*
(e) No medal	0.0092	0.2769	0.3922	0.0000*
(iii) Sub data set II				
(a) Gold medal	0.0107	0.3000	3.2164	0.0000*
(b) Silver medal	0.0071	0.2000	2.4679	0.0002*
(c) Bronze medal	0.0061	0.1714	1.7518	0.0349*
(d) Any medal	0.0240	0.6714	2.4669	0.0000*
(e) No medal	0.0117	0.3286	0.4514	0.0000*
(iv) Sub data set III				
(a) Gold medal	0.0189	0.3130	2.2373	0.0000*
(b) Silver medal	0.0111	0.1832	1.2968	0.1552
(c) Bronze medal	0.0134	0.2214	1.5670	0.0066*
(d) Any medal	0.0434	0.7176	1.6986	0.0000*
(e) No medal	0.0171	0.2824	0.4890	0.0000*

Note(s): *Significant at the level of 5%

Table 2.
Result of the association rule

The interpretation of the result is as follows. For the full data set, *supp* (*LHS* ⇒ *gold medal*) = 0.0105, meaning that only 1.05% of the participants who were assessed by the same university judges (from all the records) have won the gold medal. This statistic seems to be trivial if we also look at *supp* (*LHS* ⇒ *silver medal*) and *supp* (*LHS* ⇒ *bronze medal*) that equal to 0.61 and 0.74%, respectively. However, the confidence measure tells a different story. The value of *conf* (*LHS* ⇒ *gold medal*), which equals 30.37% implies that about one-third of the records containing only participants who were assessed by the same university judges, the rule is correct. Further, if we inspect *conf* (*LHS* ⇒ *any medal*), we would be surprised since the probability of the participants who got any medal where the judges were from the same university is 0.6963. If the event of getting any medal for a participant who competed with other competitors in a particular panel room is assumed to be equally likely, then the probability of the event is only 5%. Comparing the huge amount of difference between these probabilities (0.6963 vs. 0.05), one might possibly doubt that the judge was fair and unbiased.

This suspicion is strengthened by looking at the lift measure. The value of *lift* (*LHS* ⇒ *gold medal*) is 3.2461. It infers that the odds of a participant getting the gold medal is three times more likely when the judge is from the same university as this participant. Since the value is more than one, it means that those two item sets (i.e. *LHS* and *gold medal*) are not independent of one another. We confirm this by conducting the chi-square test of independence (Alvarez, 2003) to check the dependence between the antecedent and the consequent. Because the *p-value* (see the last column of Table 2) is less than the significance level of 5%, we conclude that the association between the antecedent and the consequent exists. It is worth noticing the value of *lift* (*LHS* ⇒ *no medal*) is less than one, meaning that the occurrence of *LHS* has a negative effect on the occurrence of *RHS*. It indicates that the participant would get the medal (i.e. the negation of "get no medal") when there is a judge from the same university taking a charge in the participant's panel room.

The relatively similar patterns are found in the sub data sets I and II: containing the oral presentation and the poster session, respectively. The discrepancy between the value of *supp* (*LHS* ⇒ *gold medal*) in the full data set with the value of the same measure in the presentation and poster sub data sets are both only 0.0002, while for *supp* (*LHS* ⇒ *any medal*) is only 0.0001. Note that the biggest discrepancy between full the data set with sub-data sets I and II occurs in *supp* (*LHS* ⇒ *bronze medal*), which is only 0.0013. For the confidence measure, the discrepancy between *conf* (*LHS* ⇒ *gold medal*) in the full data set with the oral presentation is 0.0040 and 0.0037 with the poster session. The anomaly occurs in the lift measure, as the discrepancy in *lift* (*LHS* ⇒ *silver medal*) and *lift* (*LHS* ⇒ *bronze medal*) are more than 0.4. This is anticipated due to the different scales among the lift measure and the support measure (and the confidence as well). The discovered patterns show that, no matter which data set we use, the conclusion inferred would be the same.

The rationale behind generating sub-data set III is that we want to investigate whether the medals won by participants were obtained because there were judges from those particular universities. The conditional probability illustrated by the value of *conf* (*LHS* ⇒ *any medal*), which equals 0.7176 is very high. Again, we could suspect that from the probability point of view with the assumption of equally likely, the occurrence of this event is not random. The values of lift measure for *LHS* ⇒ *gold medal*, *LHS* ⇒ *silver medal*, *LHS* ⇒ *bronze medal* and *LHS* ⇒ *any medal* are all more than one, implying that the events of getting medals (gold, silver, bronze and any medal) with the condition that the judges are from the same university are not independent. Finally, from all four classes of the data above mentioned, we could suspect that the universitarian bias has occurred. This claim is supported by the fact that the probability of participants getting a medal under the condition that the judge is from the same university is much higher than if the event is assumed to be equally likely. Moreover, by looking at the values of the lift measure (which are more than one), the dependency between those two events did exist.

Discussion

The main challenge in investigating judging bias is that we never observe the *objective* measure of performance. As a human, a judge, who assesses the performance of the participants in a competition is neither objective nor free from bias. As there are observed six major biases in the literature (Auweele *et al.*, 2004), we investigate the most discussed bias in the literature, which is due to the patriotism effect. We broaden the definition of this bias as we propose the term universitarian bias. In a student competition, students as contestants bring honor to their alma maters (or universities). Their performances are evaluated by lecturers who act as judges and seem to have a mission similar to the contestants. Therefore, the judges are suspected of whether they favor contestants from the same alma mater.

The nonparametric test is an alternative to identify the existence of this patriotism effect bias by comparing the scores of an observed judge with other judges. Ansorge and Scheer (1988) used three panel judges to evaluate the observed judge's scores in the Artistic Gymnastics Competition at the 1984 Olympic Games. The sign test was employed to detect whether the judge scored the contestants from his/her own country higher, lower, or the same as the panel judges. Popovic (2000) adopted the same methodology to investigate the nationalistic bias in the Rhythmic Gymnastics Competition at the 2000 Olympic Games. The sign test was also used by Campbell and Galbraith (1996) who contrasted the judge's scores with the median scores given by all judges. Whissell *et al.* (1993) compared the judge's scores with other judges' scores by using four nonparametric tests, namely, average-related, maximum- and minimum-frequency as well as rank-deviation methods. Finally, Emerson and Meredith (2011) used the permutation test to inspect the difference between the observed judge's score with the untrimmed mean of all other judges' scores in the 2000 Olympic Diving Competition. This study does not adopt this perspective since it could raise questions about other biases. For instance, how can we confirm that the panel—or other—judges are free from bias; and if there is a dispute between panel judges and the observed judge, how can we assure that the dispute is due to the bias of the observed judge?

The parametric method through the use of regression analysis was pioneered by the work of Zitzewitz (2006) who studied the bias that might exist during the 2002 Olympic winter sports. He assumed that the score given by a particular judge is influenced by the nationalistic bias (a dummy variable), other predictors (such as athlete fixed effect, judge fixed effect and judge country-fixed effect), and the error term, which varies among judge, athlete and performance. The normality assumption of each component in addition to zero expectation of the error term is essential for the estimation procedure. Other studies adopted this method and modified the model to exploit nationalistic bias in many different sports competitions including diving (Emerson *et al.*, 2009), surfing (Sampaio, 2012), football (Pope and Pope, 2015), gymnastics (Callahan *et al.*, 2016), dressage (Sandberg, 2018) and ski jumping (Lyngstad *et al.*, 2020). A slightly different model was proposed by Myers *et al.* (2006) who employed the use of the multi-level model (Goldstein, 2011) to find the evidence of nationalistic bias in Muaythai.

Although the parametric procedure has more statistical power than its counterpart (i.e. nonparametric), the normality assumption posed in the model is quite strong. The violation of the assumption is often neglected even though it is somewhat impractical. For an illustration, when the score ranges from one to 10 and the nature of the scoring is discrete, it is unrealistic to assume the score is normally distributed with say, $\mu = 5$. This is because the normal distribution is continuous, and the judge's score tends to have a negative skew (the normal distribution, on the other hand, has a symmetric shape).

The use of multinomial (logistics) regression is also appealing since it can be used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables. The only covariate in this regression is that at least one of the participants and a judge in the particular room are coming from the same university (in this study, it is labeled as *LHS*). Given the nature (characteristics) of the data as the only information revealed is about the participants (or the teams) and the judges in each room, this procedure has technical drawback about the significance of the covariate. If the covariate is not significant, obviously, the value of the covariate is pointless. The interpretation would be that the covariate cannot predict the judging bias. However, it is not as straightforward as it is; there is a chance that if the number of the data is increased, we can get the significant result. As a consequence, this procedure would be very sensitive to the amount of data—so as other parametric procedure. In a competition, sometimes we cannot obtain a sufficient enough data.

The unavailability of judges' scores, as in this research [1] and the strong assumption of the parametric procedure made the rise of different approaches; hence, this research employs the association rule of data mining. This procedure is claimed to be the first attempt in implementing data mining in the field of judging bias. Unlike other statistical procedures, the result does not exhibit the statistical significance of a particular hypothesis, instead, it shows the ratio, the probability and the odds. While the first is revealed by the support measure, the second and the last are indicated by the confidence and lift measures, respectively. Also, by the confidence function $conf\ (LHS \Rightarrow RHS)$, the conditional probability $P(Y|X)$ as in the parametric procedure can be identified with the same manner (where $LHS = X$ and $RHS = Y$).

As has been shown in the Result section, the value of $conf\ (LHS \Rightarrow any\ medal)$ in the full data set is 0.6963, which is considered high. (Also, the high values are found in the first, second and third sub-data set, i.e. 0.7231, 0.6714 and 0.7176, respectively). To give more information about the tendency of judging bias, we generate another rule, i.e. LHS^C or the complement of the *LHS*. It is defined as an event that the participant and the judges are not from the same university. Using the full data set, the measure $conf\ (LHS^C \Rightarrow no\ medal)$ equals 0.7318. This result reveals that the probability of the participant who is not from the same university as the judge getting no medal is more than 73%. Again, this statistic is higher compared to the assumption of equally likely event. Next, we also calculate the measure of $lift\ (LHS^C \Rightarrow any\ medal)$, which equals 0.9478. The fact that the value is less than one shows that the participants who are not from the same university as the judge would not get any medal (or would get no medal, as the negation of "get any medal"). The information obtained from both using *LHS* and LHS^C is not contradictory; thus, we strongly believe that the universitarian bias has occurred in the student competition being investigated.

Conclusion

This study examined whether there is a universitarian bias in the student competition. This new type of bias is proposed due to the suspicion that a lecturer who takes charge as a judge in the student competition tends to give a higher score to a participant from the same university. The association rule of data mining is used to investigate the bias. To exhibit the applicability of the proposed method, we extract the data from the annual national university student competition held in Indonesia. There are five rules generated in this study: (1) $LHS \Rightarrow gold\ medal$, (2) $LHS \Rightarrow silver\ medal$, (3) $LHS \Rightarrow bronze\ medal$, (4) $LHS \Rightarrow any\ medal$, and (5) $LHS \Rightarrow nothing$. The *LHS* refers to the event that the participant and the judge are from the same university, while the consequents refer to the outcome achieved by the participants. The result shows that the bias tends to exist; it is illustrated by the high values of the measures of the rule, i.e. the confidence and the lift (see the result section for the detail).

One of the limitations of using the association rule is that the minimum threshold of the support and confidence measures have to be defined subjectively by the decision-maker (Coenen and Leng, 2007). In this research, we only consider that those measures are more than zero but does not consider their magnitudes. For instance, is it fair to say that the rule is interesting if the confidence is only 10%? Therefore, it is recommended to conduct further research with more data to define the minimum threshold so that the rule appears to be interesting. A higher threshold is desired as the decision-maker would have a stronger belief that the bias was present. The cut-point of the logistics regression could be used that if the probability is more than 50%

(Ulkhaq *et al.*, 2018), it can be said that the rule is interesting (or the judging bias happened). Another issue in this research is due to the confidentiality of the data used. The committee does not reveal the data to the public, so it is very hard to evaluate the transparency. Also, the confidentiality issue makes the finding of this study is hard to be verified and validated.

As the implication of the presence of the universitarian bias, the committee should remove all the university features attributed to the participants. The papers submitted must be double-blind assessed, meaning that the participants do not know who has assessed their works, as well as the judges must not know who is being assessed. At the presentation session, if possible, there should not be judges from the same university as the participants in a particular panel room. These endeavors are expected to minimize the universitarian bias that might happen.

This study calls for future research to extend the application and implication of this study. First, this research only uses the data of 2020; further research that accommodates the longitudinal data is suggested to detect whether this negative practice has occurred over time or just in this particular time. Second, due to its flexibility, the association rule can be applied to other applications when they are suspected of judging bias, for example, in sports competitions, art contests, orchestra auditions, or even cooking tournaments. The method—with adjustment—is also possible to identify other types of judging biases, for instance, gender bias (Feld *et al.*, 2016; Sandberg, 2018), racial or ethnic bias (Larsen *et al.*, 2008; Parsons *et al.*, 2011; Price and Wolfers, 2010), cultural bias (Callahan *et al.*, 2016) and reputation bias (Findlay and Ste-Marie, 2004).

Note

1. The committee of the competition did not reveal the judges' scores to the public.

References

Agrawal, R. and Srikant, R. (1994), "Fast algorithms for mining association rules in large databases", *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94)*, pp. 487-499.

Alvarez, S.A. (2003), "Chi-squared computation for association rules: preliminary results", Technical Report BC-CS-2003-01, pp. 1-11.

Anderlucci, L., Lubisco, A. and Mignani, S. (2021), "Investigating the judges performance in a national competition of sport dance", *Social Indicators Research*, Vol. 156, pp. 783-799, doi: 10.1007/s11205-019-02256-z.

Ansorge, C.J. and Scheer, J.K. (1988), "International bias detected in judging gymnastic competition at the 1984 Olympic Games", *Research Quarterly for Exercise and Sport*, Vol. 59 No. 2, pp. 103-107, doi: 10.1080/02701367.1988.10605486.

Ansorge, C.J., Scheer, J.K., Laub, J. and Howard, J. (1978), "Bias in judging women's gymnastics induced by expectations of within-team order", *Research Quarterly. American Alliance for Health, Physical Education and Recreation*, Vol. 49 No. 4, pp. 399-405, doi: 10.1080/10671315.1978.10615552.

Auweele, Y.V., Boen, F., De Geest, A. and Feys, J. (2004), "Judging bias in synchronized swimming: open feedback leads to nonperformance-based conformity", *Journal of Sport and Exercise Psychology*, Vol. 26, pp. 561-571, doi: 10.1123/jsep.26.4.561.

Bar-Eli, M., Plessner, H. and Raab, M. (2011), *Judgement, Decision Making and Success in Sport*, Wiley-Blackwell, Chichester.

Brijs, T., Vanhoof, K. and Wets, G. (2000), "Defining interestingness for association rules", *International Journal of Information Theories and Applications*, Vol. 10 No. 4, pp. 370-376.

Borman, W.C. (1975), "Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings", *Journal of Applied Psychology*, Vol. 60, pp. 556-560, doi: 10.1037/0021-9010.60.5.556.

Brin, S., Motwani, R., Ullman, J.D. and Tsur, S. (1997), "Dynamic itemset counting and implication rules for market basket data", *Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data (ACM SIGMOD '97)*, pp. 255-264, doi: 10.1145/253262.253325.

Callahan, B.P., Mulholland, S.E. and Rotthoff, K.W. (2016), "Cultural bias: gymnasts, judges, and bilateral trade agreements", *The Journal of SPORT*, Vol. 5 No. 1, pp. 35-45, doi: 10.21038/sprt.2016.0512.

Campbell, B. and Galbraith, J.W. (1996), "Nonparametric tests of the unbiasedness of Olympic figure-skating judgments", *The Statistician*, Vol. 45 No. 4, pp. 521-526, doi: 10.2307/2988550.

Coenen, F. and Leng, P. (2007), "The effect of threshold values on association rule based classification accuracy", *Data and Knowledge Engineering*, Vol. 60 No. 2, pp. 345-360, doi: 10.1016/j.datak.2006.02.005.

Eich, E. (1984), "Memory for unattended events: remembering with and without awareness", *Memory and Cognition*, Vol. 12 No. 2, pp. 105-111, doi: 10.3758/BF03198423.

Eiser, J.R. (1990), *Social Judgment*, Open University Press, Buckingham.

Emerson, J.W. and Meredith, S. (2011), "Nationalistic judging bias in the 2000 Olympic diving competition", *Math Horizons*, Vol. 18 No. 3, pp. 8-11, doi: 10.4169/194762111X12954578042812.

Emerson, J.W., Seltzer, M. and Lin, D. (2009), "Assessing judging bias: an example from the 2000 Olympic Games", *The American Statistician*, Vol. 63 No. 2, pp. 124-131, doi: 10.1198/tast.2009.0026.

Faulkner, J. and Loken, N. (1962), "Objectivity of judging at the national collegiate athletic association gymnastic meet: a ten-year follow-up study", *Research Quarterly. American Association for Health, Physical Education and Recreation*, Vol. 33 No. 3, pp. 485-486, doi: 10.1080/10671188.1962.10616481.

Feld, J., Salamanca, N. and Hamermesh, D.S. (2016), "Endophilia or exophobia: beyond discrimination", *The Economic Journal*, Vol. 126 No. 594, pp. 1503-1527, doi: 10.1111/ecoj.12289.

Findlay, L.C. and Ste-Marie, D.M. (2004), "A reputation bias in figure skating judging", *Journal of Sport and Exercise Psychology*, Vol. 26 No. 1, pp. 154-166, doi: 10.1123/jsep.26.1.154.

Fiske, S. and Taylor, S.E. (2008), *Social Cognition: From Brains to Culture*, McGraw-Hill, New York.

Goethals, B. and Zaki, M.J. (2004), "Advances in frequent itemset mining implementations: report on FIMI'03", *ACM SIGKDD Explorations Newsletter*, Vol. 6 No. 1, pp. 109-117, doi: 10.1145/1007730.1007744.

Goldstein, H. (2011), *Multilevel Statistical Models*, 4th ed., John Wiley & Sons, Chichester.

Hahsler, M., Grün, B. and Hornik, K. (2005), "Arules – a computational environment for mining association rules and frequent item sets", *Journal of Statistical Software*, Vol. 14 No. 15, pp. 1-25, doi: 10.18637/jss.v014.i15.

Heineger, S. and Mercier, H. (2021), "Judging the judges: evaluating the accuracy and national bias of international gymnastics judges", *Journal of Quantitative Analysis in Sports*, Vol. 17 No. 4, pp. 289-305, doi: 10.1515/jqas-2019-0113.

Hipp, J., Güntzer, U. and Nakhaeizadeh, G. (2000), "Algorithms for association rule mining — a general survey and comparison", *ACM SIGKDD Explorations Newsletter*, Vol. 2 No. 1, pp. 58-64, doi: 10.1145/360402.360421.

Johnson, M. (1971), "Objectivity of judging at the National Collegiate Athletic Association gymnastic meet: a twenty-year follow-up study. Research Quarterly", *American Association for Health, Physical Education and Recreation*, Vol. 42 No. 4, pp. 454-455, doi: 10.1080/10671188.1971.10615096.

Kingstrom, P.O. and Mainstone, L.E. (1985), "An investigation of the rater-ratee acquaintance and rater bias", *Academy of Management Journal*, Vol. 28 No. 3, pp. 641-653, doi: 10.5465/256119.

Kunda, Z. (1999), *Social Cognition: Making Sense of People*, MIT Press, Cambridge.

Larsen, T., Price, J. and Wolfers, J. (2008), "Racial bias in the NBA: implications in betting markets", *Journal of Quantitative Analysis in Sports*, Vol. 4 No. 2, Article 7, doi: 10.2202/1559-0410.1112.

Leskošek, B., Čuk, I., Pajek, J., Forbes, W. and Bučar-Pajek, M. (2012), "Bias of judging in men's artistic gymnastics at the European championship 2011", *Biology of Sport*, Vol. 29 No. 2, pp. 107-113, doi: 10.5604/20831862.988884.

Lyngstad, T.H., Härkönen, J. and Rønneberg, L.T.S. (2020), "Nationalistic bias in sport performance evaluations: an example from the ski jumping world cup", *European Journal for Sport and Society*, Vol. 17 No. 3, pp. 250-264, doi: 10.1080/16138171.2020.1792628.

Murphy, K.R., Balzer, W.K., Lockhart, M.C. and Eisenman, E.J. (1985), "Effects of previous performance on evaluations of present performance", *Journal of Applied Psychology*, Vol. 70 No. 1, pp. 72-84, doi: 10.1037/0021-9010.70.1.72.

Mussweiler, T. (2003), "Comparison processes in social judgment: mechanisms and consequences", *Psychological Review*, Vol. 110 No. 3, pp. 472-489, doi: 10.1037/0033-295X.110.3.472.

Myers, T.D., Balmer, N.J., Nevill, A.M. and Al-Nakeeb, Y. (2006), "Evidence of nationalistic bias in MuayThai", *Journal of Sports Science and Medicine*, Vol. 5, pp. 21-27.

Neilson, W.S. (1998), "Reference wealth effects in sequential choice", *Journal of Risk and Uncertainty*, Vol. 17 No. 1, pp. 27-48, doi: 10.1023/A:1007791217751.

Novemsky, N. and Dhar, R. (2005), "Goal fulfillment and goal targets in sequential choice", *Journal of Consumer Research*, Vol. 32 No. 3, pp. 396-404, doi: 10.1086/497551.

Page, L. and Page, K. (2010), "Last shall be first: a field study of biases in sequential performance evaluation on the idol series", *Journal of Economic Behavior and Organization*, Vol. 73 No. 2, pp. 186-198, doi: 10.1016/j.jebo.2009.08.012.

Parsons, C.A., Sulaeman, J., Yates, M.C. and Hamermesh, D.S. (2011), "Strike three: discrimination, incentives, and evaluation", *American Economic Review*, Vol. 101 No. 4, pp. 1410-1435, doi: 10.1257/aer.101.4.1410.

Plessner, H. (1999), "Expectation biases in gymnastics judging", *Journal of Sport and Exercise Psychology*, Vol. 21 No. 2, pp. 131-144, doi: 10.1123/jsep.21.2.131.

Plessner, H. and Betsch, T. (2002), "Refereeing in sports is supposed to be a craft not an art: a response to Mascarenhas, Collins and Mortimer (2002)", *Journal of Sport and Exercise Psychology*, Vol. 24, pp. 334-337, doi: 10.1123/jsep.24.3.334.

Pope, B.R. and Pope, N.G. (2015), "Own-nationality bias: evidence from UEFA champions league football referees", *Economic Inquiry*, Vol. 53 No. 2, pp. 1292-1304, doi: 10.1111/ecin.12180.

Popovic, R. (2000), "International bias detected in judging rhythmic gymnastics competition at Sydney-2000 Olympic Games", *FACTA UNIVERSITATIS Series: Physical Education and Sport*, Vol. 1 No. 7, pp. 1-13.

Price, J. and Wolfers, J. (2010), "Racial discrimination among NBA referees", *The Quarterly Journal of Economics*, Vol. 125 No. 4, pp. 1859-1887.

Rosenthal, R. and Jacobson, L. (1968), *Pygmalion in the Classroom*, Holt Rinehart & Winston, New York.

Sampaio, B. (2012), "Three essays on applied microeconomics", Doctoral dissertation, University of Illinois, available at: https://core.ac.uk/download/pdf/4837557.pdf.

Sandberg, A. (2018), "Competing identities: a field study of in-group bias among professional evaluators", *The Economic Journal*, Vol. 128 No. 613, pp. 2131-2159, doi: 10.1111/ecoj.12513.

Scheer, J.K., Ansorge, C.J. and Howard, J. (1983), "Judging bias by viewing contrived videotapes: a function of selected psychological variables", *Journal of Sport Psychology and Exercise Psychology*, Vol. 5 No. 4, pp. 427-437, doi: 10.1123/jsp.5.4.427.

Ste-Marie, D.M. (2003), "Memory biases in gymnastic judging: differential effects of surface feature changes", *Applied Cognitive Psychology*, Vol. 17, pp. 733-751, doi: 10.1002/acp.897.

Ste-Marie, D.M. and Lee, T.D. (1991), “Prior processing effects on gymnastic judging”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 17 No. 1, pp. 126-136, doi: 10.1037/0278-7393.17.1.126.

Ste-Marie, D.M. and Valiquette, S.M. (1996), “Enduring memory-influenced biases in gymnastic judging”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 22 No. 6, pp. 1498-1502, doi: 10.1037/0278-7393.22.6.1498.

Ste-Marie, D.M., Valiquette, S.M. and Taylor, G. (2001), “Memory-influenced biases in gymnastic judging occur across different prior processing conditions”, *Research Quarterly for Exercise and Sport*, Vol. 72 No. 4, pp. 420-426, doi: 10.1080/02701367.2001.10608979.

Tan, P.-N., Steinbach, M. and Kumar, V. (2016), *Introduction to Data Mining*, 2nd ed., Pearson, London.

Ulkhaq, M.M., Widodo, A.K., Yulianto, M.F.A., Widhiyaningrum Mustikasari, A.K. and Akshinta, P.Y. (2018), “A logistic regression approach to model the willingness of consumers to adopt renewable energy sources”, *IOP Conference Series: Earth and Environmental Science*, Vol. 127, 012007, doi: 10.1088/1755-1315/127/1/012007.

Wanderer, J.J. (1987), “Social factors in judges’ rankings of competitors in figure skating championships”, *Journal of Sport Behavior*, Vol. 10 No. 2, pp. 93-102.

Whissell, R., Lyons, S., Wilkinson, D. and Whissell, C. (1993), “National bias in judgments of Olympic-level skating”, *Perceptual and Motor Skills*, Vol. 77 No. 2, pp. 355-358, doi: 10.2466/pms.1993.77.2.355.

Witherspoon, D. and Allan, L.G. (1985), “The effect of a prior presentation on temporal judgments in a perceptual identification task”, *Memory and Cognition*, Vol. 13 No. 2, pp. 101-111, doi: 10.3758/BF03197003.

Zhang, C. and Zhang, S. (2002), *Association Rule Mining: Models and Algorithms*, Springer-Verlag, Berlin.

Zitzewitz, E. (2006), “Nationalism in winter sports judging and its lessons for organizational decision making”, *Journal of Economics and Management Strategy*, Vol. 15 No. 1, pp. 67-99, doi: 10.1111/j.1530-9134.2006.00092.x.

Corresponding author
M. Mujiya Ulkhaq can be contacted at: ulkhaq@live.undip.ac.id