Open access • Journal Article • DOI:10.1080/00036846.2013.868592

A new international database on education quality: 1965–2010 — Source link

Nadir Altinok, Claude Diebolt, Jean Luc De Meulemeester

Institutions: Université libre de Bruxelles

Published on: 07 Feb 2014 - Applied Economics (Routledge)

Topics: Standardized test and Primary education

A new International Database on Education Quality: 1965-2010

Nadir Altinok, Claude Diebolt, Jean-Luc de Meulemeester

Les Documents de Travail de l'IREDU
Working Papers

Institut de Recherche sur l'Education
Sociologie et Economie de l'Education
Institute for Research in the Sociology and Economics of Education

A New International Database on Education Quality: 1965-2010

Nadir ALTINOK, Claude DIEBOLT, Jean-Luc DEMEULEMEESTER
Novembre 2013
DT 2013/5

Pôle AAFE – Esplanade Erasme – B.P. 26513 - F 21065 Dijon Cedex
Tél.+33 (0)3 80 39 54 50 - Fax +33 (0)3 80 39 54 79
iredu@u-bourgogne.fr- http://iredu.u-bourgogne.fr

A New International Database on Education Quality: 1965-2010

Nadir ALTINOK*, Claude DIEBOLT, Jean-Luc DEMEULEMEESTER

14/11/2013

Abstract: The aim of this paper is to propose a new database allowing a comparative evaluation of the relative performance of schooling systems around the world. We measure this performance through pupils achievement in standardized tests. We merge all existing regional and international student achievement tests by using a specific methodology. Compared to other existing databases, our approach innovates in several ways, especially by including regional student achievement tests and intertemporal comparable indicators. We provide a dataset of indicators of quality of student achievement (IQSA) for 103 countries/areas in primary education and 111 countries/areas in secondary education between 1965 and 2010.
Keywords: Quality, Human Capital, Education, Database, PISA, TIMSS.

JEL Classification: C8, I2, N3.

*Corresponding author, Nadir Altinok, BETA (University of Lorraine), IREDU (University of Bourgogne). Address: BETA (Bureau d'Economie Théorique et Appliquée), 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France. Tel: +33 387 54 71 38 Fax: +33 390 24 20 70. Email nadir.altinok@univ-lorraine.fr

Claude Diebolt, University of Strasbourg Institute for Advanced Study & Bureau d'Economie Théorique et Appliquée (BETA/CNRS), 61 avenue de la Forêt Noire, 67085 Strasbourg Cedex, France. Tel: +33 368 85 20 69 Fax: +33 368 85 20 71. Email: cdiebolt@unistra.fr

Jean-Luc De Meulemeester, Université Libre de Bruxelles, Avenue F.D. Roosevelt 50, 1050 Bruxelles. Email: jldemeul@ulb.ac.be

Acknowledgments: We are most grateful to Geeta Kingdon and Jean Bourdon and two anonymous referees for their helpful comments. Financial support from the University of Strasbourg Institute for Advanced Study (USIAS) is gratefully acknowledged by Claude Diebolt.

Introduction

In this paper we propose a new database allowing the measurement of the evolution of students' achievements in a large set of countries. The main reason for focusing and providing quantitative estimates of students learning achievements lies in the desire to identify the direction and strength of the causal link between schooling quality and economic performance – as it is now acknowledged that educational quality matters both for wage levels (at the micro level) and for economic growth (at the macro level).[1]

What we actually develop here is a standardized measure of pupils' achievements in several key competencies (reading, mathematics and sciences) in primary and secondary education at several points in time. We use the results of the various international assessments available (by the way also enlarging the geographical scope of such analysis as usually carried out by using results of tests carried out in Africa and Latin America not often used in the literature so far) to build a standardized score of those achievements as a proxy of the education system quality.

Our methodology aims at improving the seminal work done by Lee and Barro (2001) and Barro (2001), and consists of a major update of a previous work by Altinok and Murseli (2007). In their pioneering paper, Lee and Barro (2001) used direct results from International Student Achievement Tests (ISATs), without any specific methodology for adjusting potential differences between all the series. They used instead a regression technique – i.e. the seemingly unrelated regression – which allows to obtain different constants between each test, and hence to allow for potential differences between tests over years and over skills. We begin our analysis by using these studies.

Another method of anchoring has been used by Hanushek and Kimko (2000). These authors adjusted ISATs between 1964 and 1995 by using results from NAEP (*National Assessment of Educational Progress*[2]). Their methodology is only based on United States scores, and the data is limited to the restricted period 1964-1995. A recent paper by Hanushek and Woessmann (2012) aimed at correcting some of these imperfections by using an approach that assumes stability over time of the variance of quality of student achievement in a restricted number of OECD countries. The authors suggest two criteria for a group of countries to serve as a standardization benchmark for performance variation over time. Firstly, the countries have to be member states of the relatively homogenous and economically advanced group of OECD countries in the whole period of ISATs observations. Second, the countries should have had a substantial enrollment in secondary education already in 1964. Then, the authors suggest 13 countries that meet both of these measures of stability which are named "OECD Standardization Group" (OSG) of countries[3]. Hanushek and Woessmann (2012) assume that the cross-country variation among the OSG countries do not vary substantially since 1964. By using this assumption, they build new indicators of student achievements and educational quality. Their main measure of cognitive skills is a simple

[1]See Demeulemeester and Diebolt (2011).
[2]We provide a description of the NAEP in the appendix B.
[3]The OSG countries are: Austria, Belgium, Canada, Denmark, France, Germany, Iceland, Japan, Norway, Sweden, Switzerland, the United Kingdom, and the United States.

average of all standardized math and science test scores of the ISATs in which a country participated. Their database includes combined measure for the 77 countries that have always participated in any of the math and science tests.

As the authors explain in their paper, a major issue with this methodology concerns countries that are far from the measured OECD performance. In particular, countries far off the scale of the original test scores may not be well represented because the tests may be too hard and thus not very informative for them. This bias may be more important when analyses are focused on developing countries, which is the case of our study.

Moreover, the methodology used – i.e. the "OSG of countries" – does not take into account several improvements made by ISATs since 1995. The International Association for the Evaluation of Educational Achievement (IEA) and the Organisation for Economic Co-operation and Development (OECD) teams prepared modern ISATs in order to allow intertemporal comparisons. By using another methodology, Hanushek and Woessmann (2012) chose a specific approach and neglected the recent improvements made by psychometricians, as the Item Response Theory (IRT). Moreover, they do not clearly show to what extent their main assumption – i.e. the variation between the OSG of countries is stable – is corroborated by results from modern ISATs, as these ones permit to compare countries performance over time.

Another limit deals with the absence of Regional Student Achievement Tests (RSATs) in their database. Hanushek and Woessmann (2012) only focused on ISATs since the methodology used is based on developed economies. In our paper, we provide an alternative methodology which enables the possibility to include several regional assessments. As these ones are focused on developing countries, our study permits to analyze more deeply the question of universal primary education, and allow for specific analyzes for developing economies.

Compared to previous research, our approach innovates in several ways. First, we use *regional* student assessments tests (RSATs), such as LLECE, SACMEQ or PASEC[4], which were never used in previous papers. This enables us to obtain original data on the quality of student achievement for a larger number of developing countries, and especially for African and Latin American countries. Our methodology consists of anchoring ISATs with the results of United States in the US NAEP (National Assessment of Educational Progress) for only the very first ISATs. In parallel to this anchoring, recent achievement tests - for instance, the Trends in International Mathematics and Science Study (TIMSS) - permit us to make over time comparisons of country performance. Last but not least, adjusting RSATs needs a new methodology which uses countries participating at least to one ISAT and a RSAT. By using these countries performance in both tests, we make an anchoring of RSATs with ISATs. We therefore combine these new series – i.e. the Regional Student Achievement Tests – by using scores for countries which took part in at least one ISAT for the same period. The main advantage of this specific methodology is both to adjust old achievement tests and to permit to

[4]Respectively the Latin American Laboratory for Assessment of the Quality of Education (LLECE), the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and the Program on the Analysis of Education Systems (PASEC).

take into account future achievement tests. Hence, our database will be updated when new international and regional achievement tests will be available in the future[5].

Differences can be found with a previous publication (Altinok and Murseli, 2007) on this database. First of all, the use of trends on pupil performance between 1965 and 1980 has been made possible by using the micro databases of IEA's pioneering assessments. In the first version, we only based our statistics on global means published in reports. Another change is our choice not to include data from an assessment for which there is no micro-data available, in order to diminish measurement error. This leads to the exclusion of International Assessment of Educational Progress (IAEP)[6] and Monitoring Learning Achievement (MLA) assessments. Moreover, the inclusion of more developing countries has been made possible by updating the data with recent regional student assessments as the SACMEQ III, LLECE II or PASEC II assessments. Another change concerns the possibility to distinguish between male and female and rural and urban areas. Finally, we propose to complement standard means of student achievements with proportions of students reaching specific competency levels in order to track for different goals (such as the Education for All Goal).

We obtain two complementary databases. The first database gives the opportunity to compare student performance between countries for each level of education, by computing an average score for quality of student achievement between 1965 and 2010 (Database 1). This database includes quality scores for 103 countries/areas in primary education and 111 countries/areas in secondary education. The second table allows us to compare the global *change* in quality of student achievement over time, for a long term period. This long term table includes quality scores from 1965 to 2010 for 150 countries/areas and more than 800 different scores divided by countries, levels and years (Database 2). Each dataset is presented for general sample, by gender, by type of location. Moreover, besides mean score for each subgroup, we provide proportion of pupils for two predefined benchmarks (i.e. the 'Minimum Level' benchmark and the 'Advanced Level' benchmark).

2. Data and methodology

2.1. A Brief Presentation of International and Regional Learning Achievement Tests

This section describes the various international learning achievement tests that may be used to construct our indicators, named "Indicators of Quality of Student Achievement" (IQSA). We identify seven groups of international surveys in which 105 countries have participated. These groups can be divided into two subgroups. The first one consists of international assessments, while the second one consists of regional assessments. Detailed information on these assessments is provided in Table 1. We will only make here a short presentation of the various existing learning assessments. More information can be obtained in Appendix A.

The *International Association for the Evaluation of Educational Achievement* (IEA) was the first body to measure individual learning achievement for international comparative

[5]The dataset and other material can be found in the following link: http://www.beta-umr7522.fr/Datasets.
[6]Unfortunately, microdata for IAEP is not available. We would like to thank ETS and NCES for their support.

purposes in the early 1960s. The surveys include the highly regarded "Trends in International Mathematics and Science Study" (TIMSS) and "Progress in International Reading Literacy Study" (PIRLS). TIMSS test aims at evaluating skills of students in grades 4 and 8[7] in mathematics and science, while PIRLS is based on a test based on reading in Grade 4. Another well-known international assessment is PISA (Programme for International Student Assessment). The Organisation for Economic Co-operation and Development (OECD) launched its PISA in 1997. More generally, PISA has assessed the skills of 15-year-old every three years since 2000 in countries that together account for almost 90% of the global economy – i.e. a major part of the World GDP. Until now, four rounds of PISA are available (2000, 2003, 2006 and 2009). Moreover, data for additional countries in PISA 2009+ has been included in our dataset.

Two other international assessments are available. Drawing on the experience of the *National Assessment of Educational Progress* (NAEP), the *International Assessment of Educational Progress* (IAEP) comprises two surveys first conducted in 1988 and 1991. Under a joint UNESCO and UNICEF project, learning achievements have been assessed as part of the *Monitoring Learning Achievement* (MLA) programme on a vast geographical scale in more than 72 countries (Chinapah, 2003). This programme of assessment is flexible and ranges from early childhood, basic and secondary education to nonformal adult literacy. However, all of the data have not been published. Supplementing national reports, a separate report on MLA I was drafted for 11 African countries (Botswana, Madagascar, Malawi, Mali, Morocco, Mauritius, Niger, Senegal, Tunisia, Uganda and Zambia; see UNESCO, 2000). As the microdata of IAEP and MLA is not available, we preferred not to include these assessments in our database.

Three major regional assessments have been conducted in Africa and Latin America. The *Southern and Eastern Africa Consortium for Monitoring Educational Quality* (SACMEQ) grew out of a very extensive national investigation of the quality of primary education in 15 African countries in 1995-1999, 2000-2002 and 2007. Following a different approach, surveys under the *Programme d'Analyse des Systèmes Educatifs* (PASEC, or "Programme of Analysis of Education Systems") of the Conference of Ministers of Education of French-Speaking Countries (CONFEMEN) have been conducted in the French-speaking countries of sub-Saharan Africa since 1993. Finally, the network of national education systems in Latin American and Caribbean countries, known as the *Latin American Laboratory for Assessment of the Quality of Education* (LLECE), was established in 1994 and is coordinated by the UNESCO Regional Bureau for Education in Latin America and the Caribbean. Assessments conducted by the LLECE focused on learning achievements in reading, mathematics and science[8] in grades 3 and 4[9] in 13 countries of the subcontinent in 1998 and for grade 3 and 6 pupils in 2006.

[7]A grade consists of a specific stage of instruction in initial education usually covered during an academic year. Students in the same grade are usually of similar age. This is also referred to as a 'class', 'cohort' or 'year' (Glossary of UIS website available at: http://glossary.uis.unesco.org/glossary/en/home).
[8]Science skill was included in the second round only.
[9]A grade is a stage of instruction usually equivalent to one complete year. Hence, grade 3 represents the third year of compulsory schooling – i.e. of primary education in most countries.

All achievements tests undertaken and the main information concerning them are summarized in Table 1 and presented in Appendix A. The methodology used to adjust them in order to yield comparable indicators is briefly presented below and in more details in Appendix B.

2.2. A methodology for the consolidated study of surveys

Given the diversity of existing achievement tests, there is no single and comparable measure of pupil achievement over all tests. On the contrary, ISATs and RSATs differ greatly in the definition of what a pupil should know in the respective skill tested. Therefore, we propose a methodology which enables the comparison between all existing assessments. Below, we present shortly this methodology. The detailed methodology can be found in Appendix B.

This paper is a major update of a previous publication in 2007 (Altinok and Murseli, 2007). Compared to the previous paper, we made six major modifications:

- We did not include surveys for which we do not have the micro dataset, in order to obtain more precise data. Hence, MLA and NAEP assessments are not included in our updated paper.
- While in the first version we used results from international reports for earlier assessments of IEA, we now use the micro dataset. It allows us to obtain less biased results for years before 1980.
- The methodology of anchoring has been improved by using recent results of the PISA assessments. Now, our methodology is also based on the measured progress of students' achievements over time, instead of only using the simultaneous participation of countries to several assessments. It brings us the possibility to track trends in students' achievements between 1965 and 2010.
- We include recent international and regional assessments which were not included in the previous version (such as PISA 2009, TIMSS 2007 or LLECE II).
- Another change concerns the possibility to distinguish between male and female and rural and urban areas, which would allow us to analyze gender issues in potential future studies.
- Finally, we propose to complement standard means of student achievements with proportions of students reaching specific competency levels in order to track for different policy goals (such as the "Education for All" Goal or the goal of "innovation" which needs to track best-performing pupils over the world).

Below, we present succinctly the methodology. For a more detailed explanation, please consult Appendix B. We proceed in five steps. Firstly, we need to track overall performance overtime for IEA assessments. We need to obtain an anchored assessment which will provide possible future comparisons over time of pupils' results. As highlighted in section 2.1., IEA organized several assessments since 1964 in mathematics, science (known as TIMSS tests since 1995) and reading (known as PIRLS tests since 2001). Unfortunately, IEA assessments prior to 1995 in mathematics and science and 2001 in reading are not directly comparable

with further IEA tests. In order to adjust all IEA assessments, we propose to anchor them on an external assessment which provides comparable data over time. For doing this, we need data from another assessment which is comparable over time and which tests approximately the same population (Hanushek and Kimko (2000) used the same methodology). The National Assessment of Educational Progress (NAEP) can be used for this purpose. NAEP is the main instrument of evaluation in education in United States. As NAEP began to test pupils aged of 4 and 8 years old from 1964 to 2007, we propose to anchor old IEA assessments on trends on US results in NAEP. Therefore, if the performance of the United States declined in NAEP and changed in another way in IEA study, it would mean that the IEA study is upward or downward biased. In order to correct for this under- or over-estimation, we propose to adjust old IEA studies according to trends on NAEP results.

However, as recent IEA assessments permit over-time comparison of performance levels, there is no specific reason to anchor results on NAEP results after this period (which is after TIMSS 1995 and PIRLS 2001 assessments). Hence, our panel data gives exactly the same trends on students' performance between 1995 and 2007 for mathematics and science which can be found in international reports of the IEA (see for instance Mullis, I.V.S., Martin, M.O., Foy, P., 2009). Moreover, we also included data from PISA assessments. As our main methodology is based on IEA assessments, our paper includes PISA results by using different anchoring procedure. Firstly, we used the same way as done for TIMSS for adjusting PISA assessment. Indeed, the anchoring methodology was made on the basis of USA results in both PISA and NAEP. As many countries took part to both PISA 2000 and TIMSS 1999/2003, we predicted scores for adjusted PISA 2000 in the TIMSS 1999 or TIMSS 2003 datasets when countries did only took part to TIMSS 2003. This is a better way of including PISA results compared to an averaging between PISA and TIMSS results. Indeed, some recent papers highlighted that PISA and TIMSS results may differ because of different population target or test contents (see for instance Wu, 2010). Another issue may concern to differences in levels of scoring between PISA and TIMSS. If we simply aggregate results from TIMSS and PISA for a given year (for instance for 2003), some slight differences may appear for some countries. In order to correct this possible issue, we based the inclusion of PISA results on the variation of performance for countries which took part at two or more rounds of PISA. Instead of predicting all PISA scores with TIMSS scores, our methodology includes results for countries based on the change of their performance in PISA assessments.

Last but not least, adjusting regional student achievement tests (RSAT) needs a new methodology which uses countries participating at least to one ISAT and a RSAT. Since the USA did not take part to any regional student assessment, it is impossible to adjust the surveys with the NAEP assessment results. Our methodology relies on the possibility to combine results between ISATs and RSATs for countries which took part to both assessments. By using these countries results in both tests, we make an anchoring of RSATs with ISATs. We therefore combine these new series – i.e. the Regional Student Achievement Tests – by using scores for countries which took part in at least one ISAT for the same period. The main advantage of this specific methodology is both to adjust old achievement tests and to permit to take into account future achievement tests. Hence, our database will be updated when new international and regional achievement tests will be available in the future.

We obtain two complementary databases. For each observation, we provide data for the whole sample, by gender (female versus male) and by type of location (rural versus urban). Moreover, while the general data concerns mean scores for each subgroup, we also provide data for the proportion of student reaching two specific levels. The first level represents one standard deviation below the international average (i.e. 100 points from 500 points). Hence, the proportion of students who perform higher than 400 points represents the first benchmark. This group can be considered as reaching the 'Minimum Level' benchmark. Our second level deals with potential innovation and research development - or at least technological adaptation capabilities. Following the initial work by Hanushek and Woessmann (2012), we define the most able pupils as pupils reaching the level of 600 points (i.e. one standard deviation above the international mean). This group of students may innovate more (or contribute more to technological adaptation) than others and hence enhance economic growth (see Hanushek and Woessmann, 2012).

Dataset 1. Panel dataset. The first dataset deals with trends on pupils' achievement. On top of that, we propose to analyze the global trends in schooling quality, by creating new scores computed by combining achievement levels for secondary education[10]. To our knowledge, this is the first study that ambitions to analyze and compare the trends in schooling quality[11]. The main methodology used in our paper for analyzing trends is based on variations between two evaluations over time for each country. Instead of aggregating results from different sources, we prefer to include variations of scores when it is possible. For instance, as some countries took part to both PISA 2003 and TIMSS 2003 assessments, one can have two options for including both results. The first one is to compute the arithmetic mean of both results. However, this methodology may lead to biases in the case of countries which took part to only one assessment, while other took part to both assessments. The alternative methodology is to base the anchoring on IEA assessments and to include only trends from other assessments like PISA. This methodology provides results quite similar to the ones found in international reports[12]. We preferred to choose this second option.

Dataset 2. Cross section dataset. The second dataset consists of a cross-section database where we compute the mere arithmetic mean of the different scores obtained by countries at each educational level. For example, if a country has taken part in three different assessments concerning secondary education, its mean score for the three tests will be calculated. This procedure may be used to obtain indicators of quality of student achievements (IQSA) at both primary and secondary levels (IQSA-PRI and IQSA-SEC, respectively for the primary and secondary education levels) as well as to avoid some inconsistencies in student achievement levels from one test to the other. By calculating the mean of a country's scores for all adjusted assessments, it is possible to evaluate its mean performance for the period 1964-2010, without

[10]While we combine both mathematics and science scores, we take into account only growth rates of scores in science for countries which did not took part to an evaluation in mathematics. Therefore, the scale always refers to mathematics, while for some countries trends are taken from science results. This is mainly the case for FIMS, SIMS, FISS and SISS achievement tests. Further assessments included both mathematics and science.

[11]Although Hanushek and Woessmann (2012) propose to use a similar analysis, their study is only graphical and very few information is available regarding to the amplitude of trends in schooling quality.

[12]For a few number of countries, we observed an opposite variation between TIMSS and PISA trends. As our methodology is based on IEA assessments, trends from TIMSS assessments are used in priority.

relying solely on one survey that might overestimate or underestimate pupils' actual achievement level.

The end-result of these computations is a new international database on pupil achievements in several key competencies that allows both cross-country comparisons as well as international comparisons of trends of students' achievements. We provide the results in the Table 2. The full dataset can be downloaded at the following link: http://www.beta-umr7522.fr/Datasets

Although our approach tries to diminish as most as possible measurement error, it may still exist due to specific issues related to our methodology. We highlight below three main issues with the methodology used in our paper. Firstly, the nature of the skills assessed may differ from survey to survey, in which case the surveys may not be readily comparable. While some assessments (such as the IEA surveys) tend to measure learning achievement in terms of knowledge, others (such as the OECD surveys) focus more on the level of pupils' skills. In this particular case, any equivalence established may be distorted, as the "products" of education measured in this way are not clearly equivalent. As in our study, we only focus on assessments based on the knowledge dimension, this bias is altered.

Moreover, the content of the tests is not strictly equivalent, which may bias cross-country comparisons when assessments are anchored. For example, IEA's TIMSS test in Mathematics included the same domains tested in SACMEQ. The domain are however different compared to PASEC tests and LLECE-SERCE tests. Therefore, the anchoring supposes that a pupil who performs well in one assessment will probably perform in the same way in another assessment, regardless to the exact content of tested skills. Despite this difference in the kind of "acquisition" measured, the surveys provide a sound assessment of pupils' attainment in mathematics, science and reading.

Another limit concerns the difference concerning the target population. As shown in section 2, the grade tested differs between anchored assessments. Despite the fact that tests are adapted for each evaluated grade, comparing assessments in different grades may distort the cross-country comparisons (Saito, van Capelle, 2009). For instance, PASEC tests evaluate grade 5 pupils whereas PIRLS focuses on Grade 4 pupils. On the contrary SACMEQ and LLECE-SERCE assess Grade 6 pupils. Since, each test prepared adapted items for each grade, cross-country comparisons within each assessment are possible. As the reference grade is Grade 4, all regional assessments are made in another grade (Grade 5 or Grade 6). Our paper supposes that the performance in one grade for a country (i.e. Grade 5 or Grade 6) would give the same level in the reference grade (here Grade 4). Unfortunately, there is no specific research in this area which compares regional student assessments and international student assessments. However, some recent research comparing results from TIMSS and PISA showed that ranking of countries are quite similar, despite the difference of the target population (Brown et al., 2005; Wu, 2010).

In the following sections, we make use of our database to answer several key questions concerning the international evolution of schooling systems and the link with the economy. We successively address the issue of the evolution of schooling quality, both in the short run

and the long run (trends). We compare our results with the ones obtained in the literature using different proxies. We highlight the peculiar benefits of using our improved indicators.

3. Analysis of trends on schooling quality

3.1. International comparison on schooling quality

Before examining trends in pupils performances over time, it is interesting to study possible differences in the average countries performance in international learning achievement tests. As explained above, a single indicator of countries performance is obtained by computing the mere arithmetic mean of the different scores obtained by countries at each educational level. This procedure may be used to obtain indicators of quality of student achievements (IQSA) at both primary and secondary levels (IQSA-PRI and IQSA-SEC, respectively for the primary and secondary education levels) as well as to avoid some inconsistencies in student achievement levels from one survey to the other. By calculating the mean of a country's scores between 1965 and 2010, it is possible to evaluate its mean performance for the last 46 years, without relying solely on one survey that might overestimate or underestimate pupils' actual achievement level. However, it should be noted that available data change from country to country, and for some countries, we only have one single score. The database contains information on primary education in 103 countries/areas and on secondary education in 111 countries/areas (see Table 2 for this database).

Figures 1 and 2 present the measured evolution of the quality of learning achievements among regions. Asian countries seem to outperform countries from other regions in the primary level. Especially, Korea, Kazakhstan, Chinese Tapei and Singapore are ranked at the top performing countries/areas, followed by Finland, Russian Federation and Hong-Kong China. It should be noted that pupils in Kazakhstan, the Russian Federation, Lithuania, and Latvia have demonstrated high level of achievement. It may thus be concluded that, among the middle income countries, those in Eastern Europe and Central Asia perform the best. This could be linked with the communist past of those countries. Such regimes tended to heavily weight the investments in human capital (as other investment in public goods). Other regions perform less than developed economies and Asian countries. A strong variability in performance can be observed in Arab states, where the top performing country/area is Dubai, while the less performing country is Mauritania. In contrast, Cuba performs very well in primary education. Cuba's performance is higher than a lot of developed economies, such as Spain or Israel. Most sub-Saharan countries have a low performing level in primary education, but the scores varies greatly, from 180 points (for Niger) to 330 points (for Cameroun). It should be noted that we based our analysis on the results of the global performance of a country, regardless to its composition. Therefore, if a country tends to attract most able students around the world, it will have a high performance in our methodology. This is exactly the same "bias" which occurs for university rankings. For instance, PISA 2009 results showed that approximately 15% of pupils from Singapore have an immigrant background, while this proportion is more than 40% in Qatar. Therefore, it could be possible that immigrant students from Singapore may increase the mean score of Singapore. However, there is no significant difference in performance between students with and without an immigrant background in Singapore,

based on PISA 2009 study (OECD, 2010, p.69). Moreover, a recent OECD study (OECD, 2013a) showed that in Qatar, Dubai, Macao-China and Singapore, recent immigrants tend to perform better than older immigrants.

Analysis of the countries performance in secondary education shows slight modifications compared to the primary education ranking (see Figure 2). The developed economies in Asia perform better, whereas those in Eastern Mediterranean Europe – Cyprus and Greece – score the lowest. Results from Shanghai of China demonstrate the high ability of a part of Chinese people to perform well in secondary schools. The performance level of developing countries in terms of learning achievement are again highest in Eastern Europe. The top four countries are all in this region, namely Estonia, Czech Republic, Hungary and Slovenia. This tends to confirm our assumption concerning the policy preferences inherited from a communist past. Conversely, most countries with the lowest scores for secondary education are in Africa (Ghana, Botswana, South Africa). Moreover, India's very low score should be put in perspective given that the country has not taken part in international surveys since 1970. Only two Indian states took part at PISA 2010 (Tamil Nadu and Himachal Pradesh). Adjusted results for these states are quite similar with sub-Saharan African countries.

Besides the mean scores for the whole population within each country, our database provides additional data for each gender (Female and Male) and each type of geographical location (Urban and Rural). A summary of these distinction can be found in Figures 3 to 6. Figures 3 and 4 present differences in schooling performance between Female and Male in primary and secondary education. Countries which are above the line are the ones where Female tend to outperform Male. In all Arab states where data is available, Male tend to perform better than Female. The difference between each gender is quite important in Sub Saharan African countries but significant differences occur within the region. In developed countries, the difference of performance between Male and Female is quite lower compared to developing countries. These results tend to confirm recent findings on gender gap research on educational attainment. Despite the benefits of female education in developing countries there are large gender differences in educational attainment in favor of males especially at the secondary and tertiary levels (OECD, 2011). There is also persistent gender gap in primary education in the Middle East and North Africa, South Asia and sub-Saharan Africa (World Bank, 1999). This contrasts with most industrialized countries where females have higher educational attainment than males (Pekkarinen, 2012). As the income levels rise across countries, gender gap against females shift from lower level of education to higher level and eventually the gap reverses (World Bank, 2012). This hypothesis is confirmed in some Arab states, like Algeria or Morocco: while we observed an advantage for Male students in primary education, it shifts toward Female student in secondary education.

Figures 5 & 6 present the difference of performance between Urban and Rural areas inside each country/economy. In most countries, pupils from urban areas tend to perform better than pupils from rural areas. This is especially the case for Sub Saharan African countries as Comoros and Togo, where the difference is higher than 60 points between the two groups. In developed economies, the difference is the highest in Portugal and Israel, where in a small number of countries, rural areas tend to perform better (England and Germany), but the

difference is much smaller (less than 35 points). In secondary education, pupils in rural areas from Belgium and England have a better performance compared to pupils in urban areas. However, in most countries, pupils from urban areas perform better than pupils from rural areas. Some recent research showed that among OECD countries, a performance gap between urban and rural schools can be observed in 19 countries and is widest in Mexico and Portugal (which confirm our findings). In general, students who attend schools in urban settings come from more advantaged socio-economic backgrounds. However, this is not the case of students from Germany or United Kingdom (based on PISA 2009 results). In particular, it appeared that in United Kingdom, urban areas differed from rural areas about the student-teacher ratio (Heath, A., Kilpi-Jakonen, E., 2012). More teachers are allocated in rural areas compared to urban areas. Therefore, one possible explanation of the difference of performance of pupils from rural areas can be attributed to differential student-teacher ratio. Concerning Germany, one possible factor is material educational resources which are more present in rural than urban schools (OECD, 2013b). Another important result is the lack of difference about socio-economic background of pupils between urban and rural areas in these two countries, where this is the case in most OECD countries. More research explaining these differences are needed. Below, we discuss about long term trends on quality of student achievement.

3.2. Long term trends (1965-2010)

It appears crucial to deeply analyze the global trends in indicators of quality of student achievement (IQSA). As it has been highlighted in the introduction, comparable data relative to schooling quality is very difficult to obtain. When we analyze the nature of the currently available data on schooling (such as the Barro and Lee database released in 2010), we must recognize that shortcomings are numerous, especially concerning the quality dimension (even if one has to recognize in parallel the pioneering role of this paper). It seems therefore important to obtain alternative indicators which partially take into account the qualitative dimension of education.

On top of that, we propose to analyze the global trends in schooling quality, by creating new scores computed by combining mathematics and science scores for secondary education[13]. To our knowledge, this is the first study that ambitions to analyze and compare the trends in schooling quality[14]. The importance of stressing mathematics and science results is quite usual in economic literature as it has bee shown that such performance greatly affects growth and competitiveness potential of nations (Hanushek and Kimko, 2000). Table 3 provides summary statistics to each year and level of schooling concerning the IQSA. As we can see in Table 4, comparable data are available for very few countries from 1980 to 2007. While only 13 countries/areas have comparable data in the long run (from 1970 to 2007), our database include 24 countries/areas for which we have been able to obtain comparable data between 1980 and 2007. Countries are ranked by the global variation of their schooling quality between 1980 and 2007.

[13]This panel data is only possible by combining science and mathematics subjects. This is the reason why we do not present separate trends for each subject.

[14]Although Hanushek and Woessmann (2012) try to make similar analysis, their study is only graphical and very few information are available regarding to the amplitude of trends in schooling quality.

We firstly detect a global convergence movement between countries in terms of schooling quality since the 1980s. For instance, whereas Hungary was ranked higher than Finland in 1980 (respectively 641 points and 513 points score), these two countries tend to converge in terms of their quality of secondary education across time[15]. Indeed, in 2007, the quality of secondary schooling is approximately equal to 577 for Hungary and 574 for Finland.

Secondly, Figure 7 presents the long term trends on schooling quality for selected countries. for most countries, a significant growth in schooling quality is present (for instance for Chile, Hong-Kong and Luxembourg), a global decline of schooling quality can be observed in countries like France. While the level of schooling quality increased between 1965 and 1980, it remained quite stable between 1980 and 2000, but dropped significantly in the last decade. Indeed, the global decline between 1980 and 2007 for France is approximately equal to -5%. Other countries as Thailand or Hungary appear to have a decline of schooling decline between 1980 and 2007. On the opposite side, Singapore, Korea and Hong-Kong have strongly improved the quality of secondary schools since 1980. The global increase of schooling quality is equal to 30% in Singapore, which represents an average increase of about 1% each year. This growth is three times higher than the ones found in the USA.

3.3. Short term trends (1995-2010)

Due to the scarcity of data for the long run, we propose to undertake a similar level and convergence analysis for the short run. We follow the same methodology as in the long run analysis. We however use more recent achievement tests (between 1995 and 2010). Our database includes trends on student achievement from 1995 to 2010 for 105 countries/areas, whereas absolute scores for 150 countries/areas are provided. It may be interesting to compare the variation of schooling quality since 1995 and identify which countries have improved their level of quality (i.e. level of pupil achievements).

Table 5 presents extensive results for the short term trends in schooling quality. We compile all variations in order to check for robustness of trends in schooling quality for countries that participate in several tests. However, it should be borne in mind that the period between two participations for a country is not always equivalent to other countries. For instance, the United States took part in 1995, 1999, 2003 and 2007 to the TIMSS survey. It is therefore possible to compute growth rates of schooling quality between 1995 and 2007, between 1999 and 2007, and between 2003 and 2007. Other countries, such as France took part to the PISA survey between 2000 and 2006. Here, the interval is shorter compared to the 1995-2007 period for USA. For controlling these differences, we computed the average annual growth rate (AAGR) of schooling quality. While periods differ between countries and achievements tests, it is then possible to compare trends in schooling quality. When we have several cases to compare trends for the same country, we compute a mean of the AAGR. First, we distinguish between the primary and the secondary education (columns 1 to 4 in Table 5). Then, we aggregate the two levels in order to check for the global variation (columns called "Primary + Secondary"). For instance, 32 possible comparisons over time are available for

[15]On the issue of convergence between East and West-Europe since the 70s in educational terms, see Demeulemeester and Rochat, 1997.

Latvia. If we compute the mean of all 32 AAGR, we obtain an average annual growth rate equal to 0.54% with a standard error (SE) of 0.11.

In order to classify short term trends, we adopt the following classification. We define two criteria: the first one is the absolute amplitude of the AAGR. If it is higher than 0.5, then we define it as high change (++ or --). If the effect is between 0.2 and 0.5, the amplitude of the effect is intermediate (+ or -). However, in order to correct for measurement error, we add a second condition to the intensity level of the growth: we accept the idea according which when the mean AAGR is at least 2 times higher than its SE, the effect can be considered as trustable. For instance, for Qatar, while the mean is equal to 4.6, its SE is only equal to 0.94, which is a more than 4 times lower. Indeed, we can say that the effect is trustable (although we have only 3 observations). Hence, regardless to the AAGR, when the ratio AAGR/standard error is lower than 2, we preferred not to accept a clear trend toward an increase or a decrease ("o").

Data on short term trends is available for 105 countries/areas. 1 404 different variations have been calculated in order to obtain these trends. Hence, there are approximately 13 different series for each country. For developed countries/areas, such as Hong-Kong of China, Hungary, there are 42 series of variation, which can allow us to increase the probability to conclude for an increase or a decrease of the quality of schooling. There is a significant increase in schooling quality for 26 countries/areas, while we observe no significant evolution for 60 countries/areas and a significant decrease for 19 countries/areas. Countries where the mean AAGR of quality of schooling is the highest are mainly countries/areas with a low level of educational quality. A strong increase is observed in Qatar (mean AAGR equal to +4.60%), Syrian Arab Republic (+2.52%) and Kyrgyzstan (+2.08%). The evolution is higher than +0.5% for 17 countries/areas, including only 2 developed countries (Luxembourg and Portugal). On the contrary, a significant decrease of schooling quality is present for 19 countries/areas. The decrease is higher than -0.5% for 11 countries/areas. All these countries are developing countries, excepting United Kingdom and Canada (Quebec). Despite an initial low level, schooling quality tends to decrease in some Sub Saharan African countries, such as Cameroon, Burkina Faso and Madagascar.

The comparison of quality of education (measured from student achievement scores) between primary and secondary levels tends to show a high correlation: when we observe an increase in the primary education quality, trends in secondary level show an increase in quality too. While the primary education quality is increasing in countries like Slovenia, Hong-Kong China or Singapore, the quality of secondary schools tend to remain stable. It is quite interesting to note that when primary education quality declines, the quality of secondary schools shows a significant decrease too. The only country where the variation of performance is opposed between primary and secondary education is Morocco. Indeed, a high decrease of performance is observed in primary schools (-0.91%) while secondary schools tend to improve their quality (+2.18%).

Moreover, in Figures 8 and 9, we compare the initial level of IQSA in each education level with its average growth rate during the period 1995-2010. The main idea is to test to what extent a convergence is present regarding the schooling quality. Results for primary education

do clearly reject any convergence between countries. Indeed, while some countries had higher education quality in 1995, their growth rate is higher than other countries with initial low education quality (for instance Singapore versus Côte d'Ivoire). In secondary education, results tend to indicate some convergence between countries. For instance, while Luxembourg performed less than Japan in 1995, its performance increased during the period 1995-2010, on the contrary of the case of Japan, where the performance slightly decreased in secondary education. However, this convergence hypothesis should not hide that some countries with initial low performance tend to decrease their performance over time (as it is the case of Iran). Hence, it could be useful to analyze to what extent there is a trade-off between quantity and quality among developing countries.

4. Comparisons with alternative measures of human capital stock

In this section, we compare our indicators of student achievement with alternative measures of quality used in the literature and traditional measures of education. In our paper, we build indexes of schooling quality on the basis of the international and regional standardized measures of pupils achievements (i.e. cognitive skills).

4.1. Alternative measures of schooling quality

Three major research papers propose an alternative measure of human capital, schooling or pupils achievement quality[16]. Firstly, Hanushek and Kimko (2000), concerned with finding a better measurement of human capital quality of the workforce, measure it using scores obtained from students participating in international assessments in science and mathematics. Starting from these test scores, they construct a unique (normalised) labour force quality measure for 31 countries covering the period from 1960 to 1990. They computed a quality measure for each country's labour force using the weighted average over all harmonised test scores, where each country's weight is calculated as the normalised inverse of its standard error. They then performed a single cross-country regression for the 31 countries over the 1960-1990 period. The authors used the surveys of the IEA and of the IEAP. In total, twenty-six series of educational performances were taken into account (by distinguishing according to age, the field of competence, namely mathematics and sciences, and year). They end up with two series that are available. The first series sets the world mean on each of the tests used equal to 50. The second series adjust all scores based on the U.S. international performance modified for the national time pattern of scores on the NAEP (see Appendix B for more explanation of this methodology).

Another contribution that directly includes the measurement of the quality of education in a model of growth is the one by Barro (2001). The data used comes from the same sources as those of Hanushek and Kimko (2000). Barro however builds different aggregate indicators of

[16]Other papers deal with this topic, however, databases provided suffer from either lack of enough data or huge methodological issues. See for instance, Lee and Lee (1995), Ciccone and Papaioannou (2005) and Coulombe and Tremblay (2006).

cognitive skills than Hanushek & Kimko (2000), including namely mathematics, sciences and reading tests results. Because of the restricted number of countries for which qualitative indicators of education are available, his sample is small and involves only 43 countries. Moreover, the methodology used by Barro (2001) does not take into account for the possible differences in test variances across assessments. Finally, Barro (2001) calculated average scores without any adjustment between achievement tests.

Recently, Hanushek and Woessmann (2012) extended measures from Hanushek and Kimko (2000). They add new international tests, more countries and changed the methodology in order to make available the possibility of tracking evolution in cognitive skills. The methodology used combines the adjustments in levels (based on the U.S. NAEP scores) and the adjustment in variances (based on the OECD Standardization Group, see Appendix B for more details). They compute standardized scores for all countries on all assessments. Each age group and subject is normalized to the PISA standard mean of 500 and individual standard deviation of 100 across OECD countries. Cognitive skills are measured by the simple average of all observed math and science scores between 1964 and 2003 for each country. Standardized scores are available for 77 countries. More information on this methodology can be found in Appendix B.

The correlation between our data and database of Hanushek and Kimko (2000) is high (R^2= 0.52 ; 77 observations). We include imputed data from Hanushek and Kimko (2000) in this correlation. The comparison with the recent database of Hanushek and Woessmann (2012) is higher than the correlation with Hanushek and Kimko (2000) paper while it includes approximately the same number of countries (R^2=0.89 ; 75 observations). However, as our database is larger than other databases, some countries are not included in this comparison. Figure 4 show the correlation between our database and the database of Hanushek and Woessmann (2012). Even if a clear correlation appears, we tend to observe some significant differences, in particular for countries having a lower score for schooling quality. This is due to the fact that we take into account more assessments than Hanushek and Woessmann (2012), especially LLECE, PASEC and SACMEQ. Information used by Hanushek and Woessmann (2012) for countries such as Jordan or Swaziland are not recent and are only based on one single score of achievement tests in 1980s. On the contrary, our database include new scores for these countries (except for India[17]), based on regional achievement tests. Since, ranking of countries differ between regional and international achievement tests, performance of some countries like South Africa tend to be over-estimated in Hanushek and Woessmann (2012) dataset, compared to our dataset. The opposite effect can also be showed for countries as Peru, where participation in LLECE tends to increase its performance.

[17]As India only took part to one assessment in 1970, we preferred not to include this country in our dataset.

4.2. Data on Educational Attainment

Three major databases are available for educational attainment. The first database is from Barro and Lee (2010, 2012), which is an update of a previous paper (Barro and Lee, 2000). Their paper presents a data set on educational attainment for the population over age 15 and over age 25 at five-year intervals between 1950 and 2010 for a broad number of countries. The data set comprises at least one observation for 146 economies, of which 109 have complete information at five-year intervals from 1960 to 2000. We obtain measures of average years of schooling for all levels in every country. The main source for such data is coming from the Unesco database on educational attainment and UNDP Human Development Office. In years when censuses or surveys are not available, Barro and Lee (2012) estimate the educational attainment using enrolment rates.

Cohen and Soto (2007) have proposed a new set of data on human capital. It is originally based upon data released by the OECD for a subgroup of 38 member and non-member countries. Cohen and Soto (2007) expanded this dataset to other developing countries. The key of their methodology is to minimize extrapolations and keep the data as close as possible to those directly available from national censuses (in the spirit of the work of De la Fuente and Domenech (2002, 2006) for OECD countries[18]). The dataset consists of 95 countries and has been computed for the beginning of each decade from 1960 to 2000, plus a projection for 2010. This projection is based upon population projections by age taken from the U.S. Census Bureau website and the estimated educational attainment for the year 2000.

The World Population Program of the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria and the Vienna Institute for Demography, Austrian Academy of Sciences developed the IIASA/VID educational attainment model for reconstruction of educational attainment distributions by age and sex from 1970 to 2000 and projections to 2050. This database is downloadable from the website of Edstats[19]. We use data from 1970 to 2005 by combining two models proposed by the IIASA/VID: the Reconstruction model and model of Projections. Projections have been carried out for 120 countries. More information concerning this database can be obtained in Lutz et al. (2007) and Kumar et al. (2010).

The correlation between our adjusted students' test scores and data on schooling quantity is not very high. The correlation of our (= IQSA) with the average Net Enrollment Ratio (= NER) between 1960 and 2010 in Secondary level, for High Income Countries, is null and includes 38 observations. While some countries have approximately the same NER in secondary education, the IQSA tend to be different (Figure 12). This is for instance the case of Hungary and Greece: the former has a higher level on schooling quality, while the NER in 1995 is equal in these countries. The comparison with the database of Barro and Lee (2010) shows a positive and quite high significant correlation of 0.62 (116 observations). However, this correlation is mainly due to a "structural" effect, indicating that the higher a country is developed, the higher is its schooling quality. When we restrict this comparison to High income countries, the correlation shrink to 0.01 (38 observations, see Figure 13). As the new

[18]We do not test the correlation between our data and the data base proposed by De la Fuente and Doménech, since the database of Cohen and Soto (2007) is quite similar and include more countries.
[19]Database can be downloaded at: http://go.worldbank.org/47P3PLE940.

data set of Cohen and Soto (2007) appear to be more precise than the one of Barro and Lee (2010), we test to what extent years of schooling are correlated with our data. When we include all countries, the correlation is very high (R^2= 0.74 ; 82 observations). However, for the high income countries group, the correlation decreases to 0.15 (27 observations). We found similar results when we compare our data with the IIASA/VID database. The correlation decreases from 0.62 to 0.09 when we restrict our analysis to high income countries (37 observations against 104 for the full sample).

Hence, the absence of strong correlation between the basic quantitative indicator of education and our data on schooling quality (pupils achievements) tends to prove the importance of taking into account the qualitative dimension of education, and by the way of human capital.

Conclusion

It is now common practice to measure the performance of education systems. The findings from international surveys of learning achievement are increasingly being used to measure the quality of education. Yet they provide a solid foundation with which to do so.

The purpose of this paper was to build a new index for the measurement of learning achievement by grouping together all existing international and regional learning assessments. Where other research has focused solely on one survey, this paper has sought to group them together and thus obtain a more comprehensive range of data on pupils' performance. The methodology used is based largely on countries' participation in several surveys simultaneously, thus permitting the computation of inter-survey equivalence indices. Comparable indicators on learning achievement can thus be devised on the basis of a variety of surveys.

We provide comparable indicators for student achievement for a broad number of countries, both for intertemporal data (1965-2010), but also for cross-country analysis. The cross-country dataset includes indicators of quality of student achievement (IQSA) for 103 countries/areas in primary education and 111 countries/areas in secondary education. Besides the general population within each country/areas, we also provide data for each gender (Female/Male), each type of location (Urban/Rural). Moreover, following the initial idea of Hanushek and Woessmann (2012), we computed proportion of pupils reaching a low level of competency (i.e. one standard deviation below the international average – 400 points), and the high level of competency (i.e. one standard deviation above the international average – 600 points). Following, the same methodology used for mean scores, it permits us to obtain similar comparable proportions both in panel data and in cross-section data.

Bibliography

Altinok, N., Murseli, H. (2007). "International database on Human Capital Quality". *Economics Letters*, 96(2), pp. 237-244.

Barro, R.J. (2001). "Education and Economic Growth". in Helliwell, J.F. (Ed.), *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being* (pp. 14-41). Paris: OECD Press.

Barro, R.J., Lee, J.W. (1993). "International Comparisons of Educational Attainment". *Journal of Monetary Economics*, 32, pp.363-394.

Barro, R.J., Lee, J.W. (2001). "International Data on Educational Attainment: Updates and Implications". Center for International Development Working Paper no. 45, Harvard University.

Barro, R.J., Lee, J.W. (2010). " New Data Set of Educational Attainment in the World: 1950-2010", *NBER Working Paper*, No. 15902.

Barro, R.J., Lee, J.W. (2012). "A New Data Set of Educational Attainment in the World: 1950-2010", updated version provided in the link: http://www.barrolee.com/papers/Barro_Lee_Human_Capital_Update_2012Oct.pdf (14/11/2012).

Brown, G., Micklewright, J., Schnepf, S.V., Waldmann, R. (2005). "Cross-National Surveys of Learning Achievement: How Robust are the Findings?". Southampton Statistical Sciences Research Institute, Applications & Policy Working paper, A05/05.

Card, D., Krueger, A.B (1992). "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *Journal of Political Economy*, 100(1), pp.1-40, February.

Chinapah, V. (2003). "Monitoring Learning Achievement (MLA) Project in Africa", *ADEA Biennal Meeting 2003.* Grand Baie, Mauritius, 3-6 December.

Ciccone A., Papaioannou E. (2005). "Human Capital, the Structure of Production, and Growth". Barcelona, Universitat Pompeu Fabra.

Cohen, D., Soto, M. (2007). "Growth and Human Capital: Good Data, Good Results". *Journal of Economic Growth*, 12(1), pp.51-76.

Coulombe, S., Tremblay, J.-F. (2006). "Literacy and Growth", *Topics in Macroeconomics*, 6(2). Berkeley Electronic Press, http://www.bepress.com/bejm /topics/vol6/iss2/art4/.

Coulombe, S. and J.F. Tremblay (2007), "Skills, Education and Canadian Provincial Disparity", *Journal of Regional Science*, vol. 47, nr. 5, pp. 965-991, December.

De la Fuente, A., Domenech, R. (2002). "Educational Attainment in the OECD, 1960-1995". *CEPR DP 3390*.

De la Fuente, A., Domenech, R. (2006). "Human Capital in growth regression: How much difference does quality data make?". *Journal of the European Economic Association*, 4(1), pp. 1-36.

Demeulemeester and Diebolt (2011), "Education and Growth: What Links for Which Policy?", *Historical Social Research*, 36, pp. 323-346.

Demeulemeester and Rochat (1997), "Convergence versus divergence between European countries: the case of higher education systems", *Brussels Economic Review*, ULB - Université Libre de Bruxelles, 153, pp. 3-19.

Gurgand, M. 2000. "Capital humain et croissance : la littérature empirique à un tournant ?", *Économie Publique*, 6, pp. 71-93.

Hanushek, E.A., Rivkin, S.G., Taylor, L.L. (1996). "Aggregation and the Estimated Effects of School Resources", *NBER Working Papers*, 5548, National Bureau of Economic Research.

Hanushek, E.A., Kimko, D.D. (2000). "Schooling, Labor-Force Quality, and the Growth of Nations". *American Economic Review*, 90(5), pp. 1184-1208.

Hanushek, E.A., Woessmann, L. (2007). "The Role of Education Quality in Economic Growth". World Bank Policy Research Working Paper, 4122. Washington, D.C.

Hanushek, E.A., Woessmann, L. (2012). "Do Better School Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation". *Journal of Economic Growth*, 17(4), pp. 267-321.

Heath, A., Kilpi-Jakonen, E. (2012). "Immigrant Children's Age at Arrival and Assessment Results", *OECD Education Working Papers*, n°75, OECD Publishing.

Heckman, J., Layne-Farrar, A., Todd, P., 1996. "Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings," *Review of Economics and Statistics*, 78(4), pp. 562-610.

Keeves, J.P. (1992). *The IEA Science Study III: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon Press.

Kumar, S., Barakat, B., Goujon, A., Skirbekk, V., Sanderson, W., Lutz, W. (2010). "Projection of Populations by Level of Educational Attainment, Age and Sex for 120 Countries for 2005-2050". International Institute for Applied Systems Analysis, *Demographic Research*, 22, pp. 383-472.

Lazear, E.A. (2003). "Teacher Incentives". *Swedish Economic Policy Review*, 10(3), pp. 179-214.

Lee, J.W., Barro, R.J (2001) "Schooling Quality in a Cross Section of Countries", *Economica*, 38(272), pp. 465-88.

Lee, D.-W., Lee, T.-H. (1995). "Human Capital and Economic Growth: Tests Based on the International Evaluation of Educational Achievement". *Economics Letters*, 47(2), pp. 219-225.

Leuven, E., Oosterbeek, H., van Ophen, H. (2004). "Explaining International Differences in Male Skill Wage Differentials by Differences in Demand and Supply of Skill". *Economic Journal*, 114(495), pp. 466-486.

Lutz, W., Goujon, A., K.C., S., Sanderson, W. (2007). "Reconstruction of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 1970-2000". International Institute for Applied Systems Analysis, *Interim Report IR-07-002*, Austria.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research Press.

Mulligan, C.B. (1999). "Galton versus the Human Capital Approach to Inheritance". *Journal of Political Economy*, 107(6), S184-S224.

Mullis, I.V.S., Martin, M.O., & Foy, P. (2009). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Murname, R.J., Willet, J.B., Duhaldeborde, Y., Tyler, J.H. (2000). "How Important Are the Cognitive Skills of Teenagers in Predicting Subsequent Earnings?". *Journal of Policy Analysis and Management*, 19(4), pp. 547-568.

OECD (2010). *PISA 2009 Results: Overcoming Social Background: Equity in Learning Opportunities and Outcomes,(Volume II)*, PISA, OECD Publishing.

OECD (2011a), *Report on the Gender Initiative: Gender Equality in Education, Employment and Entrepreneurship*, OECD Publishing.

OECD (2013a). "Do immigrant students' reading skills depend on how long they have been in their new country?", *PISA in Focus*, n°29, June, Paris.

OECD (2013b). "What makes urban schools different?", *PISA in Focus*, n°28, Paris.

Pekkarinen, Tuomas (2012), "Gender Differences in Education", IZA Discussion Paper 6390.

Pritchett, L. (2001). "Where has all the education gone?". *World Bank Economic Review*, 15(3), pp. 367-391.

Psacharopoulos, G., Patrinos, H. (2004). "Returns to Investment in Education: A Further Update". *Education Economics*, 12(2), pp. 111-134.

Robitaille, D.F., Garden, R.A. (Eds.) (1989). *The IEA Study of Mathematics II: Context and Outcomes of School Mathematics*. Oxford: Pergamon Press.

Ross, K.N., Postlethwaite, T.N. (1991). *Indicators of the Quality of Education: A Study of Zimbabwean Primary Schools*. Harare: Ministry of Education and Culture; Paris: International Institute for Educational Planning.

Saito, M., van Capelle, F. (2009). "Approaches to Monitoring the Quality of Education in Developing Countries – Searching for Better Research-Policy Linkages", Paper based on the presentation during The International Symposium on Quality Education for All – Approaches to Monitoring and Improving the Quality of Education, Berlin, 11-12 May 2009.

Sakellariou, C. (2006). "Cognitive Ability and Returns to Schooling in Chile". Background Paper for Vegas, E., Petrow, J. (Eds.) (2008), *Raising Student Learning in Latin America. The Challenge for the 21st Century*. World Bank. Latin American Development Forum series.

UNESCO (2000). *With Africa for Africa. Toward quality education for all*. UNESCO. Paris: Human Sciences Research Council.

UNESCO (2004). *EFA Global Monitoring Report 2005: The Quality Imperative*. Paris.

Vandenbussche, J., Aghion, P., Meghir, C. (2006), "Growth, distance to frontier and composition of human capital", *Journal of Economic Growth*, 11(2), pp.97-127.

Vegas, E., Petrow, J. (2008). *Raising Student Learning in Latin America. The Challenge for the 21st Century*. World Bank, Latin American Development Forum series.

Wolf, A. (2002), *Myths about Education and Growth*. London, Penguin Books.

World Bank (1999), *Education Sector Strategy*, Human Development Network Series, World Bank, Washington, DC, 1999.

World Bank (2012), *World Development Report 2012: Gender Equality and Development*, Washington

Wu, M. (2010), "Comparing the Similarities and Differences of PISA 2003 and TIMSS", *OECD Education Working Papers*, No. 32, OECD Publishing. http://dx.doi.org/10.1787/5km4psnm13nx-en

Abbreviations

AAGR	Average annual growth rate
CONFEMEN	Conference of Ministers of Education of French-Speaking Countries
ETS	Education Testing Service
FIMS	First International Mathematics Study
FISS	First International Science Study
IAEP	International Assessment of Educational Progress
IEA	International Association for the Evaluation of Educational Achievement
IIASA	World Population Program of the International Institute for Applied Systems Analysis
IIEP	International Institute for Educational Planning
IQSA	Indicators of quality of student achievement
IRT	Item Response Theory
ISAT	International Student Achievement Tests
LLECE	Latin American Laboratory for Assessment of the Quality of Education
MLA	Monitoring Learning Achievement
NAEP	National Assessment of Educational Progress
NCES	National Center for Education Statistics
NER	Net Enrollment Ratio
OECD	Organisation for Economic Co-operation and Development
OSG	OECD Standardization Group
PASEC	Programme on the Analysis of Education Systems
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
RLS	Reading Literacy Study
RSAT	Regional Student Achievement Tests
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SIMS	Second International Mathematics Study
SISS	Second International Science Study
SSA	sub-Saharan African countries
TIMSS	Trends in International Mathematics and Science Study
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
UNDP	United Nations Development Programme
VID	Vienna Institute for Demography, Austrian Academy of Sciences

APPENDIX A: A SYNOPSIS OF INTERNATIONAL AND REGIONAL TESTS ON EDUCATIONAL ACHIEVEMENT

In this appendix, we present international and regional student achievement tests (respectively ISAT and RSAT) used in this study, in order to obtain the indicators of quality of student achievement (IQSA).

A.1. International learning assessments

The International Association for the Evaluation of Educational Achievement (IEA) was the first body to measure individual learning achievements and conduct recurrent surveys for international comparative purposes as soon as in the early 1960s. The surveys include the highly regarded "Trends in International Mathematics and Science Study" (TIMSS) and "Progress in International Reading Literacy Study" (PIRLS).

A. TIMSS. The major survey series from IEA is the "Trends in International Mathematics and Science Study" (TIMSS). The central goal of TIMSS is to assess pupils performance in both subjects and to describe the environment in which they acquired these skills. With this second objective in view, those who launched TIMSS firmly took a policy-oriented approach, since pupils' scores were correlated with the various factors involved in teaching and affecting them. Five TIMSS rounds have been held to date. The first, conducted in 1995, covered 45 national educational systems and three groups of learners[20] The second round covered 38 educational systems in 1999, examining pupils from secondary education (grade 8). The third round covered 50 educational systems in 2003, focusing on both primary and secondary education (grades 4 and 8). In 2007, the fourth survey covered grades 4 and 8 and more than 66 educational systems. The last round was performed in 2011 and covered 77 countries/areas[21]. The precise content of the questionnaires can vary quite a lot but remains systematic across countries. Each topic is given a specific weight (as for example, numbers, algebra and geometry in mathematics subject and life sciences, physical sciences and the history of science in science subject).

B. PIRLS. The other major IEA survey is the "Progress in International Reading Literacy Study", also known as PIRLS. Three major rounds of PIRLS have been held up to 2011: in 2001, in 2006 and in 2011. PIRLS survey tests pupils from primary schools in reading proficiency[22]. For instance, the 2006 PIRLS survey involved 41 countries/areas, only two of which were African countries (Morocco and South Africa). This round included 5 lower-middle-income countries (Georgia, Indonesia, Moldova, Morocco) and 7 upper-middle-

[20]IEA assessments defined populations relative to specific grades, while assessments as PISA focus on age of pupils. In IEA studies, three different group of pupils are generally assessed: pupils from grade 4, grade 8 and from the last grade of secondary education. Some Canadian provinces or states in the United States of America have occasionally taken part in the IEA surveys. For the sake of simplicity, these regions are not included in the number of countries participating in the surveys.

[21]Since the micro data is not yet released, we did not included TIMSS 2011 dataset in our database. However, the database will be updated regularly with new released data.

[22]Similarly to TIMSS, pupils from Grade 4 were chosen.

income countries (Bulgaria, Islamic Republic of Iran, Lithuania, Macedonia, Federal Yugoslavian Republic, Romania, Russian Federation, South Africa) took part in PIRLS 2006. The last PIRLS round was done together with TIMSS (2011) and included 60 countries/areas.

This paper has drawn on virtually all IEA studies in the three skills (Mathematics, sciences and reading/literacy). Data for 1964 to 1990 have been taken from major publications dealing with IEA studies and the scores for other years are taken from official reports (Harmon, Smith et al., 1997; Martin, Mullis et al., 2000; Mullis, Martin et al., 2000; Mullis, Martin et al., 2003; Mullis, Martin et al., 2004; Martin, Mullis et al., 2004; Mullis, Martin et al., 2007, Mullis, Martin et al., 2009).

C. PISA. The Organisation for Economic Co-operation and Development (OECD) is another international organisation that has carried out standardised international comparisons of pupils achievements. The OECD launched its "Programme for International Student Assessment" (PISA) in 1997 to meet the need for readily comparable data on student performance. The basic principles underlying PISA studies are the use of an extended concept of "literacy" and an emphasis on lifelong learning. Literacy is considered more largely because PISA studies are concerned with the pupils' capacity to extrapolate from what they have learnt and apply their knowledge to new settings. More generally, PISA has assessed the skills of 15-year-old pupils every three years since 2000 in countries. PISA concentrates on three key areas, namely mathematics, science and literacy. In all PISA cycles, the domains of reading, mathematical and science skills are assessed. The main focus of PISA 2000 was on reading literacy, in the sense that it included an extensive set of tasks in this domain. In PISA 2003, the emphasis was on mathematical skills and in 2006 the focus was on scientific skills. The framework of evaluation remains the same across time so that one cycle's findings can be compared with those of the others[23]. In 2009, 72 countries/areas participated, while 64 countries/areas expect to take part to PISA 2012. Unlike the IEA surveys, PISA assesses only 15-year-old pupils, whatever their school level, whereas the grade is the main criterion in selecting pupils for IEA assessments (and over all student achievement tests).

A.2. Regional learning assessments

Three major regional assessments have been conducted in Africa and Latin America.

D. SACMEQ. The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) grew out of a very extensive national investigation into the quality of primary education in Zimbabwe in 1991, supported by the UNESCO International Institute for Educational Planning (IIEP) (Ross and Postlethwaite, 1991). Keen to follow up this successful initiative, several education ministers in southern and Eastern African countries expressed an interest in the study and wished to take part in such an assessment. Planners from seven countries therefore met in Paris in July 2004 and established SACMEQ as a special group. The 15 SACMEQ-member education ministries are those of Botswana,

[23]As it is explained in the PISA 2006 technical report, this is only the case for reading between 2000-2009, for mathematics between 2003-2009 and for science between 2006-2009. See OECD (2010) and Annex B for more details.

Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, United Republic of Tanzania, United Republic of Tanzania (Zanzibar), Uganda, Zambia and Zimbabwe.

The first SACMEQ round took place between 1995 and 1999. SACMEQ I covered seven different countries and assessed performance in reading at grade 6. The participating countries were Kenya, Malawi, Mauritius, Namibia, United Republic of Tanzania (Zanzibar), Zambia and Zimbabwe. The studies, albeit mainly national in scope, had an international dimension and shared many common features (research issues, instruments, target populations, sampling and analytical procedures). A separate report was prepared for each country. In the second round, which was held between 2000 and 2002 and covered 14 countries and one territory (Zanzibar), performance in mathematics and reading was assessed. The target cohort consisted of grade 6 pupils, as under SACMEQ I. The participating countries were Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, United Republic of Tanzania, United Republic of Tanzania (Zanzibar), Uganda and Zambia.

Several SACMEQ II items were replicated from the TIMSS survey to secure comparable results. The questionnaires were used to collect information on educational inputs, the educational environment and issues relating to the fair allocation of human and material resources. Information about the socio-economic context was gleaned from the pupils' questionnaires. More generally, SACMEQ II included items selected from four previous surveys, namely the *Indicators of the Quality of Education* (Zimbabwe) study, SACMEQ I, TIMSS and the 1985-94 IEA *Reading Literacy Study*.

The third SACMEQ round (SACMEQ III) covers the same countries as in 2002 (plus Zimbabwe) and focuses on achievements levels of grade 6 pupils. Unfortunately, the micro database for SACMEQ III has not yet been released by the SACMEQ consortium. We will use the macro results of SACMEQ III in our study (SACMEQ, 2010)., besides the micro data from SACMEQ II.

E. PASEC. Surveys under the "Programme d'Analyse des Systèmes Éducatifs" (PASEC, or "Programme of Analysis of Education Systems") of the Conference of Ministers of Education of French-Speaking Countries (CONFEMEN) have been conducted in the French-speaking countries of sub-Saharan Africa. This assessment contains results for primary school performance in mathematics and in French. In both CP2 (the second grade in primary school) and CM1 (grade 5), between 2,000 and 2,500 young learners in about 100 schools, along with their teachers and school heads, have been surveyed in each of the evaluated countries. Some countries have taken part in the PASEC survey on several occasions. The following is a list of participating countries in chronological order: Djibouti (1994), Congo (1994), Mali (1995), Central African Republic (1995), Senegal (1996), Burkina Faso (1996), Cameroon (1996), Côte d'Ivoire (1996), Madagascar (1997), Guinea (2000), Togo (2001), Mali (2001), Niger (2001), Chad (2004), Mauritania (2004), Guinea (2004), Benin (2005), Cameroon (2005), Madagascar (2006), Mauritius (2006), Congo (2007), Senegal (2007), Burkina Faso (2007), Burundi (2009), Ivory Coast (2009), Comoros

(2009), Togo (2010). It should be noted that the findings of the first four assessments are not available because data relative to assessments are not available.

In order to simplify the analysis, we will consider two different rounds of PASEC: the first round includes assessments carried out between 1996 and 2003, whereas the PASEC II takes into account evaluations which have been done between 2004 and 2010. The next round of PASEC, namely the PASEC III study, is currently under preparation by the CONFEMEN. Moreover, as scores are not directly and fully comparable between each assessment, an anchoring of major items has been made in order to allow for international comparability[24].

F. LLECE. The network of national education systems in Latin American and Caribbean countries, known as the Latin American Laboratory for Assessment of the Quality of Education (LLECE), was formed in 1994 and is coordinated by the UNESCO Regional Bureau for Education in Latin America and the Caribbean. The main aim of this survey is to collect information on pupil performance and performance-related factors likely to help policy makers to design better educational policies. For this purpose, the LLECE seeks to answer the following questions: What do pupils learn? At what level is learning achieved? What skills are developed? When does learning occur? Under what circumstances does it occur? (Casassus et al., 1998).

Assessments conducted by the LLECE focused therefore on learning achievement in reading and mathematics in grades 3 and 4 in 13 countries of the subcontinent (Casassus et al., 1998, 2002), namely Argentina, Bolivia, Brazil, Chile, Columbia, Costa Rica, Cuba, Dominican Republic, Honduras, Mexico, Paraguay, Peru and the Bolivarian Republic of Venezuela (Casassus et al., 1998). In each country, samples of about 4,000 pupils in grade 3 (ages 8 and 9) and grade 4 (ages 9 and 10) were assembled. These surveys covered over 50,000 children, amounting to at least 100 classes per country. In 2006, the second round of the LLECE survey was initiated in the same countries as LLECE I. Data between the two rounds are therefore not directly comparable. Moreover, grades tested partly changed compared to the first study: pupils from grade 3 and grade 6 took part to the LLECE II study. Our analysis will include only LLECE II results, since the grade tested is the last grade in all countries.

Data sources

Altinok, N., Murseli, H. (2007). "International database on Human Capital Quality". *Economics Letters*, 96, pp. 237-244.

Burnstein, L. (Ed.) (1992). *The IEA Study of Mathematics III: Student Growth and Classroom Processes*. Oxford: Pergamon Press.

[24] We are very grateful to the PASEC team, and especially to Jean-Marc Bernard, Antoine Marivin and Vanessa Sy for their help in the providing of the data. More details concerning the adjustment of the PASEC database is provided in Appendix B.

Casassus, J., Froemel, J.E., Palafox, J.C., Cusato, S. (1998*). First International Comparative Study of Language, Mathematics, and Associated Factors in Third and Fourth Grades. First Report*. Santiago, Chile: Latin American Laboratory for Evaluation of the Quality of Education.

Casassus, J., Froemel, J.E., Palafox, J.C., Cusato, S. (2002). *First International Comparative Study of Language, Mathematics, and Associated Factors in Third and Fourth Grades. Second Report.* Santiago, Chile: Latin American Laboratory for Evaluation of the Quality of Education.

Elley, W.B. (Ed.) (1992). *How in the World do Students Read?* Hamburg: Grindeldruck GMBH.

Harmon, M., Smith, T.A., Martin, M.O., Kelly, D.L., Beaton, A.E., Mullis, I.V.S., Gonzales, E.J., Orpwood, G. (1997). *Performance Assessment in IEA's Third International Mathematics and Science Study (TIMSS)*. Boston: Boston College Press.

Husén, T. (Ed.) (1967). *A Comparison of Twelve Countries: International Study of Achievement in Mathematics* (Vols. 1-2). Stockholm: Almquist & Wiksell.

Keeves, J.P. (1992). *The IEA Science Study III: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon Press.

Lee, J.W., Barro, R.J. (2001). "Schooling Quality in a Cross-Section of Countries". *Economica*, 38(272), pp. 465-488.

LLECE-UNESCO (2008). *Student Achievement in Latin America and the Caribbean. Results of the Second Regional Comparative and Explanatory Study (SERCE)*, Unesco, Paris.

Martin, O.M., Mullis, I.V.S., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K., Chrostowski, S.J., Smith, T.A. (2000). *TIMSS 1999 International Science Report*. Boston: Boston College Press.

Martin M.O., Mullis, I.V.S., Gonzales, E.J., Chrostowski, S.J. (2004). *TIMSS 2003 International Science Report.* Boston: Boston College Press.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K., Chrostowski, S.J., Smith, T.A. (2000). *TIMSS 1999 International Mathematics Report*. Boston: Boston College Press.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kennedy, A.M. (2003). *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools*. Chestnut Hill, MA: Boston College.

Mullis, I.V.S., Martin, M.O., Gonzales, E.J., Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report*. Boston: Boston College Press.

Mullis, I.V.S., Martin, M.O., Kennedy, A.M., Foy, P. (2007). *IEA's Progress in International Reading Literacy Study in Primary School in 40 Countries.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., & Foy, P. (2009). *TIMSS 2007 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

OECD (2001). *Knowledge and Skills for Life. First Results From the OECD Programme for International Student Assessment (PISA) 2000*. Organisation for Economic Co-operation and Development. Paris.

OECD (2004). *Learning for Tomorrow's World. First Results PISA 2003*. Organisation for Economic Co-operation and Development. Paris.

OECD (2007). *PISA 2006 Science Competencies for Tomorrow's World*. Organisation for Economic Co-operation and Development. Paris.

OECD (2010). *PISA 2009 Results: Learning Trends: Changes in Student Performance Since 2000 (Volume V).* Organisation for Economic Co-operation and Development. Paris.

OECD, UNESCO Institute for Statistics. (2003). *Literacy Skills for the World of Tomorrow: Further Results from PISA 2000*. Paris and Montreal. OECD and UNESCO Institute for Statistics.

Robitaille, D.F., Garden, R.A. (Eds.) (1989). *The IEA Study of Mathematics II: Context and Outcomes of School Mathematics*. Oxford: Pergamon Press.

SACMEQ (2010), "SACMEQ III Project Results: Pupil Achievement levels in reading and mathematics", Working Document Number 1, SACMEQ.

UNESCO (2000). *With Africa for Africa. Toward quality education for all*. UNESCO. Paris: Human Sciences Research Council.

Walker, D.A. (1976). *The IEA Six-Subject Survey: An Empirical Study of Education in Twenty-One Countries.* Stockholm: Almquist & Wiksell; New York: John Wiley & Sons.

APPENDIX B: DETAILED METHODOLOGY FOR THE MAIN DATABASE

We hereafter present the methodology used to compare the main database underlying our indicators of quality of student achievement (IQSA).

Step 1: Distinction between assessments (A.1, A.2 and B)

As the methodology of anchoring differs between assessments, we allocate each assessment to three different groups of surveys (groups A.1, A.2 and B). Surveys grouped in survey series A are from IEA and OECD, while assessments from survey series B are from RSATs (PASEC, SACMEQ and LLECE). In Table 1, we present the different assessments used in our study with the classification used.

Concerning surveys A.1. and A.2., two complementary methods of anchoring can be used in order to obtain comparable scores. The first one is related with the permanent anchoring of each score with the results of USA. Since the USA took part to all international assessments, it remains possible to obtain comparable scores between assessments by anchoring the performance of this country with a national assessment. The surveys included in this group are mainly the ones which occurred until 1995. More precisely, we include all ISATs in reading until 2001, and all ISATs in mathematics and science until 1995 for IEA surveys, while surveys for PISA are included until 2000[25].

Other assessments (PIRLS [2006], TIMSS [1999, 2003, 2007], PISA [2003, 2006, 2009]) are grouped in the survey series A.2. This second group of assessments is adjusted with another methodology (presented in step 3).

Step 2: Adjustment of Survey series A.1.

The methodology used in survey series A.1 involved recurrent adjustment to the *US National Assessment of Educational Progress* (NAEP) survey in the way described by Hanushek and Kimko (2000). The NAEP has been the main instrument used to measure the learning achievement of pupils in the United States of America since 1969. In order to obtain comparable ISATs, we need information on test performance that is comparable over time. Such information is available in NAEP. At various times since 1970, pupils aged 9, 13 and 17 in the United States have been tested on their achievement in science, mathematics and reading. To our knowledge, this is the only information on educational performance that is consistently available for comparisons over time. Moreover, the United States is the only country that participated in every ISAT. Based on the Hanushek and Kimko (2000) methodology, we can scale the level of each ISAT relative to the comparable test performance

[25]In the case of mathematics, we also include PISA 2003 survey, while in science, PISA 2003 and 2006 surveys are included too, since the PISA datasets does not allow direct comparability of scores between 2000 and 2003 for mathematics and between 2000 and 2006 for science.

of the United States, by using the time series evidence on the performance of U.S. students. More specifically, the procedure involves tracking the level of "difficulty" of a survey in accordance with the successive scores recorded in the United States of America.

The example below shows how achievement scores in mathematics under TIMSS 1995, in which the United States of America took part, can be anchored to the NAEP score (Table 1). The United States of America tested its own pupils under the NAEP mathematics survey. The score of the United States of America was 548 points under NAEP but 492 points under TIMSS. This differential can be used to adjust the TIMSS survey to the NAEP benchmark. Table B.1. shows the scores of five countries under TIMSS and the calculation used to adjust them to the benchmark.

Table B.1 : Example n°1 - Adjustment of TIMSS 1995 based on NAEP assessment

Country	TIMSS 1995	NAEP	Adjustment	TIMSS 1995 adjusted
Australia	509		$509 \times (548/492)$	567
Canada	521		$521 \times (548/492)$	580
Japan	581		$581 \times (548/492)$	647
Norway	498		$498 \times (548/492)$	555
United States	492	548	$492 \times (548/492)$	548

Another example is PISA 2000 in which the United States of America also took part (see Table B.2). The idea is similar and aims at determining the difference between the score of the United States of America under PISA and NAEP respectively. As PISA 2000 seems to underestimate the level measured by the NAEP, the PISA 2000 results have been increased by an adjustment coefficient of 518/505 = 1.03.

Table B.2 : Example n°2 - Adjustment of PISA 2000 based on NAEP assessment

Country	PISA 2000	NAEP	Adjustment	PISA 2000 adjusted
Australia	528		$528 \times (518/505)$	542
Canada	534		$534 \times (518/505)$	548
Japan	522		$522 \times (518/505)$	536
Norway	505		$505 \times (518/505)$	518
United States	505	518	$505 \times (518/505)$	518

The same methodology has been applied to all assessments in which the United States took part and for which this is not possible to directly compare trends over time. Hence, following assessments are included in this group: FIMS, FISS, RLS, TIMSS 1995, PIRLS 2001 and PISA 2000 for reading. However, it should be noted that PISA scores are not directly comparable in mathematics between 2000 and 2003, while this is the case only between 2003 and 2009. Similarly for science, scores are not comparable between 2000 and 2006. Indeed, PISA items differed greatly between 2000 and 2003 in mathematics. This is the

reason why the anchoring process for mathematics and science is based on results from both PISA and TIMSS assessments. By comparing countries which took part to these assessments, our methodology avoids the problem of potential difference of performance between assessments (this methodology take the same approach of the one presented in step 4).

Step 3: Adjustment of survey series A.2.

If recent assessments such as PISA 2009 or PIRLS 2006 were to be adjusted according to the above procedure, all survey scores would be based on scores obtained in the USA. However, recent surveys have been designed to allow analysis of country across time. Pupils are given test pieces of the same level of difficulty in survey series conducted over the years, which makes it possible to analyse trends in pupil performance directly over time. Any adjustment of the series A.2 assessments to the NAEP survey may thus distort the analysis of country performance trends, in comparison with official trends published in PISA or TIMSS Trends Reports. For example, if the United States of America score increases in the NAEP survey but decreases in another survey, such as PISA, the adjustment may lead to a fairly significant distortion. Yet the level of difficulty in pupil performance assessments may vary significantly from one type of survey to another – for example, marking under TIMSS may be stricter than under PISA. Any adjustment should thus be such as to result in comparable scores under both types of survey.

The same adjustment coefficients as those calculated for the series A surveys are used in order to achieve a single linear conversion of country scores, and this procedure does not compromise the comparability of the scores obtained by countries participating in the same survey series.

As highlighted above, this main difference with the methodology used by Hanushek and Woessmann (2012) allows us to avoid a potential bias in estimating trends in pupils performance for countries for which scores are already comparable over time. As our anchoring methodology is a simple linear transformation of surveys, the trends of schooling performance observed in PISA reports are still present in our database (see OECD, 2010)[26]. As it is shown in section 4.1., our dataset therefore shows some differences with the dataset provided by Hanushek and Woessmann (2012).

Moreover, another difference with our previous publication concern the information relative to trends between FIMS and SIMS and between FISS and SISS, published in IEA Reports (see Keeves, 1992; Robitaille and Garden, 1989). This additional information permitted us to simply adjust assessments on NAEP anchoring, but also to reproduce the trends found in these reports. Since, a significant number of items were used in both assessments, the trends found in these reports may reduce the bias which occurs when we only use NAEP anchoring.

[26]However, a problem occur when for some countries, we detect a big difference in trends between IEA and PISA assessments. For instance, in can be possible that for a low number of countries, the performance of pupils increased in TIMSS while we observed a decrease in PISA. Instead of merging both variations - which would lead to a stagnation of score - we prefer to focus primarily to IEA results.

Step 4: Adjustment of survey series B

For the last group of surveys (called « Series C »), we cannot use a simple anchoring method on the NAEP, since the United States did not participate to any regional assessment. We selected countries which participated in at least two different surveys so as to establish a comparison between the surveys. The IEA surveys have been chosen as reference surveys because they cover most of the countries and as the economic levels of the participating countries are the most heterogeneous.

As some countries took simultaneously part to several assessments, we can assume that the differences found between assessments are exogenous to the performance of these countries. For instance, if Colombia performed at a level of approximately 400 points in LLECE II assessment and obtained 321 points in TIMSS 2007 in mathematics, it is possible to suppose that independently to the performance of Colombia, the LLECE II study over-estimated the performance of its participating countries of about 24% ([400/321 – 1]×100), compared to the level of complexity in TIMSS 2007 assessment. Therefore, in order to adjust the LLECE II assessment to the TIMSS test, we need to correct this over-estimation. This is the main methodology used in order to anchor each assessment with the IEA assessment (considered here as the reference assessments). Firstly, we proceed with the adjustment of the LLECE II study to anchored IEA studies. Concerning mathematics, it is possible to anchor LLECE II assessment by adjusting its level with TIMSS 2007 assessment. We use the participation of Colombia and El Salvador in both tests in order to make this anchoring. The mean performance of these countries in LLECE II is equal to 491 points, while in TIMSS 2007, their performance declines to about 309 points. A linear transformation of LLECE II results in order to take into account this over-estimation is then made[27]. The anchoring process for reading is based on PIRLS 2001 study. As Argentina and Colombia took part to both LLECE II and PIRLS 2001 study, the same methodology is used to anchor the LLECE II assessment in reading[28]. Concerning this anchoring, we use both grade 3 and grade 6 data in order to test for a specific "grade-effect". As tested grades differ between PIRLS and LLECE II, we try to adjust independently both grades. We do not find a strong modification in countries performance between grade 3 and grade 6. Indeed, there is no strong bias relating to differences in grade level tested between assessments.

The next step consists of anchoring SACMEQ results with IEA assessments. For the anchoring process in reading, we use results from South Africa in both SACMEQ III and PIRLS 2006 database, as this country took part to both assessments. By computing the anchoring coefficient between SACMEQ III and PIRLS 2006, and having in mind that the SACMEQ scores are comparable over time, it becomes possible to anchor all SACMEQ rounds with PIRLS database. In mathematics, it is possible to indirectly adjust SACMEQ scores with countries which took part to the TIMSS 2007 assessment. In SACMEQ database, already adjusted data with TIMSS is available. However, using directly these scores may

[27]SERCE scores are multiplied by 309/491 points. This represented a decrease of about 37% of SERCE performance.

[28]As PIRLS 2001 and PIRLS 2006 assessments have been already adjusted in the first step, it is not needed to make another specific anchoring. This would have been better to anchor on PIRLS 2006 data, since the SERCE study have been conducted in 2006. However, there is no Latin American country in PIRLS 2006 test.

over-estimate the overall performance of SACMEQ countries, due to the difference of grades tested between SACMEQ and TIMSS (grade 6 for SACMEQ and grade 4 for TIMSS). Hence, we also used as anchor country Botswana that participated in both TIMSS 2003 and SACMEQ II assessments as adjustment of the differences between grades[29]. As Botswana took part to both TIMSS 2003 and SACMEQ II study, we adjust the difference of the two scores in order to rescale the SACMEQ database. After having taken into account the anchoring in the first step, the performance of Botswana declines from 512 points to 298 points (which represents a decline of about 40% of original SACMEQ performance level)[30]. The anchoring process with other SACMEQ assessments is not needed since the SACMEQ study provides comparable data between each round. Therefore, we use the same coefficient found in each skill with all SACMEQ rounds.

The last step concerns the anchoring of PASEC data. We base our methodology by anchoring the PASEC test with the adjusted SACMEQ test by comparing results of Mauritius. As this country took part to both assessments (in 2006 for PASEC and 2007 for SACMEQ), we anchor PASEC assessment by adjusting the scores of countries which took part in PASEC with the difference of performance of Mauritius between PASEC and SACMEQ. As PASEC scores are based on a scale between 0 and 100, the transformation of these scores in order to adjust them on the SACMEQ scale is equal to 5.97 in reading and 8.46 in mathematics[31]. This anchoring gives the same performance level for Mauritius in both PASEC and SACMEQ tests. As PASEC is an assessment for Francophone countries, while SACMEQ focuses on Anglophone countries, the anchoring process may lead to biases in adjusted reading scores. However, Mauritius has been tested in both languages in PASEC (English and French)[32]. By using both scores for Mauritius in PASEC, it gives us the adjusted coefficient of anchoring between PASEC and SACMEQ in reading. Similarly to SACMEQ, since PASEC survey permits over-time comparability after some adjustment[33], we use the same coefficient of adjustment between the two rounds of PASEC for each skill.

Step 5. Panel database

The main objective of our paper is to build a cross-country and over-time comparable dataset for schooling performance. The adjustment made on previous steps allowed us to obtain

[29]This adjustment was made with a prediction of grade 4 pupils' results from grade 8 pupils' results for developing countries only, based on TIMSS results.

[30]It should be noted that the score of 298 points is the adjusted scores of Botswana in TIMSS 2003 assessment, and not its direct score. As we corrected all IEA studies with National Assessment of Educational Progress (NAEP) anchoring, this mean performance has to be corrected too. It changes from 366 to 298 points in the case of Botswana. Therefore, this performance level becomes comparable with countries' performance in TIMSS 2007 study and all other TIMSS datasets.

[31]It would have been possible to begin by standardizing the PASEC database as other assessments, which would give another adjustment coefficient. However, we preferred to keep it intact before adjustment. As we use a linear translation, both methods would give the same results.

[32]Since the results of Mauritius are over-estimated of about 9% in French regarding to English language, we chose to change the coefficient of adjustment of both mathematics and reading with this coefficient.

[33]We rescaled all PASEC datasets by anchoring with common items in each skill. For instance, concerning reading, we included only 24 items for the adjustment, whereas 33 items are present inside the test. For mathematics, 35 items were selected, among 37 items. Since, the items are similar between the countries tested, the PASEC scores become comparable over-time and between countries.

comparable scores for each skill (mathematics, science and reading) and each level of education (primary and secondary). However, merging all this data without any adjustment may generate severe bias. Indeed, it may exist differences relative to assessments which may generate different scores for each country. For instance, while the performance of a country may be high in TIMSS, its performance can be lower in PISA, in comparison to other countries. Our methodology consists of using all available information on performance for each country in order to obtain a panel dataset between 1964 and 2010.

Most assessments evaluate mathematics and science. Moreover, reading competence appear to be less correlated with science and mathematics. The main explanation refers to differences in languages spoken in each country, which may distort the evaluation of reading competence.

Our adjustment is firstly based on countries which took part to FIMS, SIMS and TIMSS assessments. However, some countries took part to FISS and SISS without participating to FIMS and SIMS. In order to take into account their results in our dataset, we predict the performance of these countries by regressing their scores in FISS on FIMS. Hence, the constant in the regression allows us to avoid a bias due to the potential difference between the two subjects. Moreover, we estimate the score of countries which took part to both FISS and SISS based on the variation of their performance, instead of only the level. This methodology is better than to compute only predictions for scores in science every year, because it allows us to better evaluate the trends in schooling performance over time. The data for trends between FIMS and SIMS comes from Robitaille and Garden (1989) while the data concerning FISS and SISS can be found in Keeves (1992). For countries which took part to only one science assessment without any mathematics assessments, we restrict our adjustment to a simple prediction of scores on mathematics. The idea behind taking into account both in mathematics and science trends relies with the possibility of increasing the number of countries and the number of score points over time, without altering the quality of the data.

For a very low number of countries, our adjustment uses the same methodology in two steps by using assessments in reading (PIRLS 2001 and PIRLS 2006). Indeed, a number of countries took part to PIRLS assessment, without a participation to TIMSS surveys. Similarly to science adjustment, we firstly predict scores for countries which took part to both PIRLS 2001 and TIMSS 2003, then we compute the countries' performance by using the growth rate between PIRLS 2001 and PIRLS 2006 (instead of predicting the PIRLS 2006 scores based on the TIMSS 2007 dataset).

The anchoring with PISA 2000 dataset was made with the same methodology. Scores from PISA 2000 were predicted in mathematics with TIMSS 1999 assessment. For countries which took part to both TIMSS 2003 and TIMSS 2007, we used the growth rate of scores between PISA assessments for predicting schooling performance. When a country took part to both PISA and TIMSS assessments, we used only results from TIMSS. It allows us to avoid the possible contradiction of schooling performance between PISA and TIMSS. For instance, Canada took part to both PISA 2003 but not to TIMSS 2003. Firstly, we adjusted results of Canada by using the NAEP anchoring. The performance score of Canada changed from 533 to 612.59 points. In a second step, we predicted results of PISA 2003 with TIMSS 2003 results. The score of Canada is then equal to 612.60, since the PISA 2003 NAEP adjusted

assessment slightly over-estimated countries performance, compared to TIMSS 2003 NAEP adjusted assessment (i.e. by about 1.69% in the case of Canada). When it is possible, PISA trends are directly used. Indeed, PISA assessment permits over-time comparability for reading between 2000 and 2009, for mathematics between 2003 and 2009 and for science between 2006 and 2009[34].

Concerning grade 4 assessments, the same adjustment was made. Since regional assessments are based on primary schooling evaluation, adjusted surveys were merged with IEA surveys (TIMSS and PIRLS). For a number of countries, we only have a participation in PIRLS without data for TIMSS. In order to take into account the potential difference of evaluation between reading and mathematics, we firstly predicted PIRLS 2001 on TIMSS 2003 (grade 4), and then scores for 2007 were based on the growth rate of scores between PIRLS 2001 and PIRLS 2006 (instead of a prediction of PIRLS 2006 based on TIMSS 2007 Grade 4 datasets).

The panel dataset for primary and secondary education can be found in the Excel spreadsheet named "Table 2".

Step 6. Cross country dataset

It may be of interest to compare countries' average performance in international and regional surveys. First, countries' average score for all surveys at the same educational level is computed. Next, each country's average score in each skill and in all primary education surveys is calculated. The same procedure is carried out for secondary education and each skill.

We eventually obtain 6 different series of cross country data for educational achievement, since we distinguish for the level of education (primary versus secondary) and the skill evaluated (mathematics, science and reading). The mean score in each level is then computed by averaging scores from each skill.

Moreover, we aggregated results from each education level in order to obtain a single general score of schooling performance for each country. Since, the dataset for each level is unbalanced, we firstly predicted scores for all countries by using all the existing information (i.e. with general scores), and then we obtained the total scores for education (primary + secondary).

The dataset can be found in the spreadsheet named "Table 1" and in Annex B.

[34]Given the fact that some countries took part to the PISA 2009 study in 2010, their results have been adjusted for 2009 by predicting their performance level in 2010.

Figure 1. Average primary school quality in the World (1964-2010)

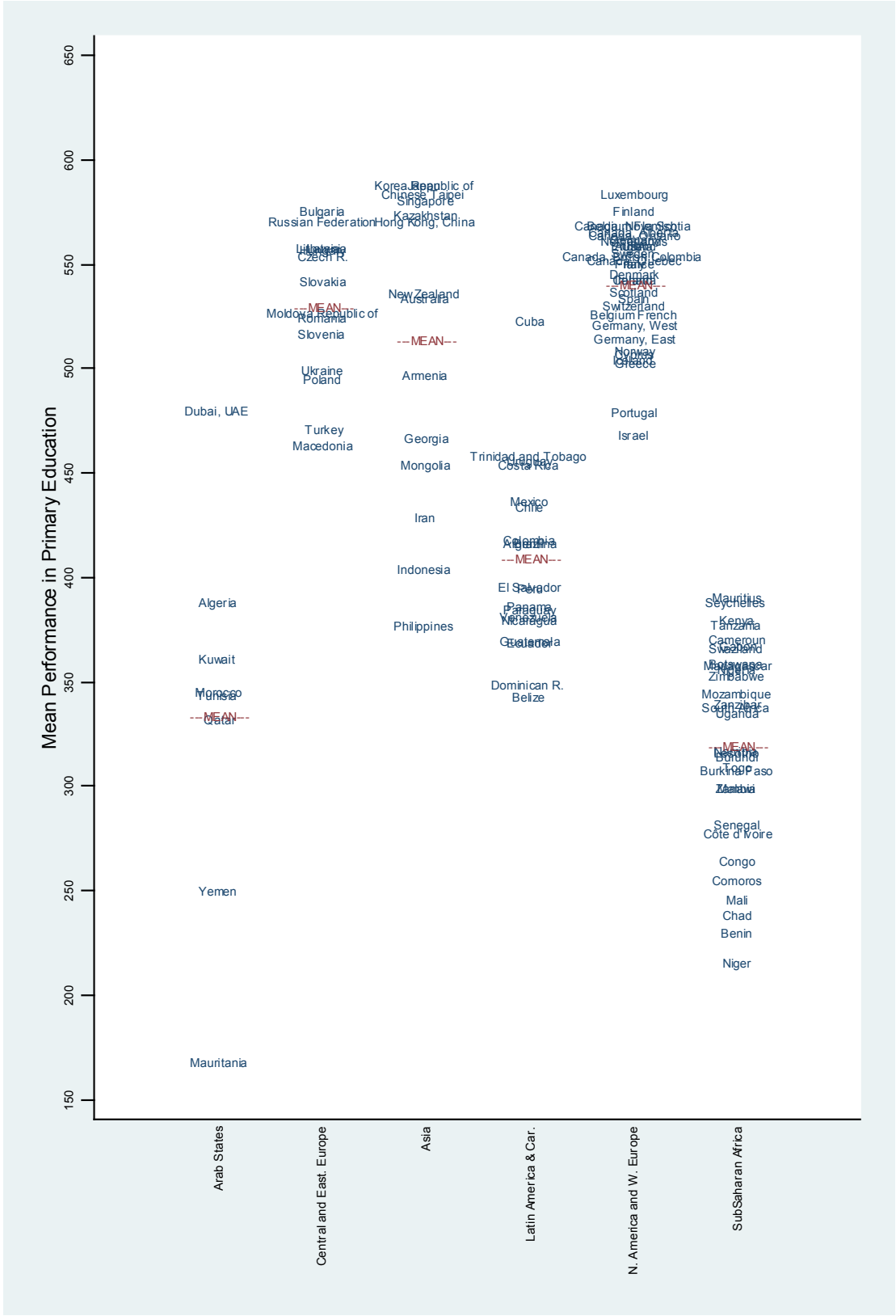

Figure 2. Average secondary school quality in the World (1964-2010)

Figure 3. Performance difference between Female and Male in Primary Education (Female – Male)

Figure 4. Performance difference between Female and Male in Secondary Education (Female – Male)

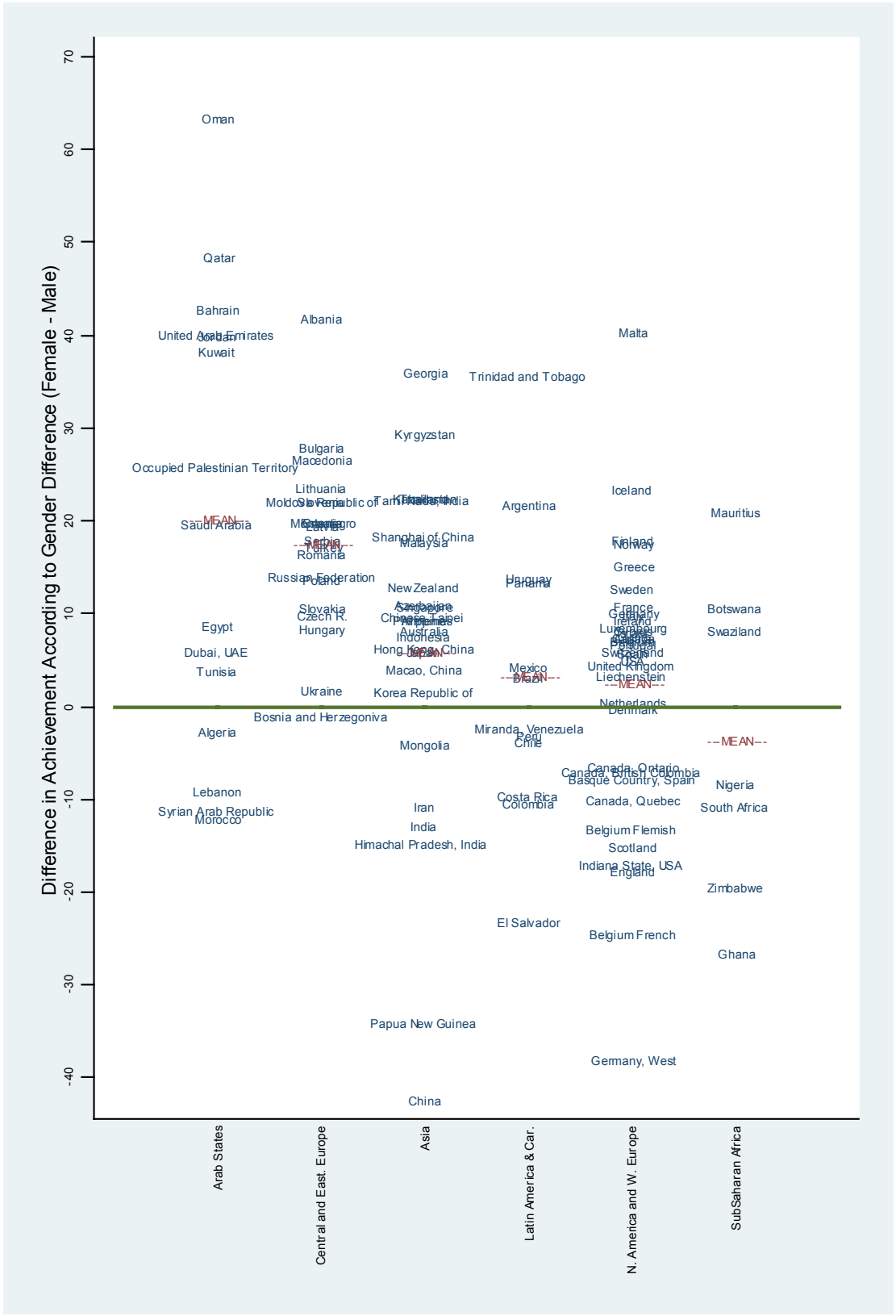

Figure 5. Performance difference between Urban and Rural Areas in Primary Education (Urban – Rural)

Figure 6. Performance difference between Urban and Rural Areas in Secondary Education (Urban – Rural)

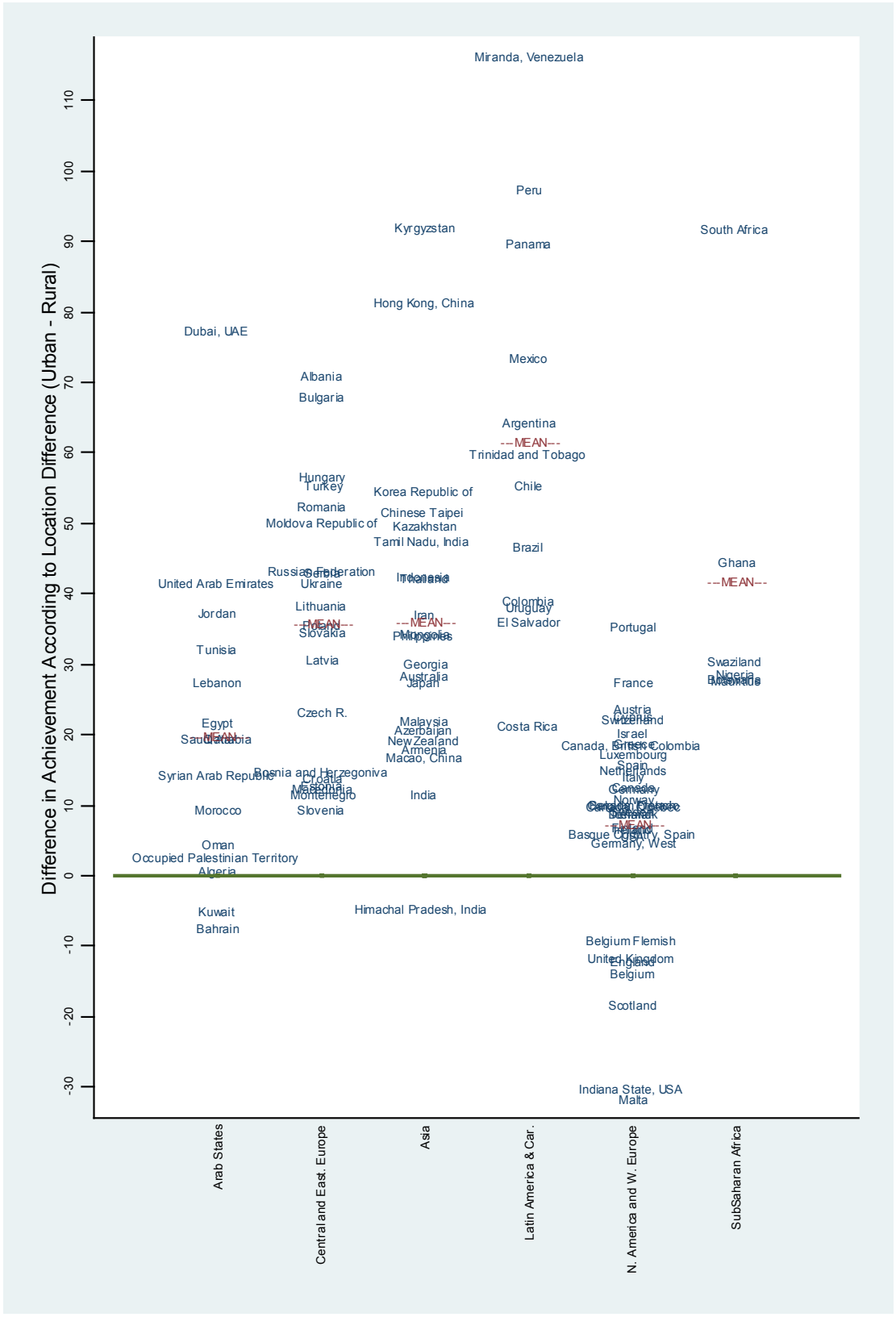

Figure 7. Long-term trends for selected countries, 1965-2010

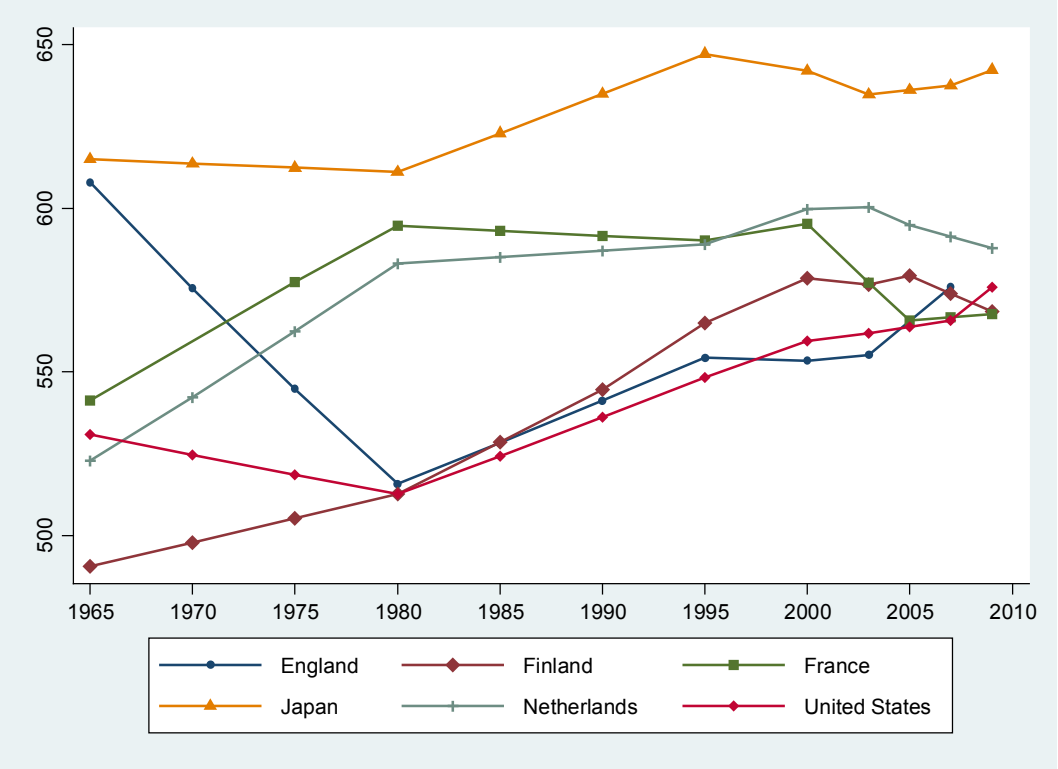

Figure 8. Correlation between mean score and average annual growth rate of IQSA-primary level (32 observations, $R^2 = 0{,}01$)

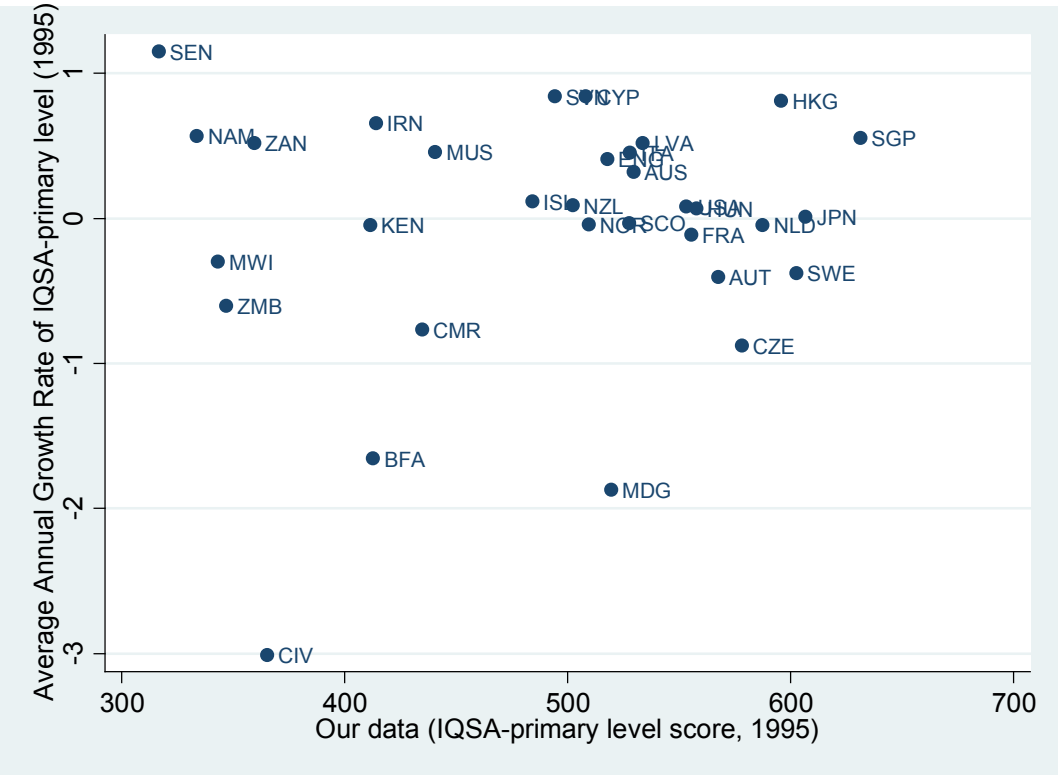

Figure 9. Correlation between mean score and average annual growth rate of IQSA-secondary level (43 observations, R^2= 0,31)

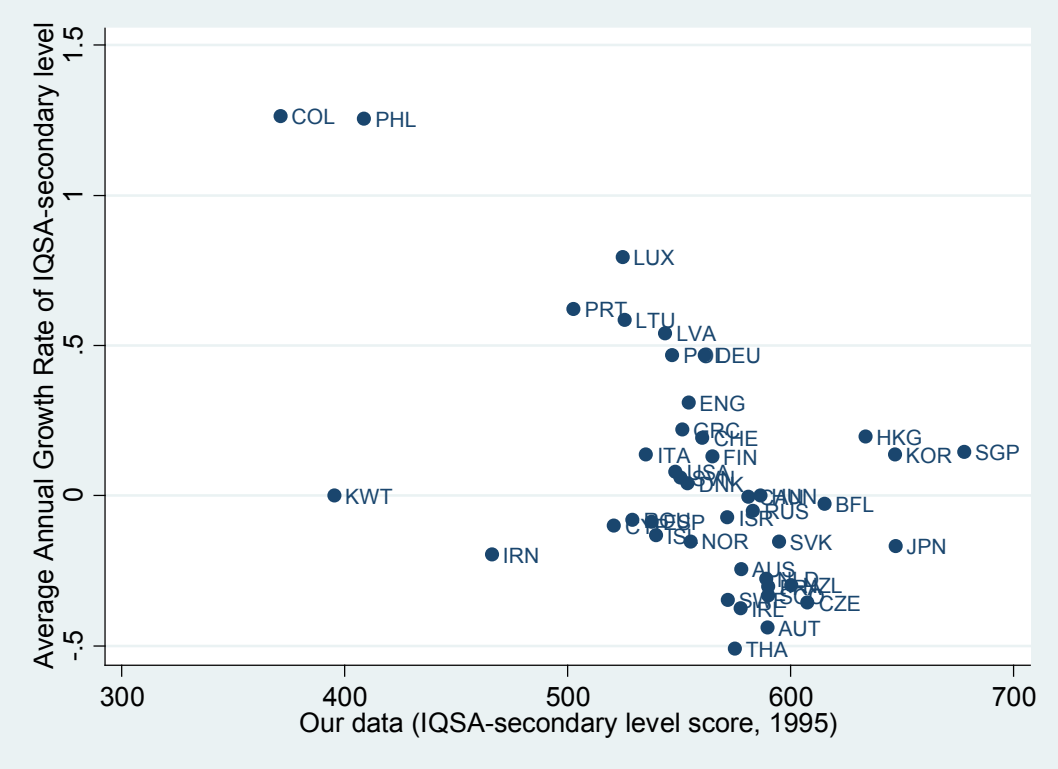

Figure 10. Correlation between IQSA and Hanushek and Woessmann (2012) dataset, (75 observations, R^2= 0,85)

Figure 11. Correlation between IQSA and Net Enrolment Rate (Primary education), high-income countries, (32 observations, R^2= 0,01)

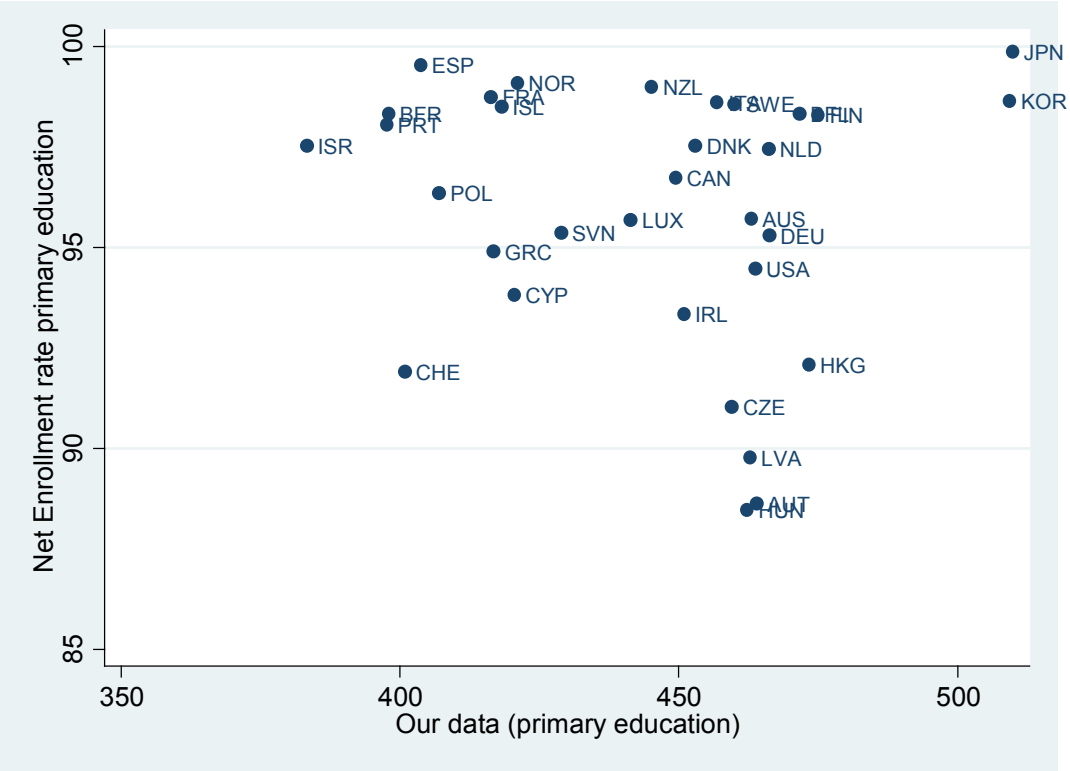

Figure 12. Correlation between IQSA and Net Enrolment Rate (Secondary education), high-income countries (38 observations, R^2= 0,00)

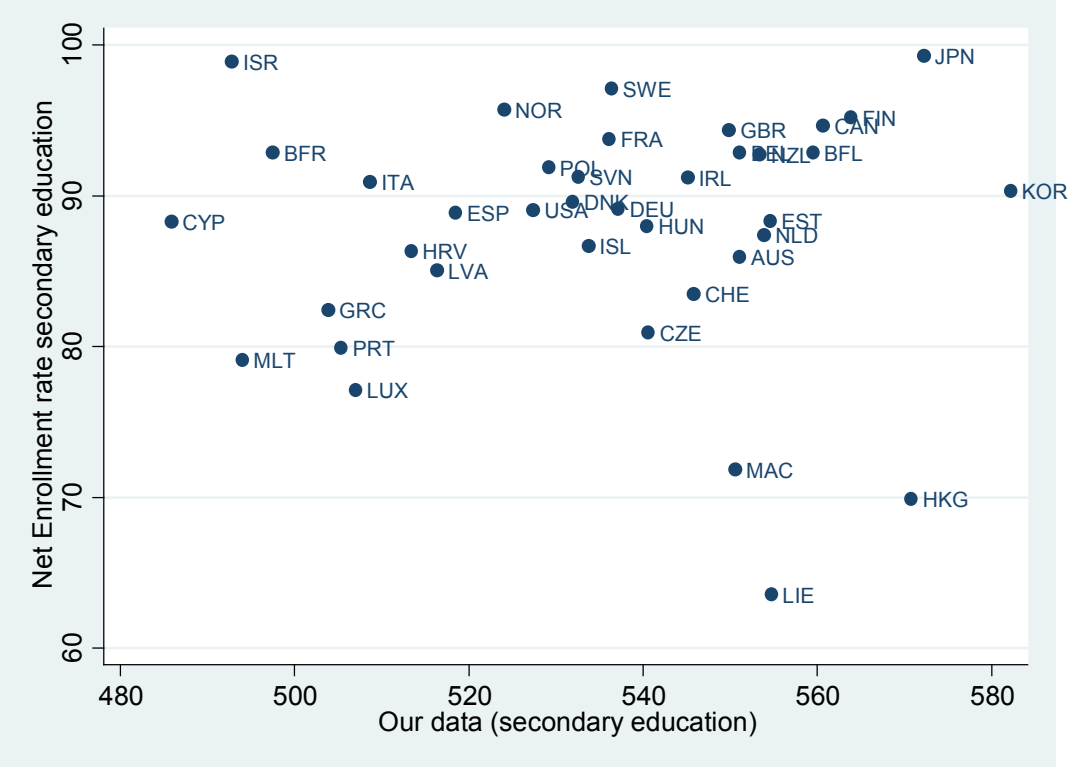

Figure 13. Correlation between IQSA and Barro and Lee (2010) Dataset, high-income countries (38 observations, R^2= 0,00)

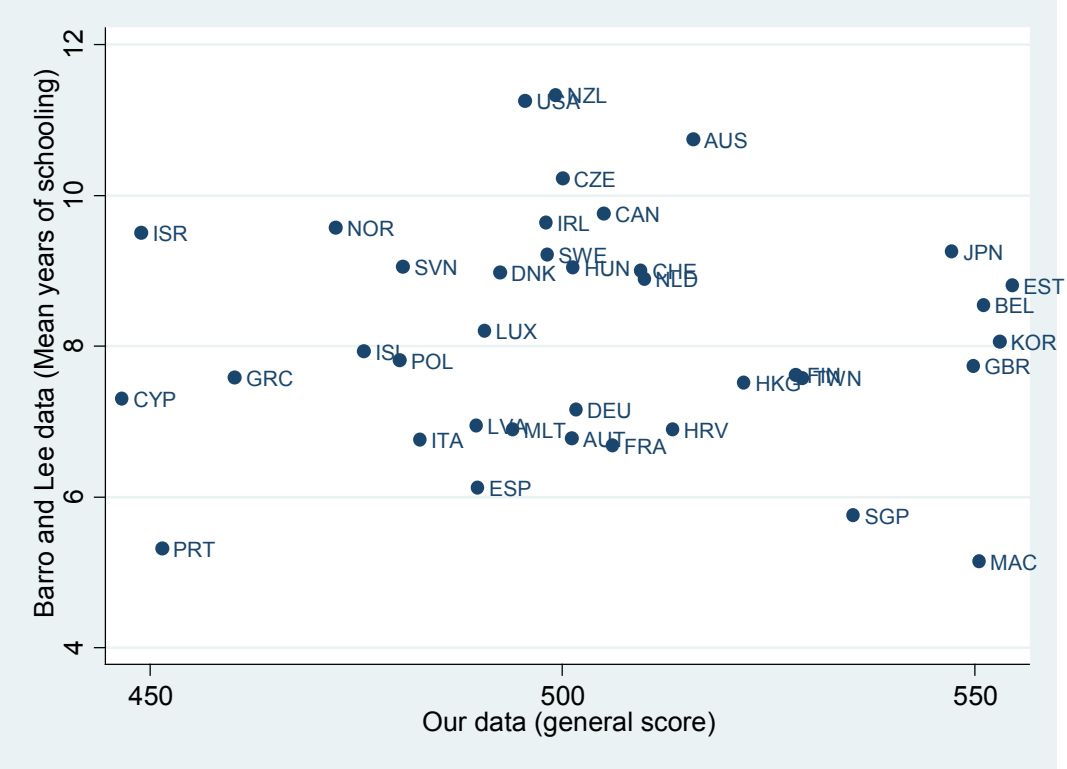

Table 1. Review of Student Achievement Tests

No	*Year*	*Organization*	*Abbr.*	*Subject*	*Countries/ Areas*	*Grade/Age*	*Included*	*Survey Series*
1	1959-1960	IEA	Pilot Study	M,S,R	12	7,8	No	-
2	1964	IEA	FIMS	M	12	7, FS	Yes	A.1.
3	1970-71	IEA	SRC	R	15	4,8, FS.	No	-
4	1970-72	IEA	FISS	S	19	4,8, FS.	Yes	A.1
5	1980-82	IEA	SIMS	M	19	8, FS	Yes	A.2
6	1983-1984	IEA	SISS	S	23	4,8, FS	Yes	A.2
7	1988, 1990-91	NCES	IAEP	M,S	6, 19	4,7-8	No	-
8	1990-1991	IEA	RLS	R	32	3-4,7-8	Yes	A.1
9	1995,1999,2003,2007,2011	IEA	TIMSS	M,S		3-4,7-8, FS	Yes	A.1 (1995), A.2. (Other years - except 2011)
10	1992-97	UNESCO	MLA	M,S,R	72	6,8	No	-
11	1997, 2006	UNESCO	LLECE	M,S,R	13	3,6	Only	B.
12	1999, 2002,2007	UNESCO	SACMEQ	M,R	7, 15,16	6	Yes	B
13	1993-2001;2002-2012	CONFEMEN	PASEC	M ;R		2,5	Yes	B
14	2001,2006,2011	IEA	PIRLS	R	35, 41, 55	4	Yes	A.1 (2001) ; A.2. (Other years - except 2011)
15	2011	IEA	prePIRLS	R	6	2,3	No	-
16	2000, 2003,2006,2009	OECD	PISA	M,S,R	43, 41, 57, 75	Age 15	Yes	A.1 (2000 for reading ; 2003 for maths ; 2006 for science); A.2. (Other years for reading ; until 2003 for maths ; until 2006 for science)

Note: For the meaning of abbreviations, please consult page 21. « FS » means "Final Secondary". Only assessments for which there is an information in "Survey Series" column are included in our dataset.

Table 2. Cross-section dataset

Country	Primary education			Secondary education			Primary + Secondary education		
	Mean score	Minimum Level	Advanced Level	Mean score	Minimum Level	Advanced Level	Mean score	Minimum Level	Advanced Level
Albania				456	0.68	0.11	438	0.62	0.09
Algeria	388	0.45	0.01	438	0.72	0.00	413	0.59	0.01
Argentina	417	0.57	0.02	476	0.72	0.18	446	0.64	0.10
Armenia	497	0.83	0.15	528	0.91	0.22	513	0.87	0.18
Australia	534	0.92	0.23	619	0.97	0.55	577	0.95	0.39
Austria	559	0.98	0.30	603	0.96	0.52	581	0.97	0.41
Azerbaijan				477	0.82	0.07	458	0.77	0.05
Bahrain				469	0.77	0.07	451	0.72	0.05
Basque C., Spain				543	0.97	0.22	520	0.95	0.18
Belgium				618	0.94	0.58	590	0.92	0.46
Belgium Flemish	569	1.00	0.30	582	0.97	0.45	576	0.98	0.38
Belgium French	526	0.93	0.18	517	0.89	0.18	522	0.91	0.18
Belize	343	0.29	0.01				374	0.43	0.06
Benin	230	0.12	0.01				269	0.30	0.07
Bosnia & Herz.				507	0.89	0.13	487	0.86	0.10
Botswana	358	0.24	0.00	399	0.50	0.01	379	0.37	0.01
Brazil	416	0.53	0.02	465	0.69	0.13	441	0.61	0.08
Bulgaria	575	0.96	0.42	529	0.85	0.29	552	0.91	0.35
Burkina Faso	307	0.25	0.03				341	0.40	0.08
Burundi	314	0.24	0.04				348	0.40	0.10
Cameroun	371	0.39	0.05				401	0.51	0.11
Canada	543	0.95	0.26	630	0.98	0.59	586	0.96	0.42
Canada (Alberta)	566	0.99	0.33				583	0.98	0.41
Canada (British Col.)	553	0.97	0.29	560	0.94	0.35	557	0.95	0.32
Canada (Nova Sco.)	569	0.98	0.37				586	0.97	0.45
Canada (Ontario)	564	0.98	0.33	572	0.98	0.37	568	0.98	0.35
Canada (Quebec)	552	0.99	0.24	582	0.99	0.40	567	0.99	0.32
Chad	238	0.13	0.01				276	0.31	0.07
Chile	434	0.64	0.03	501	0.80	0.20	468	0.72	0.12
China				540	0.93	0.29	517	0.90	0.23
Chinese Taipei	584	0.99	0.43	639	0.98	0.66	612	0.99	0.54
Colombia	418	0.58	0.02	462	0.70	0.12	440	0.64	0.07
Comoros	255	0.15	0.02				292	0.33	0.08
Congo	264	0.17	0.03				301	0.34	0.08
Costa Rica	454	0.73	0.04	518	0.89	0.20	486	0.81	0.12
Côte d'Ivoire	277	0.16	0.01				313	0.34	0.06
Croatia				577	0.95	0.41	552	0.93	0.33
Cuba	523	0.85	0.24				543	0.87	0.31
Cyprus	507	0.88	0.15	506	0.87	0.15	506	0.87	0.15
Czech Republic	554	0.97	0.28	605	0.97	0.51	580	0.97	0.40
Denmark	545	0.93	0.28	596	0.96	0.49	571	0.94	0.38

Country	Primary education			Secondary education			Primary + Secondary education		
	Mean score	Minimum Level	Advanced Level	Mean score	Minimum Level	Advanced Level	Mean score	Minimum Level	Advanced Level
Dominican Rep.	349	0.19	0.00				380	0.36	0.05
Dubai, UAE	480	0.78	0.11	526	0.87	0.24	503	0.83	0.18
Ecuador	369	0.32	0.00				399	0.46	0.05
Egypt				448	0.66	0.07	431	0.59	0.06
El Salvador	396	0.46	0.01	400	0.49	0.01	398	0.48	0.01
England	559	0.94	0.35	570	0.95	0.39	564	0.95	0.37
Estonia				621	0.99	0.58	594	0.97	0.46
Finland	575	0.97	0.44	635	0.98	0.61	605	0.97	0.52
France	551	0.98	0.26	601	0.95	0.51	576	0.96	0.39
Gabon	367	0.39	0.02				398	0.51	0.07
Georgia	467	0.77	0.06	462	0.71	0.10	465	0.74	0.08
Germany	562	0.98	0.31	602	0.94	0.52	582	0.96	0.42
Germany, East	514	0.85	0.23				535	0.87	0.30
Germany, West	521	0.87	0.25	572	0.98	0.37	547	0.92	0.31
Ghana				339	0.29	0.01	328	0.17	0.01
Greece	503	0.87	0.13	566	0.92	0.39	534	0.90	0.26
Guatemala	370	0.30	0.00				400	0.44	0.06
Hima. Prad., India				395	0.45	0.02	381	0.35	0.01
Hong Kong (China)	571	0.96	0.40	640	0.98	0.66	605	0.97	0.53
Hungary	557	0.96	0.33	605	0.97	0.52	581	0.96	0.42
Iceland	504	0.89	0.12	599	0.96	0.49	552	0.92	0.30
India				435	0.63	0.04	419	0.56	0.03
Indiana State, USA				572	0.99	0.35	548	0.97	0.28
Indonesia	404	0.48	0.01	469	0.77	0.09	436	0.62	0.05
Iran Islam. Rep.	429	0.62	0.03	472	0.80	0.07	451	0.71	0.05
Ireland	543	0.93	0.28	612	0.97	0.54	578	0.95	0.41
Israel	468	0.69	0.14	552	0.87	0.36	510	0.78	0.25
Italy	551	0.95	0.29	571	0.93	0.39	561	0.94	0.34
Japan	588	0.99	0.46	640	0.98	0.66	614	0.98	0.56
Jordan				499	0.81	0.19	479	0.76	0.15
Kazakhstan	574	0.98	0.39	482	0.78	0.14	528	0.88	0.26
Kenya	380	0.33	0.01				409	0.47	0.06
Korea Rep.	588	0.97	0.50	653	0.99	0.69	620	0.98	0.60
Kuwait	361	0.38	0.01	424	0.61	0.03	393	0.49	0.02
Kyrgyzstan				380	0.39	0.03	367	0.28	0.02
Latvia	558	0.98	0.30	579	0.96	0.41	569	0.97	0.36
Lebanon				465	0.76	0.07	447	0.71	0.05
Lesotho	317	0.10	0.00				350	0.29	0.05
Liechenstein				621	0.97	0.58	594	0.95	0.46
Lithuania	558	0.99	0.28	574	0.96	0.39	566	0.97	0.34
Luxembourg	584	1.00	0.42	569	0.92	0.40	576	0.96	0.41
Macao, China				616	0.99	0.58	589	0.97	0.46

Country	Primary education			Secondary education			Primary + Secondary education		
	Mean score	Minimum Level	Advanced Level	Mean score	Minimum Level	Advanced Level	Mean score	Minimum Level	Advanced Level
Macedonia F.Y.R.	464	0.71	0.09	484	0.78	0.14	474	0.74	0.12
Madagascar	358	0.38	0.09				389	0.51	0.15
Malawi	303	0.04	0.00				337	0.24	0.05
Malaysia				536	0.92	0.26	513	0.89	0.20
Mali	246	0.13	0.02				284	0.31	0.07
Malta				553	0.86	0.38	529	0.83	0.30
Mauritania	168	0.06	0.00				211	0.25	0.06
Mauritius	391	0.40	0.07	501	0.81	0.20	446	0.60	0.13
Mexico	437	0.64	0.04	495	0.81	0.16	466	0.73	0.10
Miranda (Venez.)				501	0.79	0.21	480	0.74	0.16
Moldova Rep.	527	0.92	0.19	503	0.85	0.17	515	0.89	0.18
Mongolia	454	0.73	0.04	485	0.85	0.07	470	0.79	0.05
Montenegro				487	0.79	0.14	467	0.74	0.11
Morocco	345	0.31	0.01	409	0.54	0.01	377	0.43	0.01
Mozambique	345	0.22	0.00				376	0.38	0.05
Namibia	317	0.11	0.00				350	0.29	0.05
Netherlands	561	0.98	0.29	621	0.98	0.57	591	0.98	0.43
New Zealand	537	0.91	0.27	622	0.97	0.55	579	0.94	0.41
Nicaragua	380	0.36	0.00				409	0.49	0.05
Niger	216	0.09	0.01				256	0.28	0.06
Nigeria	356	0.27	0.00	441	0.70	0.03	399	0.48	0.02
Norway	508	0.89	0.14	589	0.96	0.44	549	0.92	0.29
Occ. Palest. Ter.				440	0.64	0.06	423	0.57	0.05
Oman				437	0.65	0.05	421	0.59	0.04
Panama	386	0.40	0.01	446	0.64	0.09	416	0.52	0.05
Papua New Guinea				503	0.94	0.09	482	0.91	0.07
Paraguay	385	0.40	0.01				414	0.52	0.06
Peru	395	0.46	0.01	416	0.53	0.07	405	0.50	0.04
Philippines	377	0.40	0.03	402	0.49	0.03	389	0.44	0.03
Poland	495	0.83	0.13	594	0.95	0.48	545	0.89	0.30
Portugal	479	0.81	0.08	568	0.94	0.38	524	0.88	0.23
Qatar	332	0.27	0.00	409	0.47	0.09	371	0.37	0.04
Romania	525	0.89	0.21	522	0.86	0.24	523	0.88	0.22
Russian Fed.	570	0.98	0.37	573	0.95	0.40	571	0.96	0.38
Saudi Arabia				402	0.50	0.01	388	0.41	0.00
Scotland	537	0.94	0.23	557	0.94	0.34	547	0.94	0.28
Senegal	282	0.21	0.02				318	0.37	0.08
Serbia				529	0.89	0.25	508	0.86	0.20
Seychelles	388	0.36	0.01				417	0.49	0.06
Shanghai (China)				696	1.00	0.82	665	0.98	0.65
Singapore	580	0.93	0.48	658	0.98	0.71	619	0.95	0.59
Slovakia	542	0.95	0.25	590	0.96	0.46	566	0.95	0.35

Country	Primary education			Secondary education			Primary + Secondary education		
	Mean score	Minimum Level	Advanced Level	Mean score	Minimum Level	Advanced Level	Mean score	Minimum Level	Advanced Level
Slovenia	517	0.92	0.14	597	0.98	0.47	557	0.95	0.31
South Africa	338	0.21	0.01	282	0.16	0.02	310	0.18	0.02
Spain	534	0.93	0.23	582	0.95	0.43	558	0.94	0.33
Swaziland	367	0.26	0.00	462	0.80	0.05	414	0.53	0.02
Sweden	555	0.97	0.29	603	0.96	0.49	579	0.97	0.39
Switzerland	530	0.89	0.28	611	0.96	0.54	571	0.92	0.41
Syrian Arab Republic				445	0.68	0.04	428	0.62	0.03
Tamil Nadu, India				417	0.54	0.03	402	0.46	0.02
Tanzania	377	0.29	0.00				407	0.44	0.06
Thailand				525	0.91	0.22	504	0.88	0.17
Togo	310	0.26	0.02				344	0.41	0.08
Trinidad & Tobago	459	0.68	0.09	500	0.76	0.23	479	0.72	0.16
Tunisia	344	0.34	0.01	472	0.77	0.10	408	0.55	0.05
Turkey	471	0.79	0.06	527	0.87	0.26	499	0.83	0.16
Uganda	335	0.18	0.00				367	0.35	0.05
Ukraine	500	0.87	0.11	521	0.90	0.20	511	0.89	0.15
United Arab Emirates				520	0.85	0.24	499	0.81	0.19
United Kingdom				617	0.97	0.56	590	0.94	0.44
Uruguay	456	0.75	0.05	515	0.82	0.24	485	0.79	0.15
USA	559	0.96	0.34	593	0.95	0.46	576	0.95	0.40
Venezuela	381	0.38	0.00				411	0.50	0.06
Yemen	250	0.14	0.00				288	0.31	0.06
Zambia	304	0.05	0.00				338	0.25	0.05
Zanzibar	338	0.14	0.00				370	0.32	0.05
Zimbabwe	353	0.25	0.01	444	0.72	0.03	398	0.48	0.02

Note: First column in each group represents the mean score. Mean scores for both primary and secondary education are calculated by aggregating both levels. We predicted scores for levels for which no data is available in order to obtain more consistent data for schooling performance. Second column in each group represent the share of pupils in each country reaching the "Minimum Level" benchmark, defined as one standard deviation below the international benchmark (i.e. 400 points in our dataset). The third column ("Advanced Level") provides the proportion of pupils reaching at least one standard deviation above the international benchmark (i.e. 600 points). See Appendix B for more details.

Table 3. Summary statistics on panel dataset

Year	Countries/areas	Mean	Standard Deviation	Minimum	Maximum
Primary level					
1995	40	481.94	99.11	259.42	631.35
2000	45	471.69	98.60	305.97	634.19
2003	64	471.46	97.87	275.15	635.89
2005	76	451.80	98.96	256.35	638.21
2007	76	455.27	108.52	212.64	648.97
Secondary level					
1965	21	540.06	50.95	453.65	623.84
1970	11	546.73	37.35	497.80	613.69
1975	13	549.96	36.76	505.20	623.76
1980	32	526.27	40.92	458.67	640.65
1985	24	544.77	35.58	442.03	622.93
1990	24	558.30	40.48	425.04	634.95
1995	44	557.08	59.78	371.13	677.87
2000	61	538.48	74.60	302.67	673.49
2003	75	530.68	76.41	293.62	674.37
2005	69	532.81	70.15	324.13	670.49
2007	87	522.56	73.32	341.42	666.63
2009	71	530.60	65.65	376.44	673.83

Table 4. Long-term trends on schooling quality for 24 economies, 1980-2007

Country	Score 1980	Score 2007	Variation (points)	Variation (%)	Annual growth rate
Singapore	513	667	154	30	0.98
Korea Republic of	522	666	143	27	0.90
Hong Kong, China	551	640	89	16	0.55
Luxembourg	497	577	80	16	0.55
Finland	513	574	61	12	0.42
England	516	576	60	12	0.41
USA	513	566	53	10	0.37
Canada	522	574	52	10	0.35
Poland	523	572	49	9	0.33
Scotland	502	548	46	9	0.32
Canada, British Colombia	529	567	39	7	0.26
Canada, Ontario	546	576	30	5	0.20
Japan	611	638	26	4	0.16
New Zealand	523	547	24	5	0.17
Australia	528	552	24	5	0.17
Netherlands	583	591	8	1	0.05
Norway	519	523	4	1	0.03
Italy	535	534	-1	0	-0.01
Israel	517	510	-7	-1	-0.05
Sweden	568	548	-20	-3	-0.13
France	595	567	-28	-5	-0.18
Thailand	528	492	-37	-7	-0.26
Hungary	641	577	-64	-10	-0.39
Ghana	480	345	-136	-28	-1.22

Table 5. Analysis of Short term trends (1995-2010)

	Primary					Secondary					Primary+Secondary				
	Score	Average Annual Growth Rate				Score	Average Annual Growth Rate				Scor	Average Annual Growth Rate			
	Mean	Mean	SE	NB	Intensity	Mean	Mean	SE	NB	Intensity	Mean	Mean	SE	NB	Intensity
Qatar	332			0		409	4.60	0.94	3	++	371	4.60	0.94	3	++
Syrian AR				0		445	2.52	0.00	2	++	445	2.52	0.00	2	++
Kyrgyzstan				0		380	2.08	0.74	3	++	380	2.08	0.74	3	++
Armenia	497	2.47	0.16	2	++	528	1.05	0.06	2	++	513	1.76	0.42	4	++
Peru	395			0		416	1.56	0.25	4	++	405	1.56	0.25	4	++
Colombia	418			0		462	1.26	0.18	7	++	440	1.26	0.18	7	++
Senegal	282	1.15	0.16	2	++				0		282	1.15	0.16	2	++
Lebanon				0		465	1.09	0.18	2	++	465	1.09	0.18	2	++
Brazil	416			0		465	0.91	0.14	19	++	446	0.91	0.14	19	++
Tanzania	377	0.81	0.00	2	++				0		377	0.81	0.00	2	++
Luxembourg	584			0		569	0.79	0.21	22	++	572	0.79	0.21	22	++
Turkey	471			0		527	0.79	0.21	12	++	513	0.79	0.21	12	++
Lesotho	317	0.72	0.20	2	++				0		317	0.72	0.20	2	++
Chile	434			0		501	0.64	0.17	12	++	474	0.64	0.17	12	++
Portugal	479			0		568	0.62	0.11	20	++	524	0.62	0.11	20	++
Swaziland	367	0.59	0.07	2	++	462			0		398	0.59	0.07	2	++
Latvia	558	0.52	0.18	6	++	579	0.54	0.12	26	++	569	0.54	0.11	32	++
Basque C, S.				0		543	0.49	0.05	2	+	543	0.49	0.05	2	+
Lithuania	558	-0.11	0.11	3	o	574	0.59	0.14	17	++	566	0.48	0.13	20	+
Canada, BC	553	0.69		1		560	0.26		1		556	0.47	0.21	2	+
Poland	495			0		594	0.47	0.07	22	+	555	0.47	0.07	22	+
Germany	562	0.29		1		602	0.47	0.08	20	+	582	0.46	0.08	21	+
Slovenia	517	0.84	0.12	7	++	597	0.06	0.13	11	o	557	0.37	0.13	18	+
England	559	0.41	0.25	6	o	570	0.31	0.05	16	+	563	0.34	0.08	22	+
Hong Kong	571	0.81	0.15	7	++	640	0.20	0.11	35	o	605	0.30	0.10	42	+
Singapore	580	0.55	0.17	7	++	658	0.15	0.13	15	o	619	0.28	0.11	22	+
Ghana				0		339	2.20	1.77	3	o	339	2.20	1.77	3	o
Philippines	377			0		402	1.26	1.02	3	o	389	1.26	1.02	3	o
Morocco	345	-0.91	0.38	3	--	409	2.18	0.91	6	++	371	1.15	0.79	9	o
Serbia				0		529	1.00	0.58	5	o	529	1.00	0.58	5	o
Bahrain				0		469	0.73	0.90	2	o	469	0.73	0.90	2	o
Albania				0		456	0.59	0.33	4	o	456	0.59	0.33	4	o
Namibia	317	0.57	0.81	3	o				0		317	0.57	0.81	3	o
Zanzibar	338	0.52	0.62	3	o				0		338	0.52	0.62	3	o
Liechenstein				0		621	0.48	0.25	19	o	621	0.48	0.25	19	o
Mauritius	391	0.46	0.48	3	o	501			0		457	0.46	0.48	3	o
Tunisia	344	-0.38	0.62	2	o	472	0.45	0.30	16	o	421	0.36	0.28	18	o
Mexico	437			0		495	0.32	0.23	19	o	471	0.32	0.23	19	o
Jordan				0		499	0.26	0.17	10	o	499	0.26	0.17	10	o
Montenegro				0		487	0.24	0.63	3	o	487	0.24	0.63	3	o
Greece	503			0		566	0.22	0.14	20	o	534	0.22	0.14	20	o
Switzerland	530			0		611	0.19	0.08	20	o	591	0.19	0.08	20	o
Italy	551	0.45	0.17	4	+	571	0.14	0.10	28	o	561	0.18	0.09	32	o
Indonesia	404	0.63		1		469	0.12	0.19	25	o	453	0.14	0.18	26	o
Korea	588			0		653	0.14	0.10	34	o	627	0.14	0.10	34	o
Finland	575			0		635	0.13	0.10	23	o	611	0.13	0.10	23	o
Saudi Arabia				0		402	0.09	0.25	2	o	402	0.09	0.25	2	o
Denmark	545	0.92		1		596	0.04	0.15	20	o	571	0.08	0.15	21	o
USA	559	0.08	0.09	7	o	593	0.08	0.09	32	o	576	0.08	0.08	39	o
Iran	429	0.66	0.20	6	++	472	-0.20	0.11	13	o	446	0.07	0.13	19	o
Argentina	417			0		476	0.04	0.33	10	o	446	0.04	0.33	10	o
Russia	570	0.96	0.23	3	++	573	-0.05	0.12	32	o	571	0.04	0.12	35	o
Cyprus	507	0.84	0.02	2	++	506	-0.10	0.14	13	o	506	0.03	0.15	15	o

	Primary					Secondary					Primary+Secondary				
	Score	Average Annual Growth Rate				Score	Average Annual Growth Rate				Scor	Average Annual Growth Rate			
	Mean	Mean	SE	NB	Intensity	Mean	Mean	SE	NB	Intensity	Mean	Mean	SE	NB	Intensity
Hungary	557	0.07	0.18	7	o	605	0.00	0.06	35	o	581	0.01	0.06	42	o
Chinese Taipei	584	0.38	0.14	2	+	639	-0.07	0.12	10	o	612	0.01	0.11	12	o
Canada	543			0		630	0.00	0.12	24	o	586	0.00	0.12	24	o
Belgium				0		618	-0.03	0.08	19	o	618	-0.03	0.08	19	o
Belgium, Fl.	569			0		582	-0.03	0.16	8	o	574	-0.03	0.16	8	o
Kenya	380	-0.04	0.08	3	o				0		380	-0.04	0.08	3	o
Trinidad &T.	459	-0.05		1	o	500			0		490	-0.05		1	o
Israel	468	0.14		1	o	552	-0.07	0.23	23	o	519	-0.06	0.22	24	o
Belgium Fr.	526	-0.06		1	o	517			0		520	-0.06		1	o
Botswana	358	0.29	0.08	2	+	399	-0.44	0.25	2	o	379	-0.08	0.24	4	o
Spain	534	0.07		1	o	582	-0.09	0.14	20	o	570	-0.08	0.13	21	o
Uruguay	456			0		515	-0.09	0.24	9	o	485	-0.09	0.24	9	o
Canada, Ont.	564			0		572	-0.10	0.16	3	o	567	-0.10	0.16	3	o
Iceland	504	0.12	0.20	2	o	599	-0.13	0.12	20	o	552	-0.11	0.11	22	o
Romania	525	-0.88		1		522	-0.08	0.15	23	o	522	-0.11	0.14	24	o
Slovakia	542	0.49		1		590	-0.15	0.12	17	o	566	-0.12	0.12	18	o
Estonia				0		621	-0.12	0.07	3	o	621	-0.12	0.07	3	o
Norway	508	-0.04	0.28	7	o	589	-0.15	0.12	28	o	549	-0.13	0.11	35	o
Seychelles	388	-0.13	0.04	2	o				0		388	-0.13	0.04	2	o
Moldova	527	0.33		1		503	-0.30	0.54	3	o	515	-0.14	0.41	4	o
South Africa	338	0.16	0.09	2	o	282	-0.45	0.56	2	o	310	-0.14	0.29	4	o
Japan	588	0.01	0.07	5	o	640	-0.17	0.11	35	o	619	-0.14	0.09	40	o
Australia	534	0.32	0.15	5	+	619	-0.25	0.08	28	-	585	-0.16	0.08	33	o
Macao				0		616	-0.19	0.05	9	o	616	-0.19	0.05	9	o
Scotland	537	-0.03	0.08	6	o	557	-0.33	0.17	9	o	545	-0.21	0.11	15	o
New Zealand	537	0.09	0.20	7	o	622	-0.30	0.14	29	-	579	-0.22	0.12	36	o
Malawi	303	-0.30	0.61	3	o				0		303	-0.30	0.61	3	o
Macedonia	464	0.01		1	o	484	-0.57	0.09	2	--	479	-0.37	0.20	3	o
Uganda	335	-0.41	0.30	2	o				0		335	-0.41	0.30	2	o
Zambia	304	-0.60	0.51	3	o				0		304	-0.60	0.51	3	o
Malaysia				0		536	-0.84	0.43	6	o	536	-0.84	0.43	6	o
Kuwait	361	-3.59		1		424	0.00	0.03	3	o	386	-0.90	0.90	4	o
Azerbaijan				0		477	-1.08	1.18	3	o	477	-1.08	1.18	3	o
Netherlands	561	-0.05	0.15	7	o	621	-0.28	0.11	29	-	591	-0.23	0.09	36	-
France	551	-0.11	0.02	2	o	601	-0.30	0.09	22	-	589	-0.29	0.08	24	-
Sweden	555	-0.38	0.05	2	-	603	-0.35	0.05	28	-	579	-0.35	0.05	30	-
Croatia				0		577	-0.37	0.13	3	-	577	-0.37	0.13	3	-
Czech Republic	554	-0.88		1		605	-0.36	0.08	26	-	580	-0.37	0.08	27	-
Ireland	543			0		612	-0.38	0.11	20	-	578	-0.38	0.11	20	-
Austria	559	-0.40		1		603	-0.44	0.18	20	-	581	-0.44	0.17	21	-
Thailand				0		525	-0.51	0.12	29	--	525	-0.51	0.12	29	--
Bulgaria	575	-0.13		1	o	529	-0.55	0.28	16	o	541	-0.52	0.26	17	--
United Kingdom				0		617	-0.63	0.14	18	--	617	-0.63	0.14	18	--
Cameroun	371	-0.77	0.16	2	--				0		371	-0.77	0.16	2	--
Canada, Quebec	552			0		582	-0.87	0.17	2	--	564	-0.87	0.17	2	--
Egypt				0		448	-0.87	0.10	2	--	448	-0.87	0.10	2	--
Mozambique	345	-1.23	0.07	2	--				0		345	-1.23	0.07	2	--
Burkina Faso	307	-1.66	0.22	2	--				0		307	-1.66	0.22	2	--
Palestine				0		440	-1.69	0.16	2	--	440	-1.69	0.16	2	--
Madagascar	358	-1.87	0.33	2	--				0		358	-1.87	0.33	2	--
Côte d'Ivoire	277	-3.01	0.47	2	--				0		277	-3.01	0.47	2	--
Yemen	250	-5.40	0.11	2	--				0		250	-5.40	0.11	2	--