ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

# How the Large Scale Assessment Tests Help in Achieving Sustainable Development Goals: A Critical Perspective

**Jizhou Liu[1,a], Songjiang Duan[2,b], Lin Liang[2,c*]**

*[1] People's Government of Jilin Province, Changchun, 130000, China*
*[2] Changchun University, Changchun, 130022, China*
*[a]13844947888@163.com, [b]songjiang.duan@foxmail.com, [c] lyn_liang52@qq.com*
**Corresponding author*

***Abstract:*** *The 2015-2030 sustainable development goal is the first developmental goal to include targets for educational equity and learning outcomes. Current globalized testing culture suggests the need to develop standardized literacy and numeracy measures for monitoring learning at the international stage. The underlying study has critically assessed the large scale assessment tests and has highlighted limitations associated with each test (including PISA 2012, PIRLS 2011, TIMSS 2011, SACMEQ 2007, TERCE 2013, PASEC 2014). The study has extracted the data from regional and international assessment tests that are widely used to monitor the progress towards attainment of Sustainable Development Goals. For this purpose, the study has assessed the ability of students to reach minimum proficiency level in science, mathematics and reading. The research has attempted to prepare comparable proficiency scores for around fifty countries by extracting and analyzing the additional data and evaluating the major indicators in the education sector. Moreover, there are certain limitations associated with the study. The research has not included hybrid or national assessments, and for the ease of analysis, only most recent large scale assessment tests have been included. The findings have suggested a need to introduce some standardized assessments to encourage the collaboration among different organizations. The main purpose behind the execution of research was to provide support and practical guidance to international organizations and policymakers to design, develop and implement largely standardized scale assessment tests that can accurately measure the educational equity and learning outcomes across the world.*

***Keywords:*** *Sustainable Development Goals (SDGs), Large Scale Assessment Tests, UNESCO, Educational Equity, Learning Outcomes*

## 1. Introduction

The policymakers and analysts nowadays are increasingly relying on the large scale assessment tests to evaluate to inclusion and equity in the education sector. However, the results proposed by these tests often contradict each other. Hence, the main purpose of executing the critical analysis of these assessment tests is to offer practical guidance for appropriate and efficient design, development and execution of large scale assessment tests. This paper offers information about sustainable development goals and widely used latest assessment tests. It also highlights the limitations associated with each test and lack of coordination between different organizations administering those tests in the same region.

The sustainable development goals (SDGs) stress upon the significance of all young and adult students to at least get the basic education by the end of 2030. The overall aim of 2030 program is not even a single person would be left illiterate. However, the accomplishment of this broad objective is not an easy task as it involves a number of variables. The millennium development goals attainment has resulted in a clear increase in enrollment of students belonging to the middle or poor class. But, the sources have also evidenced the increasing inequality in the provision of learning opportunities. UNESCO (2014) has reported that approximately 250 million children worldwide are unable to get an even basic education. There is a need to conduct a critical assessment of issue to hand so that the real reasons behind this inequality could be highlighted. The global community needs to know which specific group of children is not getting the education and at what stage the learning gaps are most prevalent.

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

There is also need to understand whether the personal family issues are responsible for growing inequality or the overall system is fostering this gap. The underlying study assesses different equity indicators to assess the learning and cognitive abilities of children belonging to different backgrounds. The SDGs contain a total of 17 goals and 169 targets to achieve those goals. The SDG 4 (*Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*) is the educational goal which has ten targets for its successful accomplishment (UNESCO 2016). Four basic monitoring levels have been offered by the United Nations Secretary General's Synthesis Report, including national, thematic, regional and global (UNSG, 2014).

The underlying study has chosen the "global monitoring approach" that depends on a carefully chosen and limited number of indicators to offer an overview of the development. Among all indicators, the most important ones are educational equality and learning outcomes. A significant amount of effort is required to collect the statistical data regarding these two indicators across different countries. However, it is important to note that collection of required data requires a standardized assessment and process to make the comparable categories and create the standardized metrics for different countries.

### *1.1 Lack of Co-ordination and Unstandardized Results*

Unfortunately, today, the organizations lack collaboration, and this lack of coordination has made high quality data collection very difficult (UNESCO, 2016). Out of 10 targets, five specifically focus on the young children, youth and adults. The SDG agenda has mainly focused on the equity. Thus, the education indicators need to be more insightful and should capture the variation across different population sections characterized by individual and group characteristics like disability, language, ethnicity, location, wealth and sex. However, the measurement of learning abilities of students is a complex task as accurate measurement needs to resolve various political and technical challenges. Though various large scale assessments are available for this purpose, the lack of coordination among different organizations has resulted in unstandardized results. No framework is available to produce the comparable data across the nations.

Target 4.1 covers the quality of primary and lower secondary education. The current global indicator for this target is the "proportion of children and young people: (i) in Grade 2 or 3; (ii) at the end of primary education; and (iii) at the end of lower secondary education who achieved at least a minimum proficiency level in (a) reading and (b) mathematics". Large scale assessments comprise two sub categories, including household surveys and school based surveys. School based assessments are comprised of cross national initiatives and national assessments. The underlying study has chosen the cross national initiative and has tried to resolve the two complementary issues. The research has defined the minimum competency and proficiency levels. Additionally, on the basis of questionnaires, the study has aimed to collect the contextual information that mostly lacks for various counties. The paper has chosen the regional and international assessment tests to explore the underlying phenomenon.

### *1.2 Minimum Learning Level and Educational Inequalities*

The Minimum Learning Level (MLL) is the strategy for enhancing the overall quality of primary and elementary education. It involves the combination of equity with quality. MLL sets the learning outcomes in the form of different learning levels and competencies for each stage of primary education. Main reason behind laying down the minimum learning level arises from the primary concern that disregarding the sex, location, creed and caste, every child should have access to education and there should be a justifiable and comparable standard to assess their proficiency (Sharma & Dua 2013). The underlying study would discuss the percentage of young children in a pre-primary, primary and secondary level that achieve the MLL. Moreover, the research has chosen three study areas, including mathematics, general science and reading comprehension. The study monitors a specified proportion of the students studying in different grades in reading and mathematics class.

Analysis of developed countries has revealed that learning inequalities become apparent before children get into the school and this inequality increases during the schooling (Cunha et al., 2010). It has been reported that children from rich backgrounds are 15 months ahead of children with poor backgrounds in the United Kingdom (Blanden and Machin, 2010). Research has further indicated that socio-economic gap creates the difference in cognitive development by approximately 22 months and this gap widens as the child passes the primary level (Feinstein, 2003). Current research has reported that three-year-old children having a rich background are 3-times more likely to have highly advanced development than the children of the same age with the poor background (Jerrim and Vignoles, 2013).

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

Research conducted in the United States and the United Kingdom has reported varying results proposing that the learning gap increases throughout the schooling (e.g. Goodman et al., 2009 and Feinstein, 2003). Some researchers have also reported that the disparities remain same (Duncan and Magnuson, 2011; Reardon, 2011). However, no study has reported that the gaps narrow down throughout the schooling.

### *1.3 Sustainable Development Goals and Minimum Proficiency Level*

Sustainable development goals have provided a reference to the percentage of students that attain the minimum proficiency in reading and mathematics. Through, the mean score of students offer the comparison basis; it cannot highlight the students' weaknesses and strengths. In this regard, the item response theory allows to differentiate the items' difficulty level and offers a comprehensive description of students' characteristics according to the respective proficiency level. Similarly, PISA scales the students' results and divides the items into different levels to record the number of correctly answered items. There are total 6 levels from basic to highly advanced. Mean scores and proportion of students at each level offer the valuable insights to participating nations regarding their students. PIRLS and TIMSS offer four points along the achievement scale, namely, advanced, high, intermediate and low international benchmark. SACMEQ offers 8 proficiency levels for mathematics and reading.

Similarly, TERCE assessment involves 4 proficiency levels. PASEC involves 3 proficiency levels in mathematics and 4 levels in reading comprehension. The study has twenty combinations of skills, grades and assessment and has the same number of tables. PISA sets the minimum proficiency level at level 2. In reading, the students have to score above 407 whereas, the threshold in mathematics is 420. The study also provides new insights regarding environmental issues and respective proficiency levels, shown in the Table 1. In TIMSS, the minimum threshold level is 400 points in mathematics for grades 4 and 8. However, it is important to note that estimation bias may occur due to lack of alignment between different benchmarks.

*Table 1: Estimation of the differences of proficiency levels between assessments*

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | | | | | |
| **PIRLS** | 9.656 | 2.219 | 2.278 | 6.161 | -3.103 |
| | (1.807)*** | (2.183) | (2.340) | (1.547)*** | (1.485)** |
| **SACMEQ** | -0.942 | -4.661 | 2.117 | 29.039 | 24.407 |
| | (4.539) | (4.601) | (6.003) | (0.891)*** | (0.921)*** |
| **TERCE** | -14.291 | -18.010 | -12.483 | 14.786 | 10.154 |
| | (4.794)*** | (4.879)*** | (5.859)** | (2.108)*** | (2.271)*** |
| **PASEC** | -35.048 | -38.767 | -30.676 | | |
| | (6.132)*** | (6.213)*** | (7.775)*** | | |
| **PISA** | -4.856 | -8.574 | -8.345 | -12.332 | -16.964 |
| | (2.806)* | (2.939)*** | (2.899)*** | (1.212)*** | (1.326)*** |
| **GDP per capita** | No | No | Yes | No | No |
| **Dummies for skills** | No | Yes | Yes | Yes | Yes |
| **Country dummies** | No | No | No | Yes | Yes |
| **Observations** | 422 | 422 | 378 | 286 | 286 |
| **Countries** | 125 | 125 | 108 | 68 | 68 |
| **R squared** | 0.257 | 0.274 | 0.303 | 0.878 | 0.905 |

Note: Cluster-robust standard errors in brackets. Clusters are countries. Dummies for TIMSS survey and mathematics used as controls.

## 2. Educational Equity and Learning Outcomes

Researchers interested in international comparative studies have shown increased inclination to explore new learning opportunities and are conceptualizing and measuring different aspects of the association between learning and teaching, for instance, the time required to complete an assigned task and extent to which the content is taught and tested (Anderson, 1990). Today, various controversies

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

have surrounded the learning processes adopted by most of the schools. These controversies arise from the concept that different schooling process measures, mainly including learning and teaching can play an important role in developing more equitable and effective schools (Traiman, 1993). Both issues, educational equality and learning outcomes have also been highlighted in The 2030 Agenda of Sustainable Development Goal 4 (SDG 4-education 2030 agenda). While responding to equity issues in education, the indicators aim to capture national averages as well as variance within different sectors of population based on disability, language, ethnicity, location, wealth or sex (UNESCO, 2016). The underlying study has reviewed different assessment techniques and has attempted to explore if such variables relate to equity dimension. Among all study variables, the language, gender, location, socio-economic condition and immigrant status are the most important variables, representing the extent to which equity prevails in the overall system.

### *2.1 Socio-Economic Status*

The Socio-economic Status (SES) is one of the widely studied contextual variables by the education researchers. The researchers have mainly examined the academic achievement with regards to the socio-economic status of the students (Bornstein & Bradley, 2003). The first meta-analysis in this study area was conducted by White (1982) where the researchers examined the relationship between academic achievement and SES. The findings reported that various factors influence the relation between two variables. After the meta-analysis conducted by White (1982), new researchers have also assessed the influence of SES on academic achievement. However, the inconsistent results have been reported, ranging from the strong correlation (e.g. Sutton & Soderstrom, 1999) to absolutely no correlation (e.g., Ripple & Luthar, 2000). Assessment tests have unfortunately offered inadequate information about the socio-economic status of the families of students. The underlying study has chosen the home resources index as a proxy for SES. Though the chosen proxy does not cover each SES dimension; home resource index offers useful insights by differentiating between children having different socio-economic backgrounds. PASEC has not developed the clear wealth index as developed by the PISA assessment.

### *2.2 Gender*

Researchers have analyzed the gender difference in different test scores and have reported different findings. For example, research conducted by Cornwell, Mustard, & Van Parys (2013) has extended the analysis of gender difference at an early age in terms of academic achievement. The research included both teacher assessment and test scores. The results of the 1998-99 ECLSK cohort reported that the grades given by the instructors didn't align with the test scores. However, the disparities in grades exceeded in test outcomes, favoring the females. Such lack of alignment of test scores and grades could be linked to the gender difference in the non-cognitive development. The study also reported that the girls outperformed the boys on reading scores and the difference was statistically significant except for 5th graders. Test scores were same in mathematics and science whereas gender gap was most prevalent in whites. The study also reported that even those boys were treated unfavorably who scored same in science, mathematics and reading. But, the difference vanished when non-cognitive skills were considered.

### *2.3 Language and Immigrant Status*

This study variable has a significant influence on academic achievement when there are a large number of immigrants or where official language differs from the home language. Review of literature has however revealed that previous researchers have not adequately explored this variable. Previous researchers have further contended that children face a significant amount of difficulty when the language spoken at home differs from the language spoken at school (Steinberg, Blinde & Chan, 1984). This is one of the major reasons behind the gap in academic achievement of immigrant and native students (Fuligni, 1997). The gap is particularly wide in early years of education and narrows down depending upon the support from family and child's own cognitive abilities (Andon, Thompson & Becker, 2014).

### *2.4 Location*

The research has reported marginal difference when the location of educational institutions is considered. Moreover, it is important to note that the large scale assessments don't offer insights

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

regarding location type where students live. Instead, only basic information about the location is given. The PASEC and SACMEQ offer a more complicated definition of urban and rural school areas than PISA. However, the issue mainly arises due to the complex geographical structure of the region.

### *2.5 Indicators to Measure the Equity and Learning Outcomes*

The underlying study uses various indicators to measure the educational equity across different nations. The research has mainly assessed the percentage of schools having well-qualified teaching staff, percentage of schools having access to a computer and the Internet, percentage of schools having access to basic sanitations facilities and percentage of schools providing HIV and sexuality education. A few more indicators have also been added to broaden the view and offer more productive insights.

First, we explore some study indicators that relate to the students. Afterwards, the research discusses the indicators that relate to educational institutions:

(1). Indicators measuring Students' Characteristics

1) Students %age in Grade 2, 3, Primary and Secondary level, achieving MPL in mathematics and reading:

Though, the research has regarded it as a most important indicator. However, obtaining the comparable data is a significantly difficult task. Hence, the study provided official proficiency test results to address this challenge.

2) Percentage of Students less than 5 years' old and:

*a. Developmentally on track of psychological wellbeing and learning.*

*b. Experiencing favorable learning environment at home.*

This indicator has been divided into two broad categories. Firstly, the study will collect the data from the teachers, caregivers or mothers and then findings would be computed to develop a composite indicator. The most appropriate way to collect the required insights is conducting household surveys. Because large scale assessment scales don't offer much information about this indicator except TIMSS and PIRLS Read Survey for parents as it might offer valuable insights about the children' basic competencies before entering into school.

Research conducted by Son & Morrison (2010) have examined the positive learning environment and development of children's academic skills and language. The results reflected an overall enhancement in the learning environment at home from 36 to 54 months with 31 percent of the preschoolers' parents showing substantial enhancement in the home environment whereas only 0.6 percent showed a decrease. Findings further reported that the change in home learning environment had a significant effect on language than academic skills. Moreover, the mothers' education level, fewer depression symptoms and less working hours were directly related to the overall home environment. This indicator will require the data of children 36 to 59 months old and take participation in under observation areas and a total number of children of same age. Assessments and surveys including Young Lives and Multiple Indicator Cluster etc. would offer the important information in this regard. Moreover, additional information could be obtained from the large scale assessments like LLESE, PIRLS and TIMSS.

3) Over-aged students' %age in Primary and Secondary level:

There can be various reasons behind the over-age, including lack of financial resources, drop out, late admission, repeating previous grades, etc. The indicator would assess whether the children have completed the 9 years of primary and secondary education and have achieved MPL in mathematics and reading as they turn fifteen years old. It has become the international norm to complete the 9 years of formal education within first fifteen years. Previous research findings have reported that educators should discourage the students from repeating the grades as it leads towards the over age and lesser learning achievement. On the basis of the repetition rate information, the over age progression could be successfully estimated. But, a significant amount of difficulty in involved in obtaining the repetition data for each nation. Almost all large scale assessments include the questions related to the repetition rate of students.

4) Percentage of Primary students having different home language:

This indicator represents the proportion of students enrolled in primary schools, whose home

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

language and instruction language is the same. Home language is the language spoken by the parents of the child, and it is the primary communication language outside the boundaries of the school. As discussed earlier, the difference between home and school language puts a significant influence upon learning effectiveness. The data about this indicator can be collected by administrating the surveys from schools on the instruction language and students' home language. However, it is important to note that schools don't necessarily maintain the record of student's first language. TIMSS and PIRLS have questions related to the test language speaking frequency at home and students' home language.

5) Students participating in the organized learning before entering into school: (Indicator n0 10)

This indicator represents the participation rate of students in the organized learning before becoming officially eligible to enter into school. It would assess the proportion of students in a particular age group who participate in various organized learning programs. It would basically assess the exposure of the children to such activities before starting the primary school. The study would mainly collect the data about this indicator through household surveys and administrative data. However, this indicator doesn't evaluate the program intensity, and only the participation is taken into consideration. It confines the ability of the researcher to propose a meaningful conclusion. Another problem is that it is hard to assess how many years a child has spent before entering into the primary school and children's exposure to such activities exactly one year before the primary entry age. PISA successfully differentiates the 1 year and above 1 year of pre-primary exposure. Whereas, PIRLS and TIMSS distinguish three different combinations including 1 year, 1-3 years and above 3 years.

6) Percentage of 15-year old students proficient in environmental and geo sciences:

This study indicator assesses the proportion of the fifteen-year-old students who demonstrate the proficiency in the knowledge of geo and environmental sciences. The indicator is the direct determinant of learning outcome in both study areas that are directly associated with the overall sustainable development. The OECD published a report in 2009 that assessed the understanding on fifteen years' old students in the two study areas that are directly related to the long term environmental sustainability. PISA in 2006 has focused more on the students' performance in science than in 2000, 2003, 2009 and 2012. While measuring the performance score in all dimensions, competencies in different science areas could be reported separately to develop separate performance scale, which can assess the proficiency level related to the students' performance in that area. PISA 2006 has not identified the conceptual framework for studying the geoscience and environmental science. However, out of total 108 questions, the assessment asked the 24 questions related to geo and environment sciences. The report published by OECD conducted an in-depth analysis of the PISA data to present the performance measures. Based on item response modeling, two performance indices were constructed from these questions, one for environmental science and the other for geoscience. Moreover, four proficiency levels were developed for each index from A to D and students falling below the level D were considerable incapable of demonstrating minimum proficiency in this study area. The results reported by the PISA 2006 calculated the indicator on the basis of a total number of fifteen-year-old students who achieve or exceed the MPL in geoscience and environmental science as a proportion of total students of same age group. The most suitable source for this indicator is the PISA 2006 assessment. However, the limitation of this indicator is that the under observation subjects are solely regarded as the key to sustainable development. Whereas, in reality, various other factors are also involved that are not assessed by this indicator.

7) Percentage of students experiencing sexual abuse and bullying:

This indicator includes the proportion of students that experience the abuse, sexual discrimination, violence, harassment, corporal punishment and bullying. Various sources have reported a clear increase in the sexual harassment and bullying in the primary and secondary schools. Here, the indicator would measure the proportion of students who unfortunately experienced any physical or mental torture during school timings. Almost all assessment tests provide useful information in this regard. For example, PISA offers useful insights about the students bullying, intimidating or hindering the learning process of other students. It also includes the students mocking or making fun of teachers. PIRLS and TIMSS contain questions related to students hit by kicking, hitting and shoving, etc. Information related to violence is also extracted from the school directors. Violence is a broad term that triggers various other issues like verbal abuse, intimidation including emailing, texting, verbal or physical fights or verbal abuse and intimidation among staff or teachers. Violence is not directly measured by the PASEC. Instead, teachers are enquired about the reasons for the students' drop out. One of the major reasons behind drop out is students abuse and inadequate security. Among all assessment tests, the TERCE offers more insights about the violence assessment at primary and secondary level. SACMEQ

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

study also asked the school heads about the occurrence of specific issues like sexual harassment and vandalism by students.

(2). Indicators measuring schools' characteristics

1) Percentage of schools giving HIV and sexuality education:

This indicator assesses the proportion of the schools providing HIV and sexuality education. The reason behind including this indicator is to examine the development towards the execution of sexuality education and HIV based life skills in secondary schools within their formal curriculum or part of extra-curricular activities. TIMSS 2011 asks questions about reproduction, and SACMEQ III involves a comprehensive analysis of HIV for students and school heads. On the basis of results of SACMEQ III, true percentage of students could be known who attend the lessons or classes on AIDS and HIV during the school test year.

2) Percentage of schools offering basic sanitation facilities:

This study indicator includes the access to the basic necessities including basic sanitation facilities, clean drinking water, and hand washing facilities. It evaluates the access of schools to these basic facilities that are essential for meeting the hygiene standards, ensuring an effective and safe learning environment. The standard for the basic drinking water facility involves the access to clean water points during the school timings. The water must be available to all students irrespective of class, gender and ethnicity. Standardized sanitation facilities should be available separately for each gender. Moreover, the sanitation facility should be available at a reasonable distance during the school timings. Additionally, in order to meet the hygiene standards, the schools need to provide standardized handwashing facilities including water, soap and washing point that would be available to all students. The data related to this indicator could be mainly collected through administrating surveys, as PISA, PIRLS and TIMSS don't offer adequate information in this regard. Toilets availability is also covered by this equity indicator.

3) Percentage of schools having access to computer, Internet and electricity:

This study indicator includes the access of the schools to the Internet, electricity and computer availability. Regional assessments reveal adequate information about access to schools to electricity. Moreover, approximately all assessment tests offer adequate information about the Internet availability. However, the tests don't reveal the purpose for which the Internet is used. Mostly, the Internet is used to enhance and enrich the learning and teaching process and is available to all students. TERCE includes the information related to the access of computers with Internet. However, the test doesn't reveal the purpose of using this resource. PISA 2012 involves only one question regarding the problem of Internet unavailability or computers shortage. Adequate access to the computer and the Internet allows the students to use different online communication services, as an exchange of data files, entertainment, news, e-mail and World Wide Web. The indicator would be assessed by dividing it into three different categories, computers, Internet and electricity.

4) Percentage of schools having qualified staff:

This indicator includes the proportion of qualified teachers according to the type of institution and level of education. The quality of education is directly dependent upon the knowledge and competencies of teachers. Hence, the role played by teachers cannot be ignored while assessing the overall education quality. The well qualified teaching staff is an asset for any country. Qualification of teachers is regarded as the most important indicator by international assessment bodies. Teachers play very important role in certifying the quality of education provided by educational institutions. Here, the indicator would measure the contribution of teachers that are well qualified. However, a limitation associated with this indicator is that each country has its own academic qualification requirements for teachers. It reduces the usability of global tracking as it will only assess the proportion of teachers meeting qualification requirements. Approximately all assessment tests offer the data required to assess this indicator except PISA as this test doesn't have any particular questionnaire to collect insights from educators. It has collected insights from school directors that might result in biased findings.

## 3. Results

This research presents the descriptive statistics and variables codification of skill, grade, year, survey and indicator. The study's main hypothesis is:

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

*H1: It is possible to compare different minimum thresholds in each assessment to assess the educational equity and learning outcomes.*

### *3.1 Results for the Proportion of Minimum Performers*

The research analysis provides the first existing database on the monitoring of SDGs for the education sector. The dataset includes 134,674 different estimations of proportions of students reaching the minimum level of proficiency. These estimations are the combination of 127 countries or localities, 7 different assessments (PISA 2012, TIMSS 2011, PIRLS 2011, PASEC 2014, TERCE 2013, SACMEQ 2007, PISA 2006), five different grades (2, 3, 4, 6 and 8), five different skills (mathematics, science, reading, environmental science and geoscience), fifteen indicators (including the proficiency levels but also additional material on the organization of the education systems), different subsamples (including gender, location of schools but also immigrant status). If we restrict the dataset to the proficiency levels for mathematics and reading and thus exclude specific statistics[1], the number of observations falls to approximately 5,000 observations.

In addition to the indicators relative to proficiency levels, the research also prepared data for mean scores. Moreover, since the SDGs focus on both quantity and quality of education, data for survival rates were also included.

It is possible to compute an average value of each indicator by combining all available data concerning the proportion of students reaching the minimum proficiency level. The different dimensions are the level of education, the skill tested, the year of evaluation and the survey. The overall proportion of students reaching the minimum proficiency level is equal to 68.91 with a standard deviation of 22.36. Data are available for 126 countries or localities. The focus on lower grades of primary education highlights the real need to expand the coverage of countries, since data are available only for 25 countries (Table 2). In fact, these countries are the ones which took part at PASEC and TERCE, since there are the only two assessments which test pupils are early grades (grade 2 for PASEC and grade 3 for TERCE). The average level of minimum performers for early grade of primary education is thus lower than the overall mean found above (50.29% versus 68.92%). On the contrary to early grades, data for upper grades of primary education is available for more than 100 countries. This is mainly due to the availability of TIMSS, PIRLS and SACMEQ assessments which include the tests for students from grades 4 and 6, in addition to the remaining assessments (PASEC, TERCE). The proportion of minimum performers inside this level is the highest among all levels (75.35% against 50.29 for lower grades of primary education). Lastly, the focus on lower secondary education provides a quite high proportion of minimum performers (67.14%), although the number of countries is lower than the upper grades of primary level (90 versus 101 countries).

*Table 2: Descriptive statistics on minimum performers, by level of education*

| Level | Countries | % Developing countries | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|---|---|
| **Lower Primary** | 25 | 100 | 50.29 | 21.49 | 09.60 | 96.70 |
| **Upper Primary** | 101 | 64 | 75.35 | 22.97 | 06.66 | 99.64 |
| **Lower Secondary** | 90 | 51 | 67.14 | 20.47 | 14.50 | 98.90 |
| **Total** | 125 | 62 | 70.96 | 20.64 | 8.03 | 97.90 |

Note: The minimum proficiency levels are not directly comparable between assessments. A direct comparison should be made with caution. Lower primary includes grades 2 & 3, upper primary includes grades 4, 5 & 6, lower secondary includes grades 7, 8 & 9.

The focus on different regions over the world highlights the low performance of Northern Africa, Western Asia and Sub-Saharan Africa (Figure 1). If we now distinguish between education levels (Figure 2), Northern Africa is the lowest performer region for upper primary education where less than 30% of pupils reach the minimum proficiency level, while almost all population from developed countries reach this level[2]. In lower secondary education, although the number of countries with data is very low, the lowest performers are located in Sub-Saharan Africa. In the meantime, only 80 percent of students from Eastern Asia and Developed countries reach the minimum proficiency level in lower

[1]The research excluded all information relative to data which is not directly related to the proportion of students reaching the minimum proficiency level in mathematics and reading.

[2]The research used the region classification provided to follow the Millennium Development Goals. However, countries in Eastern Asia (China, Hong-Kong, Mongolia and Republic of Korea) have been included inside the South-Eastern Asian countries. We called this new sub-group "Eastern Asia".

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

secondary schools. This important findings highlights the need to not only focus on developing countries but also on the richest economies like France or the United States, where approximately 12% of students do not reach the minimum proficiency level.

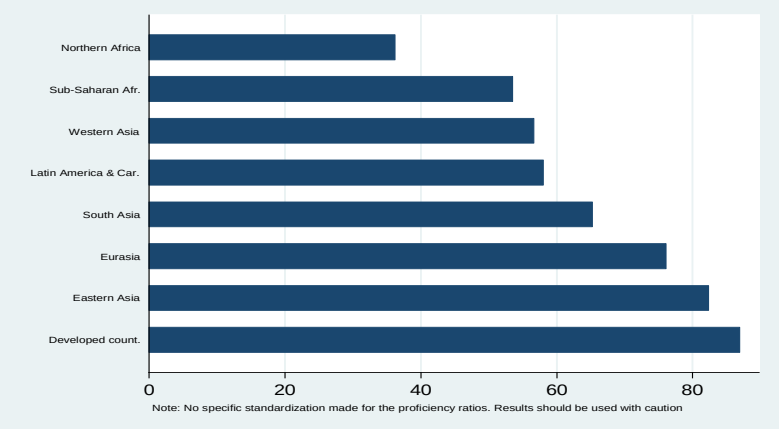



*Figure 1: Proportion of minimum performers around regions, all levels, mean of maths and reading scores*

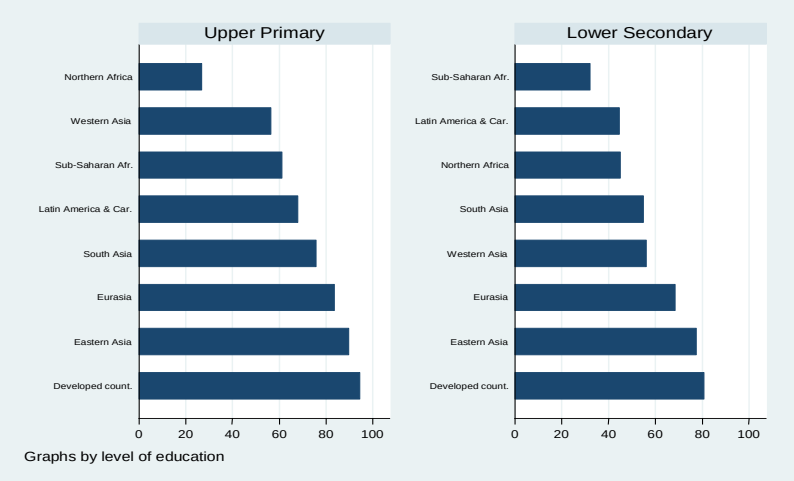



*Figure 2: Proportion of minimum performers by regions and levels, mean of maths and reading scores*

These findings clearly show that results may not be generalized to all countries, including developed economies. The research therefore compute average proportions of students reaching the minimum proficiency levels for each country. Countries with the highest proportion of minimum performers are Belgium Flemish, Canada (Quebec) and Republic of Korea. Most importantly, results for some developing countries show a high performance, although data are only available for primary education. For instance, approximately 90% of pupils from Tanzania and Kenya reach the minimum proficiency level in upper primary education, while it is the case for only 23% in Morocco. Indeed, the countries with the lowest performance are mainly North African and Sub-Saharan African countries. In Yemen, less than 10% of students from grade 4 reach the minimum proficiency level, while most of pupils from the Netherlands reach this threshold. The SDGs aimed at focusing on both quantity and quality education, it may be interesting to analyze the relationship between the proportion of minimum performers and survival ratio to the last grade of each level of education. Figures 3 (a) & 3 (b) show this relationship for respectively primary and secondary education. As expected, the relationship is positive in both cases (R squared equal to 0.30 and 0.08 in primary and secondary education respectively). However, countries with a quite high survival rate may not be always able to permit to all their children to reach the minimum proficiency level. An interesting comparison may be between Morocco and Costa Rica. While in both countries, most pupils can be enrolled until the last grade of primary education (i.e. almost 90%), only one fifth of them reach the minimum proficiency level in Morocco, compared to more than four-fifths in Costa Rica. The same comparison can be made for lower secondary education with the cases of Indonesia and Czech Republic, where the proportion of minimum performers are around 80% in the latter countries, compared to less than 40% for Indonesia

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

(Figure 3 (a)).

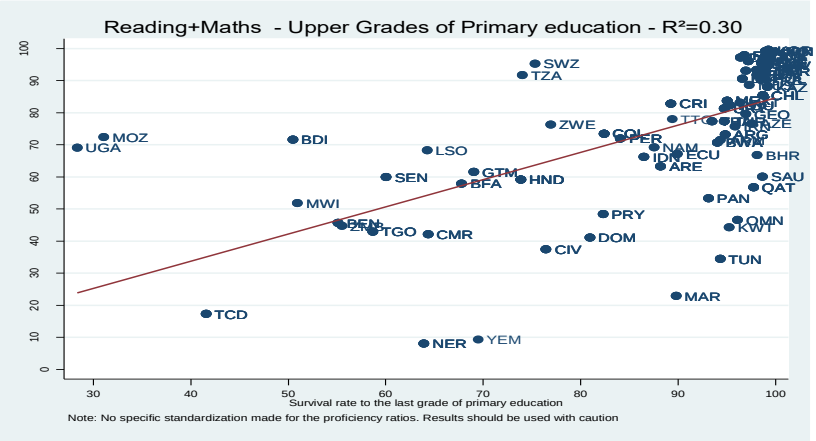



*Figure 3 (a): Relationship between proportion of pupils reaching the minimum proficiency level and survival rate to last grade, primary education*

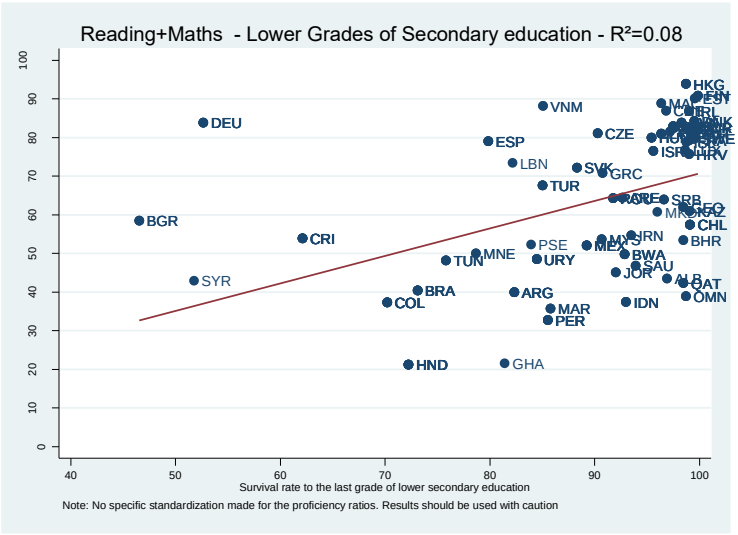



*Figure 3(b): Relation between proportion of pupils reaching the minimum proficiency level and survival rate to last grade, lower secondary education*

In order to obtain a universal education measure with a quality dimension, we computed quality adjusted survival rate to the last grade of primary (Figure 4 (a)) and secondary (Figure 4 (b)) education by multiplying the proportion of minimum learners with the survival rates. It should be noted that no specific adjustment was made in order to make comparable the proportion of pupils reaching the minimum proficiency levels. Therefore, these results should be taken with caution. Among all sub-Saharan African countries with data, Mauritius appears to be the most performing country with approximately 85% of pupils reaching both the last grade and the minimum level in primary education. On the contrary, results for mainly Francophone countries are highly dramatic: in Niger and Chad, less than 10% of pupils can be considered as reaching the goal of a qualitative and complete primary schooling. Very few SSA countries took part to at least one achievement in secondary education. Therefore, in Figure 4 (b), we report results for secondary level for both SSA and Latin American countries. Even in the top-performing country – Chile – only half of students can reach the last grade of secondary education and the minimum proficiency level. In countries like Honduras or Colombia, less than one fifth of students are in this case. Results for secondary education highlight the need to provide more data and to expand the quality of education in that level.

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

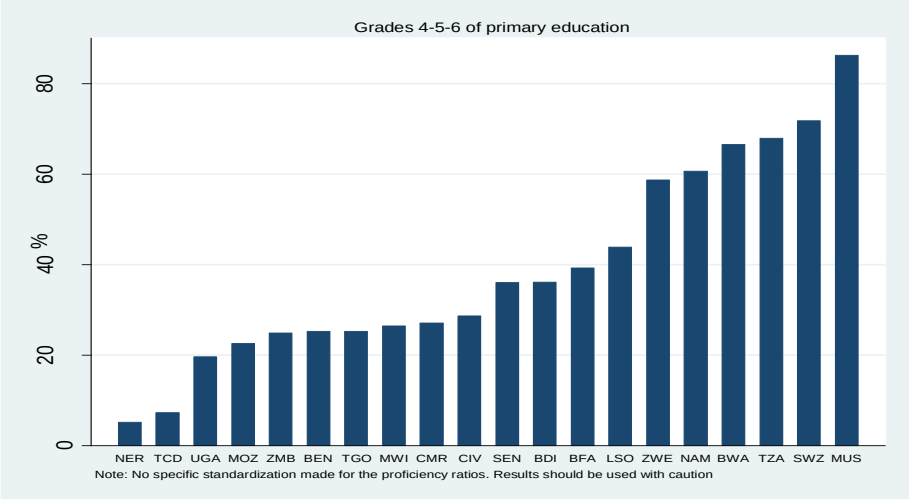



*Figure 4 (a): Distribution of the quality adjusted survival rate to the last grade, Primary education, Sub-Saharan African countries*

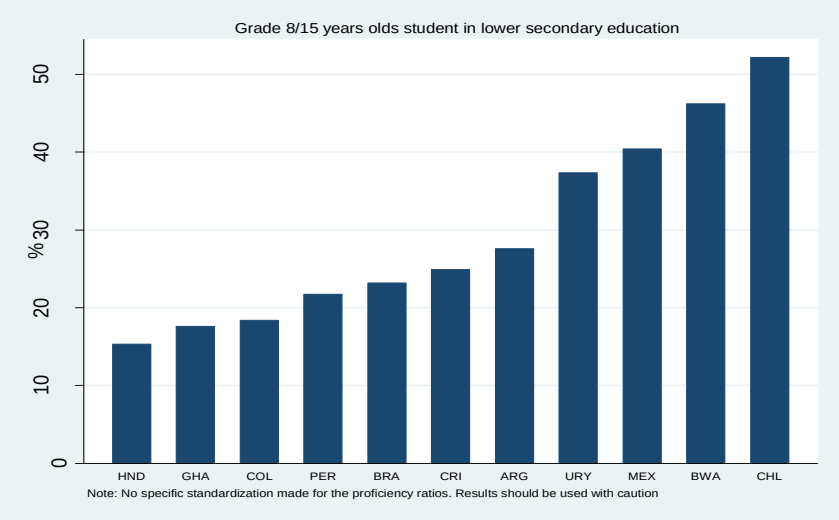



*Figure 4 (b): Distribution of the quality adjusted survival rate to the last grade, Lower Secondary education, Sub-Saharan African & Latin American countries*

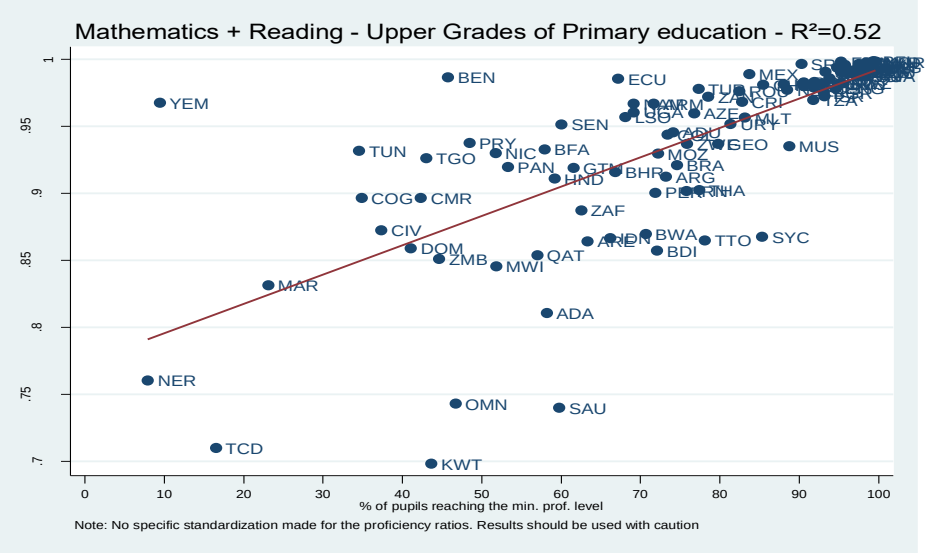



*Figure 5: Minimum Learning and Gender Parity Index, Upper Primary Education*

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

### *3.2 Results for Equity Issues*

The SDGs not only focused on school improvement but also on solving equity issues. A higher proportion of pupils reaching the minimum proficiency level may be correlated with a higher gender parity ratio, since if almost all students are able to learn the basics, then there should not be a high difference between each gender. As shown in Figure 5, there is clear positive and significant relationship between the two concepts for upper primary education (R squared equal to 0.52). This relationship holds in other levels. However, countries with the same proportion of minimum performers may have different gender parity ratios. A good comparison can be done between Oman and Benin. In both countries, almost half of pupils reach the minimum proficiency level in primary education. However, while the parity is obtained in Benin, the GPI is equal to 0.75 in Oman.

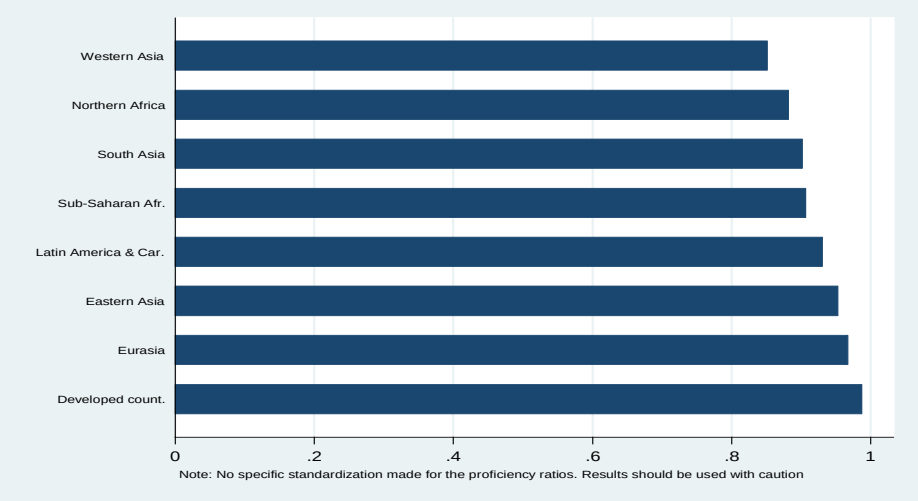



*Figure 6: Minimum Learning and Gender Parity Index, Upper Primary Education, by regions*

In this case, a GPI equal to 0.75 means that while only 40% of boys reach the minimum proficiency level, this is the case of about 53% for girls. A focus on performance among regions show that the parity is almost achieved in developed countries whereas a great improvement should be done in regions like Western Asia (for instance Jordan, Kuwait and Oman) and Northern Africa (Algeria, Egypt or Morocco). (Figure 6).

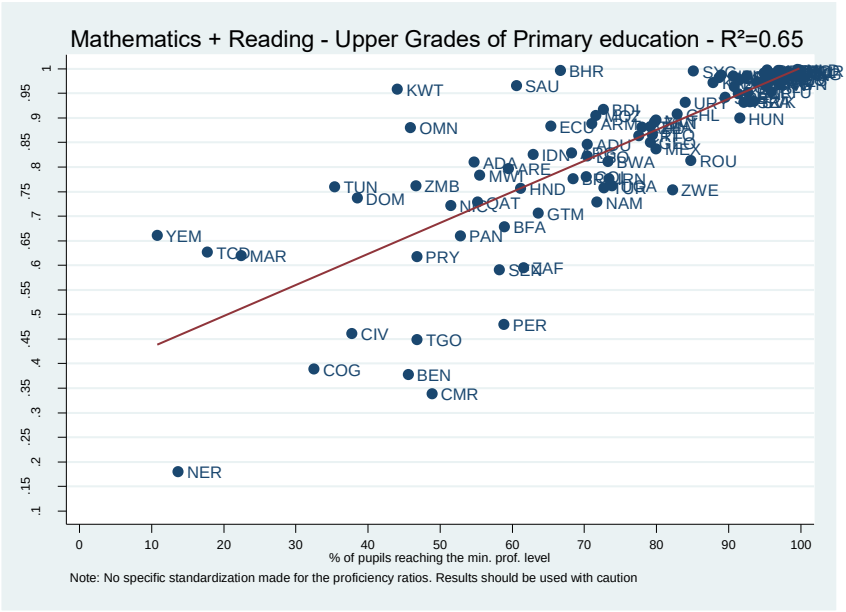



*Figure 7: Minimum Learning and Location Parity Index, Upper Primary Education*

The gender parity should not be the only dimension of equity, although it is widely used in the literature. The distinction between rural and urban areas reveals higher inequalities within countries

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

(Figure 7). Similarly to the GPI, we observe a clear and positive relationship between the proportion of minimum performers and the location parity index (R squared equal to 0.65 in upper grades of primary education). However, significant differences still exist between countries with the same proportion of minimum learners. A comparison between Niger and Chad shows that while the proportion of minimum learners is quite low in both countries (less than 20% of pupils), the location parity index diverges greatly between the two countries. In Niger, only 4% of pupils from rural schools reach the minimum proficiency level, against 21% of pupils from urban schools. Hence, the location parity index is equal to 4/21=0.19 for Niger. These results are confirmed for both reading and mathematics. On the contrary, the equity issues are less pronounced for Chad, although there are still very large (respectively 22% and 16% for urban and rural areas).

### *3.3 Results for other Indicators*

Besides the usual performance of students in student achievement tests, it is also possible to obtain additional information related to schools, teachers and students' characteristics. Based on the background questionnaires distributed to school principals, teachers and pupils, we extracted some important information regarding to statistics which are needed in order to monitor the SDGs related to the education systems. We reported the different indicators for which some alternative data can be found from regional and international student achievement tests. As it can be seen, very often the variables are proxies of each indicator. For instance, there is useful information about the overaged students in Primary and Secondary level in nearly all assessments. This indicator is based on pre-primary education: "Participation rate in organized learning (one year before the official primary entry age)". However, this participation does include not only public pre-primary education but also alternative possibilities of learning. Another criteria is that this learning should be made one year before the official primary entry age. There is no specific data which covers all these criteria in assessments. In the meantime, there is some possibilities to approximate a value for this indicator. For instance, in PISA 2012, students are asked if they attended ISCED 0. Three different answers are available: 1 = No, 2 = Yes, for one year or less 3 = Yes, for more than one year. We considered that regardless to the number of years devoted to pre-primary education, an answer which was reporting "one year or less" could be considered as fitting with the criteria of being enrolled in pre-primary education. In other assessment, similar data are available. In TIMSS and PIRLS, a derived variable provides useful information for both reading and mathematics but only for grade 4 pupils. Regional assessments also include questions dealing with pre-primary education. While in PASEC, pupils were only asked whether they took part or not at pre-primary education, in SACMEQ and TERCE, the duration of such learning was also informed. Similarly to the minimum proficiency levels, results for these indicators should be used with caution, since comparability between assessments is not directly possible. A direct comparison between each assessment is made in Table 3.

*Table 3: Proportion of pupils who attended pre-primary education*

| Assessment | Grade | Number of countries (% Developing countries) | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|---|---|
| **TIMSS 2011** | 4 | 35 (40) | 74.20 | 21.86 | 23.27 | 99.12 |
| **PIRLS 2011** | 4 | 51 (41) | 76.00 | 20.89 | 23.43 | 99.15 |
| **SACMEQ III** | 6 | 15 (100) | 59.80 | 25.24 | 25.65 | 97.96 |
| **TERCE** | 3 | 15 (100) | 58.25 | 11.76 | 37.48 | 75.91 |
| **TERCE** | 6 | 15 (100) | 58.01 | 12.66 | 35.67 | 77.65 |
| **PASEC** | 2 | 10 (100) | 27.12 | 11.89 | 10.90 | 49.90 |
| **PASEC** | 6 | 10 (100) | 28.21 | 11.62 | 12.00 | 46.60 |
| **Mean*** | | 87 (66) | 65.48 | 25.10 | 11.45 | 99.15 |

* Since some countries took part to different assessments and grades, a specific computation was made for obtaining the global mean. Mean values are obtained by using the arithmetic mean within each country firstly, regardless to the assessment, and then by computing a global mean of country means.

Mean proportion of students attending pre-primary education show strong differences between regions (Figure 8). A direct comparison with the data available from the UNESCO Data Center permit to show high differences. Although a positive correlation exists between the gross enrolment ratio and our measure of overaged students in Primary and Secondary level (R squared equal to 0.50), strong differences can be found for specific countries. This is especially the case of Latin American countries like Nicaragua where the difference is equal to 50 percentage points: while the official GER is equal to 70%, only 18% of pupils report to have participated in an organized learning before primary education. A closer look to the TERCE data shows that almost 27% of data are missing values. However, only 37% of pupils from grade 3/6 actually reported to have participated at pre-primary education, while the pre-

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

primary enrolment rate is equal to 58% according to the UNESCO.

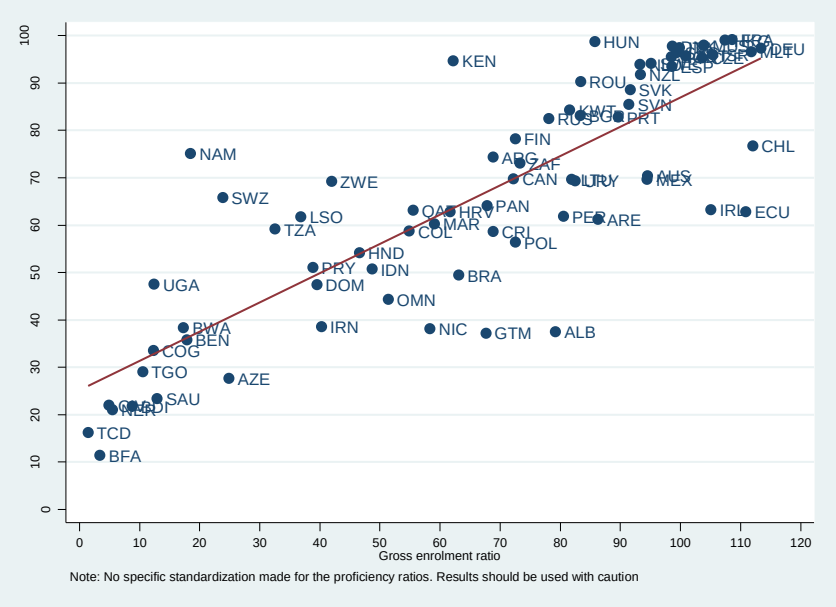



*Figure 8: Comparison between gross enrollment and participation rate to pre-primary education ($R^2$= 0.69)*

The research has extracted further insights related to the characteristics of students, teachers and schools. The survey has been conducted to collect data from school principals, teachers and students and important insights have emerged to monitor the sustainable development goals in the educational sector. Data from regional and international assessment tests has been collected related to different indicators chosen by the research. However, in most of the cases, it was not possible to collect the data for original indicator and proxies have been used for their presentation. The data has been collected on the basis of criteria that learning should be made one year before the official primary entry age. There is no specific data which covers all these criteria in assessments. In the meantime, there are some possibilities to approximate a value for this indicator. PIRLS and TIMSS have a derived variable that offers meaningful insights for mathematics and reading for early primary students. While in PASEC, students were only asked whether they took part or not at pre-primary education, in SACMEQ and TERCE, the duration of such learning was also informed. Similarly, to the minimum proficiency levels, results for these indicators should be used with caution, since comparability between assessments is not directly possible.

## 4. Discussion and Limitations

The previous section has reported the results on the basis of different proficiency levels. Overall, the research has highlighted that all assessment tests are significantly different from each other and analysts need to be cautious while conducting the comparison over the assessments made by these tests. First of all, it is important to note that different assessment tests might have taken insights from a population having significantly different characteristics. The difference is high between PISA and other assessment tests. The main reason for this difference is that PISA assessment is primarily based on the students' age, whereas, all other assessment tests base their assessments on grades.

The difference becomes even more prevalent when countries with high drop-out or repetition rate are involved. Hence, the assessments made on the basis of groups might ignore a substantial percentage of students repeating the class whereas the PISA assessment made on age would consider this group as well. Moreover, PISA focuses on the schools instead of classes whereas SACMEQ and TIMSS take classes and schools, giving more preference to the class based assessment. PISA doesn't stratify the classes and employs the random sampling technique while collecting the insights.

Another significant difference between these assessment tests is that benchmark for these tests might differ in terms of difficulty. Students might find easy to score higher in TIMSS as compared to TERCE. Each assessment test has unique psychometric methodology due to which significant difference prevails between the items selected by each assessment test. The difficulty level of each item

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

also differs and hence the proficiency level is defined differently by each assessment test, resulting in varying results for participating countries.

Different assessment tests might include different content to be tested. For example, PASEC and PISA focus more on competency skills whereas other assessment tests focus on the common curriculum of each country. This difference results into significant variance among countries that are more focused on the content knowledge instead of competencies. Mostly, the developing countries tend to focus on the content knowledge whereas advanced countries tend to focus on skills and competencies. Another source of difference is the difference between the time and duration.

The difference between assessment tests results into a different comparison of subpopulations. For example, the difference among students with varying socioeconomic backgrounds might not produce same findings even if the tests assess the same items. The difference mainly arises from the prevailing context. The underlying study has highlighted these differences for indicators that focus on the pre-primary education. These all differences result in wide variation in findings of these assessment tests for the percentage of students meeting the MPL. The underlying study has presented the difference by computing the means percentage of each assessment. The PASEC assessment test has been proved as the most difficult assessment, passing only forty percent of the total population. Whereas, TIMSS has been the most lenient assessment test, allowing around eighty percent of the students to reach the MPL. However, the difference between the two tests is explained by the difference in the economic condition of participating countries. The countries participating in TIMSS assessment test have average GDP of $33,000 in 2011 whereas, countries participating in the PASEC assessment in 2014 have average GDP of around $2,200 and less.

A closer look at prevailing differences suggests that the unique characteristics of each assessment test are mainly responsible for it. The underlying study conducted a regression estimation of the percentage of students who attained the MPL. The first column represents that percentage of students achieving minimum proficiency level are approximately 35 percent lower for the countries that participated in the PASEC test. Whereas, in the case of PISA, the difference was low yet statistically significant. The column two represents the dummy variable for reading test, producing the similar results. However, in the case of PIRLS, the effect is insignificant. In order to present the differing proportion of developing countries, column 3 presents the GDP. There is no significant change for PISA assessment as the coefficients carry a low value in regional assessments. The column 4 represents the country fixed regression effects. The study has restricted the sample for countries taking part in a minimum of 2 different assessments. The estimation has not included the PASEC countries as they didn't participate in other assessments. Additionally, the dummy inclusion for the reading allows the research to control fixed effects. PIRLS and PISA have produced the lower performance of students and TERCE, and SACMEQ made an over estimation of the results as compared to the TIMSS control assessment. Table 3 represents the analysis of participating countries. The study has grouped the participating countries in two assessments and has compared the percentage of students attaining the minimum proficiency level. Botswana presents the highest difference as SACMEQ assessment has over -estimated the results of 33 percent of the students attaining the minimum proficiency level as compared to the PIRLS assessment. Moreover, Colombia has also participated in two assessments, and there is a significant difference between both assessments. But the difference is insignificant for Honduras as the students attaining minimum proficiency are same among both assessments. Another important insight is that the difference is more prevalent in the case of developing countries like Tunisia and Kazakhstan than the advanced countries like Finland and New Zealand. Moreover, the TIMSS has consistently produced over-estimated results than the PISA assessment. Overall, the PISA assessment has provided approximately 17 percent low students' performance than TIMSS.

The unique characteristics of each assessment such as the content or grade of the test are mainly responsible for the difference in the results. But, the difference between the girls and boys achieving minimum proficiency level is unexplainable between different assessments. The difference would have remained similar even if the content being assessed differed marginally. The study only included the countries having a statistically significant difference between different assessments. Chile and Indonesia have reported strong differences. Whereas, other countries have not offered strongly contradictory results, yet the difference cannot be ignored. The examples are Turkey and Tunisia. The study has only included the "gender". The difference might be larger of research would have considered other variables as well. The hypothesis is confirmed upon calculating the urban/rural parity index. On average, PISA results tend to over-estimate the equity issues related to the location of schools, compared to TIMSS assessment.

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

## 5. Recommendations

After discussing the limitations associated with each assessment test, this section of research would offer useful insights for practical implementation. These recommendations, if applied, can significantly enhance the current large scale assessment tests. The potential enhancements would be helpful in conducting a better assessment of SDGs towards the RSATs and ISATs.

### *5.1 The Need of a Standardized Use of Assessments*

In most of the cases, the organizations don't actively cooperate during the assessment process. For example, currently, the IEA and OECD don't share common perspective to offer unified results about the learning skills at primary and secondary level. Numerous countries participate in both assessment tests without any such cooperation and collaboration. Unfortunately, both organizations have failed to propose a common agenda though both cover the African region. Hence, there is a need to introduce some standardized assessments to encourage the collaboration among different organizations. The underlying study would confer about the databases codification to bring the standardization in current regional and international achievement tests. Such standardization would offer an easier way to analyze and interpret different assessment simultaneously. The table 4 presents the codification of an important study variable. The survey name would be coded with specific label (idsvy) and numerically. Additionally, accurate results are dependent upon various other variables (including skill, level, countries, regionv2). The study has also divided the "level" variable into five sub categories including pre-primary, lower primary, upper primary, lower secondary and upper secondary. In our study, the lower primary level includes grades 1-3 of primary education, while the upper primary level focuses on grades 4-6 of the same level. Similarly, for secondary education, we consider lower secondary as grades 7-9 while upper secondary education will include the remaining grades until grade 12. The study has also named the "equity" that has divided the overall sample into various sub categories.

*Table 4: Codification of variables used in the study*

| Dimension coded | Name | Description of the codification |
|---|---|---|
| **Assessment** | idsvy | 1=TIMSS; 2=PIRLS; 3=SACMEQ; 4=ELCE; 5=PASEC; 6=PISA |
| **Skill** | skill | 1=Mathematics; 2=Science; 3=Reading; 4=Environmental Science; 5=Geoscience |
| **Level** | level | 1=Pre-Primary; 2=Lower Primary; 3=Upper Primary; 4=Lower Secondary; 5=Upper Secondary |
| **Countries** | cnt | ISO classification |
| **Regions** | regionv2 | Millennium Development Goals classification: 1=Developed countries; 2=Eurasia (countries in CIS); 3=Northern countries; 4=Sub-Saharan Africa; 5=Latin America & Caribbean; 6=Eastern Asia; 7=South Asia; 8= South-Eastern Asia; 9=Western Asia; 10=Oceania |
| **Equity** | equity | 1=Full population ; 2=Gender; 3=Location of schools; 4=Language at home; 5=Socio-Economic Status; 6=Immigrant Status; 7=Indigene population; 300=Original values of location of school; 600= Original values of immigrant status; 400= Original values of language at home |
| **Proficiency level** | plevel | Depend on each assessment. Varies between 1 and 10. Additional code: 0=Not applicable |
| **Proportion of students at the threshold** | value & value_se | Value of the proportion of students reaching the given threshold for the subpopulation. In addition, variable "value_se" provides the standard error of the given "value". |
| **Missing values** | mv | Proportion of missing values in the category considered. Can vary between 0 and 100% |
| **No. of observations** | obs | Number of (unweighted) observations included in the combination. |
| **Confidence intervals** | confintl & confindu | Left and Right values of confidence intervals computed for the proportion of students reaching the given proficiency levels ("value"). |

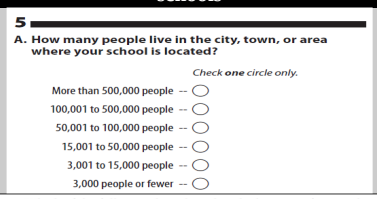
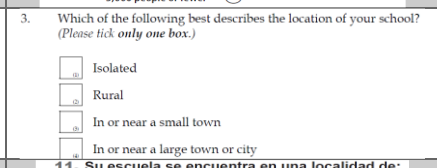
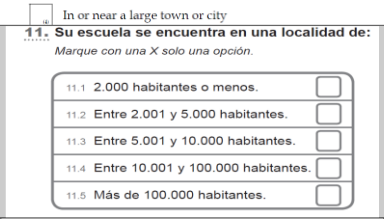
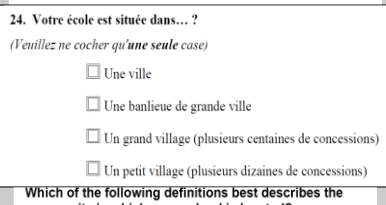
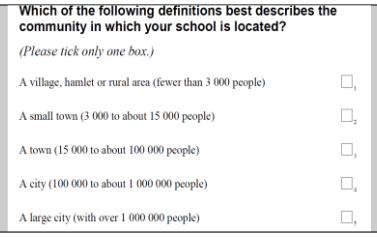
The sample has been divided by gender (equity=2), type of location (equity=3), language at home (equity=4), socio-economic status (equity=5), immigrant status (equity=6) and indigenous populations (equity=7). The research has also coded the original values by multiplying the value by hundred. The data base offers the findings for MPL and has also presented a new variable "p level" that indicates which MPL is concerned by the statistics as well as general characteristics of education systems. Here, the value of the variable p level = 0. The percentage of students reaching the MPL has been presented by "value" along with the standard error "value_se".

Table 5 presents the chosen methodology to develop a global indicator on variable "location" by differentiating between the rural and urban schools. The questions for each assessment differ, hence, the specific rule has been followed by the recorded version of the variable, "location". The school directors have been consulted to collect the information to conduct the assessment tests like TIMSS.

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

PIRLS and PISA. Whereas, PASEC and SACMEQ mostly referred to different meaning of rural and urban areas. The main reason behind this difference is that in African region, it is difficult to differentiate the schools on the basis of location.

*Table 5: Example of standardization of the equity dimension relative to the location*

| Assessment | Original question relative to the location of schools | Proposed standardized codification |
|---|---|---|
| **TIMSS/PIRLS 2011** |  | Rural=”Fewer than 3,000 people” or “3,001 to 15,000 people”. Urban =”15,001 to 50,000 people” or “50,001 to 100,000 people” or “100,001 to 500,001 people” or “More than 500,000 people” |
| **SACMEQ III** |  | Rural=”Isolated” or “Rural”<br>Urban=”In or near a small town” or “In or near a large town or city” |
| **TERCE** |  | Rural=”2.000 habitantes o menos” or “Entre 2.001 y 5.000 habitantes” or “Entre 5.001 y 10.000 habitantes”. Urban =”Entre 10.001 y 100.000 habitantes” or “Más de 100.000 habitantes” |
| **PASEC 2014** |  | Rural=”Un grand village (plusieurs centaines de concessions” or “Un petit village (plusieurs dizaines de concessions” Urban=”Une ville” or “Une banlieue de grande ville” |
| **PISA 2012** |  | Rural=”A village, hamlet or rural area (fewer than 3 000 people)” or “A small town (3 000 to about 15 000 people)” Urban=”A town (15 000 to about 100 000 people)” or “A city (100 000 to about 1 000 000)” or “A large city (with over 1 000 000 people)” |

### *5.2 Developing a Standardized Database at International Stage*

*Table 6: Standardized codification of international and regional student assessments*

| Order | Meaning | Values |
|---|---|---|
| **First digit** | Name of the assessment | 1=TIMSS; 2=PIRLS; 3=SACMEQ; 4=ELCE; 5=PASEC; 6=PISA |
| **Second digit** | Round for each assessment | For TIMSS: 1=FIMS/FISS; 2=SIMS/SISS; 3=TIMSS 1995; 4=TIMSS 1999; 5=TIMSS 2003; 6=TIMSS 2007; 7=TIMSS 2011; 8=TIMSS 2015<br>For PIRLS: 1=SSS; 2=RLS; 3=PIRLS 2001; 4=PIRLS 2006; 5=PIRLS 2011; 6=PIRLS 2016<br>For SACMEQ: 1=SACMEQ I; 2=SACMEQ II; 3=SACMEQ III; 4=SACMEQ IV<br>For ELCE: 1=LLECE 1; 2=SERCE; 3=TERCE<br>For PASEC: 1=PASEC between 1996 & 2005; 2=PASEC between 2006 and 2013; 3=PASEC 2014<br>For PISA: 1=2000; 2=2003; 3=2006; 4=2009; 5=2012; 6=2015 |
| **Third digit** | Skill assessed | 1=Mathematics; 2=Science; 3=Reading; 4=Environmental science; 5=Geoscience |
| **Fourth and fifth digit** | Grade assessed | 02=Grade 2; 04=Grade 4, and so on. |

The underlying research has mainly focused upon the latest large scale assessment tests. Some adjustments might be required to allow a comparison over the period of time. For example, there is need to rename certain variables to include the previous TIMSS assessments. The study has proposed the specific codification for this purpose. Table 6 discusses the matter in detail. For example, the code 17104 has the following meaning: 1=TIMSS; 7=2011; 1=Mathematics; 04=Grade 4. Results provided

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

for this code will hence include results for students in mathematics who are in grade 4 and tested in the TIMSS 2011 assessment.

Besides standardizing, there is a need to make suitable adjustments for all understudy variables. For example, the codification of name of the equity variable might change with the passage of time. Analysts cannot make an effective comparison unless some specific adjustments are made in each assessment. For instance, the codification of the variable "location" changed with time. Table 7 represents this change.

*Table 7: Example of over-time standardization of the equity dimension relative to location in TIMSS assessments*

| Year of the assessment | Original question relative to the location of schools | Proposed standardized codification |
|---|---|---|
| **1995** | **In what type of community is your school located?**<br>A geographically isolated area<br>Village or rural (farm) area<br>One on the outskirts of a town/city<br>One close to the center of a town/city | Rural= "A geographically isolated area" or "Village or rural (farm) area" ;<br>Urban ="One on the outskirts of a town/city" or "One close to the center of a town/city" |
| **1999** | **In what type of community is your school located?**<br>A geographically isolated area<br>Village or rural (farm) area<br>One on the outskirts of a town/city<br>One close to the center of a town/city | Rural= "A geographically isolated area" or "Village or rural (farm) area" ;<br>Urban ="One on the outskirts of a town/city" or "One close to the center of a town/city" |
| **2003** | 3<br>**How many people live in the city, town, or area where your school is located?**<br>*Fill in **one** circle only*<br>More than 500,000 people<br>100,001 to 500,000 people<br>50,001 to 100,000 people<br>15,001 to 50,000 people<br>3,001 to 15,000 people<br>Fewer than 3,000 people | Rural= "Fewer than 3,000 people" or "3,001 to 15,000 people".<br>Urban = "15,001 to 50,000 people" or "50,001 to 100,000 people" or "100,001 to 500,001 people" or "More than 500,000 people" |
| **2007** | 2<br>**How many people live in the geographic area where your school is located?**<br>*Fill in **one** circle only*<br>More than 500,000 people ①<br>100,001 to 500,000 people ②<br>50,001 to 100,000 people ③<br>15,001 to 50,000 people ④<br>3,001 to 15,000 people ⑤<br>3,000 people or fewer ⑥ | Rural="Fewer than 3,000 people" or "3,001 to 15,000 people".<br>Urban = "15,001 to 50,000 people" or "50,001 to 100,000 people" or "100,001 to 500,001 people" or "More than 500,000 people" |
| **2011** | 5<br>**A. How many people live in the city, town, or area where your school is located?**<br>*Check **one** circle only.*<br>More than 500,000 people<br>100,001 to 500,000 people<br>50,001 to 100,000 people<br>15,001 to 50,000 people<br>3,001 to 15,000 people<br>3,000 people or fewer | Rural= "Fewer than 3,000 people" or "3,001 to 15,000 people".<br>Urban = "15,001 to 50,000 people" or "50,001 to 100,000 people" or "100,001 to 500,001 people" or "More than 500,000 people" |

**Acknowledgement:**

This study is supported by Natural Science Foundation of China Ministry of Education (Youth Program), Monitoring Learning Outcomes towards Sustainable Development Goals and Indicators in the Basic Education of Developing Countries ((Research on the indicator monitoring and development trend of the United Nations 2030 Sustainable Development Goals in the field of basic education in developing countries)) Grant NO. 18YJC880020, 2018-2020.

ISSN 2706-6827 Vol. 3, Issue 21: 47-65, DOI: 10.25236/IJFS.2021.032107

**References**

*[1] Sharma, L., & Dua, M. B., (2013). A Study of the Minimum Level of Learning in Biology for Ninth Class of UP Board of Secondary Education. International Journal of Educational Planning & Administration. Vol. 3, No. 1, pp. 35-50.*

*[2] Cornwell, C., Mustard, D. B., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. Journal of Human Resources, 48(1), 236-264.*

*[3] Anderson, L. W. (1990). Opportunity to learn and the National Assessment of Educational Progress: An analysis with recommendation. Unpublished manuscript.*

*[4] Traiman, S. L. (1993). The Debate on Opportunity-To-Learn Standards. National Governors' Association, 444 North Capitol Street, Washington, DC 20001-1512..*

*[5] Ensminger, M. E., Fothergill, K. E., Bornstein, M. H., & Bradley, R. H. (2003). A decade of measuring SES: What it tells us and where to go from here. Socioeconomic status, parenting, and child development, 13-27.*

*[6] White, K. R. (1982). The relation between socioeconomic status and academic achievement. Psychological bulletin, 91(3), 461.*

*[7] Sutton, A., & Soderstrom, I. (1999). Predicting elementary and secondary school achievement with school-related and demographic factors. The Journal of Educational Research, 92(6), 330-338.*

*[8] Andon, A., Thompson, C. G., & Becker, B. J. (2014). A quantitative synthesis of the immigrant achievement gap across OECD countries. Large-scale Assessments in Education, 2(1), 7.*

*[9] Fuligni, A. J. (1997). The academic achievement of adolescents from immigrant families: The role of family background, attitudes, and behavior. Child development, 68(2), 351-363.*

*[10] Steinberg, L., Blinde, P. L., & Chan, K. S. (1984). Dropping out among language minority youth. Review of Educational Research, 54(1), 113-132.*

*[11] Son, S. H., & Morrison, F. J. (2010). The nature and impact of changes in home learning environment on development of language and academic skills in preschool children. Developmental psychology, 46(5), 1103.*