TECHNICAL NOTE N° IDB-TN-03003

The State of Education in Latin America and the Caribbean: Learning Assessments

Elena Arias Ortiz
Ximena Dueñas
Cecilia Giambruno
Angela López

Inter-American Development Bank
Education Division

September 2024

The State of Education in Latin America and the Caribbean: Learning Assessments

Elena Arias Ortiz
Ximena Dueñas
Cecilia Giambruno
Angela López

Inter-American Development Bank
Education Division

September 2024

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
The state of education in Latin America and the Caribbean 2024: Learning assessments / Elena Arias Ortiz, Ximena Dueñas, Cecilia Giambruno, Ángela López.
p. cm. — (IDB Technical Note; 3003)
Includes bibliographic references.
1. Education-Evaluation-Latin America. 2. Education-Evaluation-Caribbean Area. 3. Education-Research-Latin America. 4. Education-Research-Caribbean Area. 5. Education and state-Latin America. 6. Education and state-Caribbean Area. I. Arias Ortiz, Elena. II. Dueñas, Ximena. III. Giambruno, Cecilia. IV. López Sánchez, Ángela. V. Inter-American Development Bank. Education Division. VI. Series.
IDB-TN-3003

Jel Codes: I21; I24

Keywords: Education; Learning; Learning assessments; Educational quality; Educational equity; Education indicators; Education policies; Large-scale assessments.

http://www.iadb.org

2024

THE STATE OF EDUCATION

in Latin America and the Caribbean

LEARNING ASSESSMENTS

Elena Arias Ortiz, Ximena Dueñas, Cecilia Giambruno y Angela López

IDB

THE STATE OF EDUCATION

in Latin America and the Caribbean

2024

The second edition of *The State of Education in Latin America and the Caribbean* 2024 examines the evolution of large-scale learning assessments in the region. The report begins with an overview of education based on a selection of indicators on resources, coverage, efficiency, and learning. It focuses specifically on socioeconomic inequalities. It then focuses on the evolution of learning measurement in 18 countries from the 1990s to this day, highlighting significant milestones in regional and international assessments and the design and implementation of large-scale national assessments. The report analyzes aspects such as test frequency and coverage, institutional changes, types of assessments and the impact of the pandemic, as well as the challenges affecting the quality and validity of the assessment. The report also examines how these assessments influence education policy making and their role in improving education quality. Finally, it provides recommendations for strengthening learning assessment systems, supporting post-pandemic learning recovery and promoting greater education equity.

CONTENTS

List of figures

List of tables

INTRODUCTION

Learning assessments is critical to improving the quality of education. Educators can identify gaps in understanding and areas that need improvement by assessing their students' learning. This data-driven approach enables targeted interventions and supports the continuous improvement of teaching practices and student performance. We can summarize the intention behind assessing with two statements attributed to the British physicist and mathematician William Thomson Kelvin: "What is not defined, cannot be measured. What is not measured, cannot be improved." Assessment is not the end: it is the means to focus on using results to improve and thus help education systems develop the skills and knowledge students need to succeed in the world of work, live in society, and lead fulfilling lives.

It is difficult to disagree with these ideas, but their truthfulness hinges on an implicit assumption: that we have appropriate tools to assess learning. We must know how education interventions, programs, or policies impact learning improvement. Therefore, we must ensure that each stakeholder—students, teachers, principals, supervisors, and policymakers–has the appropriate measurement tools to determine whether we are moving in the right direction. Robust national educational assessment systems are crucial to identifying the schools and students with the most significant challenges and directing investment to those most in need. In Latin America and the Caribbean, learning assessments establish this as the pillar of education policy: to improve learning and promote greater equity in learning. PISA 2022 (Programme for International Student Assessment) results underscore the learning crisis in Latin America and the Caribbean: 1 in 2 students (55%) lacks basic reading skills, and 3 in 4 (75%) lack basic mathematics skills. The situation is even more critical among the poorest students: 9 in 10 (88%) lack basic mathematics skills (Arias Ortiz et al., 2023a).

Large-scale assessments are essential in education systems worldwide. They are used by international organizations (OECD and UNESCO), national governments, states, and municipalities to measure and report learning achievements in an education system. Large-scale learning assessments measure the learning outcomes of a specific group of students, based on their age or course level, in a given academic year. These assessments are characterized by their uniformity and standardization in terms of content, administration process, dates, and scoring systems (UNESCO, 2019). Assessments of this kind, generally summative,[1] provide an overall summary of academic performance in specific areas and are implemented nationally and cross-nationally (regionally or internationally). Indeed, these assessments have some disadvantages as they measure a small part of what students learn in school, generally only mathematics, language and science, excluding other subject areas (Vegas and Petrow, 2008). Non-academic knowledge and behaviors, such as life skills, ethical and moral values, artistic and creative skills, and a sense of civic or social responsibility, are more difficult to quantify and often fall outside the scope of such assessments. Even "objective" knowledge, such as facts and basic reasoning skills, may vary between cultural or value systems and therefore be difficult to assess.

Despite these limitations, large-scale assessments have many advantages as a monitoring tool for objectivity and data generation for system-level decision making and education policy making in critical areas such as curriculum, resource allocation and pedagogical practices. Additionally, the comparability of results makes it possible to improve equity. Connecting context, resources, and per-

1 Summative assessments are a type of learning assessment administered at the end of a unit or sequential units of instruction to measure to what extent learners have achieved desired outcomes. (Bernard, 2009).

formance allows us to identify equity gaps in learning and to promote policies that encourage a more equitable distribution of educational opportunities. Unfortunately, the reality is more complex, as the diversity of learning experiences and contexts in which students learn—including differences in socioeconomic background, cultural influences, and school characteristics—poses a considerable challenge for developing universal assessment tools that fairly measure student progress.

This second issue of *The State of Education in Latin America and the Caribbean* presents a detailed analysis of the current status of large-scale learning assessments in the region, their role in education policy decision making, and the challenges and main opportunities they offer for improving education quality.

The first part of *The State of Education*, as in every edition, includes a comprehensive update of the educational landscape in the region based on indicators taken from the CIMA (Spanish acronym for Learning Improvement Information Center) education statistics portal. The status of each country's financial and physical resources, education coverage, internal efficiency and student learning achievements are analyzed. This edition especially focuses on inequalities between and within countries. We provide a detailed analysis of gaps by socioeconomic status (SES). The results show considerable investment differences per student compared to the OECD average. There are also significant differences in access to digital resources among the countries in the region, which especially impacts the most socioeconomically disadvantaged students. Secondary and tertiary education coverage remains limited, and the low efficiency of the systems is reflected in high dropout and low completion rates, particularly in the most vulnerable sectors. Finally, PISA 2022 data show persistent gaps in learning equity, both within countries and compared to the OECD.

The second part of the report focuses on the evolution and current state of educational assessment systems in Latin America and the Caribbean. First, we analyze the participation of countries in regional and international assessments, their relevance to national education policies and their role in monitoring SDG4's targets. We then analyze the major stages of the evolution of large-scale national learning assessments[2] with an innovative methodology that includes a quantitative analysis of the assessments, education levels and their scope from 1990 to 2023. We primarily focus on eighteen Latin American countries (Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Uruguay, and Venezuela). We analyze the coverage of the target population (sample or census), the areas and levels evaluated, the psychometric model, and the assessment frequency. We also discuss the impact of COVID-19 on the continuity of assessments, considering the characteristics of assessment systems and types of assessments in countries. Finally, the last section summarizes the main findings and discusses lessons learned and challenges toward a true assessment culture.

The results show a diverse evolution but with common challenges. First, monitoring learning remains essential in education policies, and most countries have national learning assessments for this purpose. However, the level of maturity of assessment systems varies considerably from country to country. International assessments, such as ERCE and PISA, are crucial to strengthening technical capacity and consolidating assessment systems. However, the need for more centers specialized in learning assessments and the scarcity of experts in psychometrics limit the development of solid assessment systems in the region. More validation and transparency regarding the quality and validity of the assessments are needed, which are critical for building trust in the assessment systems.

2 The analysis focused on large-scale summative national learning assessments administered between 1990 and 2023 in primary and secondary school.

Second, delivering and using results efficiently remains a significant challenge, as does the balance between different types of assessments, which, as we said, is still very much focused on large-scale assessments. Third, national assessments require constant funding, but reduced resources threaten their continuity. This is the case, particularly, of innovation in terms of methodologies and areas assessed, such as measuring socioemotional elements and using technology in assessments. These areas are growing, although the limited infrastructure hinders their large-scale implementation. Despite the challenges, it is essential to resume the work and strengthen assessment systems in the post-pandemic context to help learning in the region recover.

Strengthening assessment systems is not limited to large-scale summative learning assessments; it also requires integrating other types of measurements that go beyond the focus of this report. These include formative assessments, designed to provide continuous feedback during the learning process and adapt teaching to the students' needs[3], diagnostic assessments, which are formative approaches and allow us to assess knowledge and skills before the school year begins to identify areas of need or potential for enrichment; and self-assessments, which encourage student reflection. Each type of assessment has a specific purpose, and when taken together, they provide the feedback needed to improve teaching practices continually. As mentioned at the beginning, the ultimate goal is to improve education to increase opportunities for everyone in the region.

3 While summative assessment focuses on the end of an instruction stage, formative assessment takes place during the entire stage, focusing on the process and allowing continuous iterations between student and teacher, which provides personalized and detailed information to adapt learning and teaching strategies (UNESCO, 2021).

PART I.

An overview of Education in Latin America and the Caribbean

An education system can be conceptualized as a process in which financial and physical inputs are provided to facilitate access, retention and learning of students in a given country[4]. Information is necessary to better understand these processes. In 2017, the IDB created **CIMA** (Spanish acronym for Learning Improvement Information Center) to improve the availability of comparable data and indicators in the region. CIMA provides data that measures each stage of the educational process and enables the disaggregation of information by dimensions such as SES, gender, and geographical location.

This chapter presents a detailed analysis of education in Latin America and the Caribbean through several indicators, focusing on the persistent SES inequality in the region. First, we examine the financial inputs available using UNESCO data[5]. We then assess the access to physical resources, focusing on access to and use of digital resources based on PISA 2022 data. The third section analyzes education coverage in the region using net attendance rates as obtained from national household surveys (2021 to 2023 data). We then evaluate the efficiency of the education system using dropout and completion rates also taken from national household surveys. Finally, we measure learning by analyzing low performance and lag in years of education compared to the OECD average, also using PISA data.lizando nuevamente datos de la prueba PISA.

I.1 Financial resources

We analyze the financial resources allocated to education by using government expenditure on education as a percentage of GDP and initial government funding per student in constant purchasing power parity (PPP) dollars by education level.

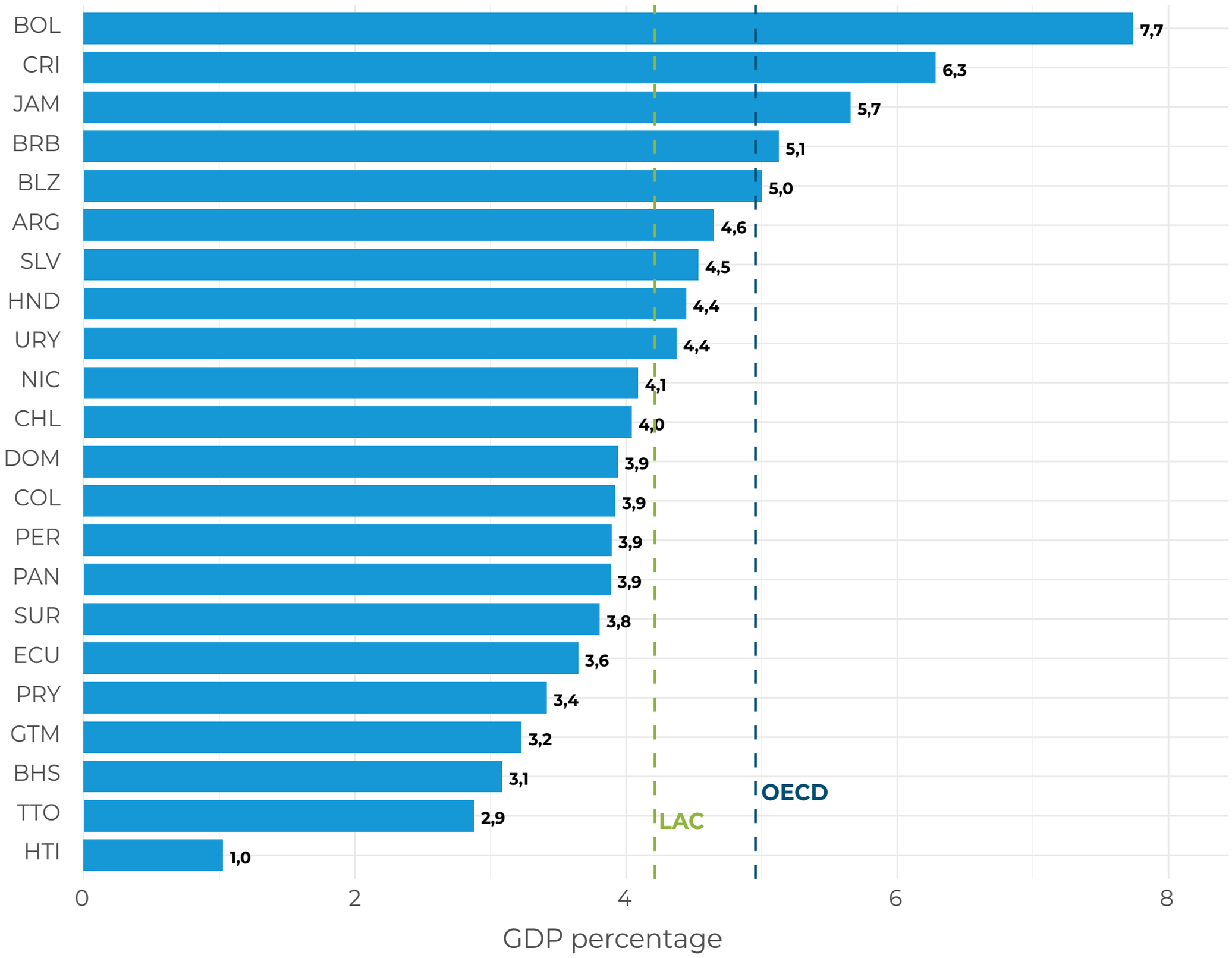
Public spending on education as a percentage of GDP reflects a government's emphasis on education relative to its national economic wealth (UNESCO Institute for Statistics, 2022). Comparing initial government expenditure per student in USD PPP allows us to examine the investment per student at a specific education level. This indicator complements the public spending on education as a percentage of GDP metric, as a higher percentage does not necessarily entail more resources available per student (Arias Ortiz, et al., 2023c).

In Latin America and the Caribbean, average spending on education equals 4.2% of GDP, a level just below the OECD average (5%). However, half the countries in the region need to reach the level recommended by UNESCO, which suggests allocating at least 4 to 6 % of GDP to education (UNESCO Institute for Statistics, 2022). Indeed, there is considerable disparity between the countries of Latin America and the Caribbean. The countries that allocate the most public spending to education as a percentage of GDP are Bolivia (7.7%), Costa Rica (6.3%) and Jamaica (5.7%); while the countries that allocate the least are Haiti (1%), Trinidad and Tobago (2.9%) and the Bahamas (3.1%) (Figure 1).

4 This chapter benefited from the valuable contribution of Melchor de la Cruz, Education Division intern at the IDB 2024 Summer Internship Program.

5 The Financial Resources data presented in this document were taken from the February 2024 update prepared by the UNESCO Institute for Statistics (UIS). Indicators from both the "SDG Global and Thematic Indicators" collection and the "Other Policy Relevant Indicators" collection are included.

Figure 1: Government expenditure on education as a percentage of GDP (circa 2023)[6]



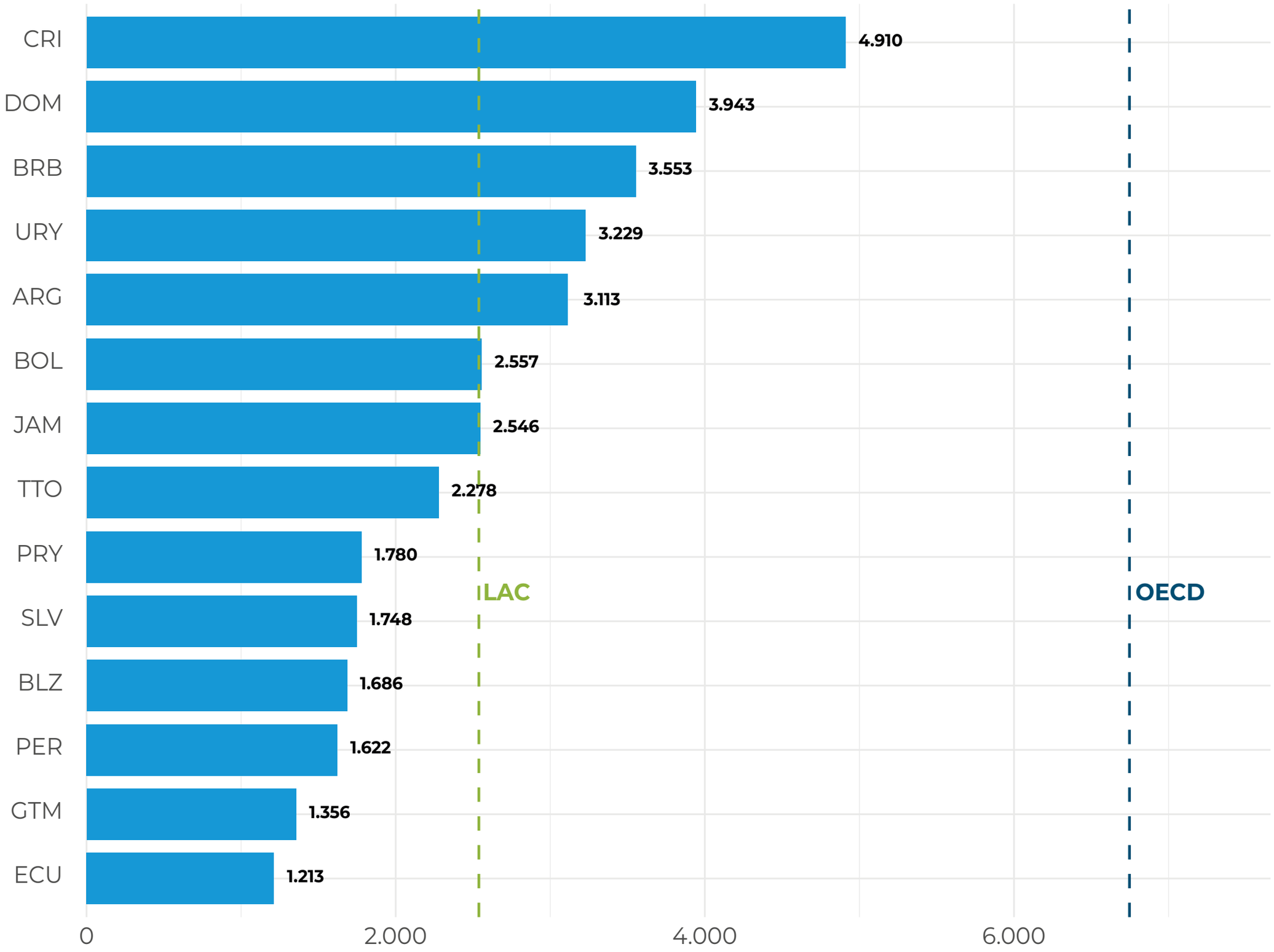
Source: Own elaboration based on CIMA.

However, primary education shows a significant gap in USD PPP in public investment per student between the region's average and the OECD average. The average OECD investment per primary student is 6,700 USD PPP, while only 2,500 USD PPP per student in Latin America and the Caribbean: 63% less.

All countries in the region show lower investment per primary student than the OECD average. There is also a significant disparity in the region. The highest initial primary spending is found in Costa Rica (4,910 USD PPP), the Dominican Republic (3,943 USD PPP) and Barbados (3,553 USD PPP). In contrast, Ecuador (1,213 USD PPP), Guatemala (1,353 USD PPP) and Peru (1,622 USD PPP) show the lowest spending (Figure 2).

6 Data for Chile (CHL), Nicaragua (NIC), Suriname (SUR), Colombia (COL), Costa Rica (CRI), and Argentina (ARG) are for 2021. Data for Haiti (HTI), Bolivia (BOL), Ecuador (ECU), Peru (PER), Dominican Republic (DOM), Panama (PAN), Paraguay (PRY), El Salvador (SLV), Trinidad and Tobago (TTO), Honduras (HND), Guatemala (GTM), and Uruguay (URY) are for 2022. The value for Latin America and the Caribbean (LAC) is calculated as a simple average of country data.

Figure 2: Initial government funding per primary school student in USD PPP (circa 2022)[7]

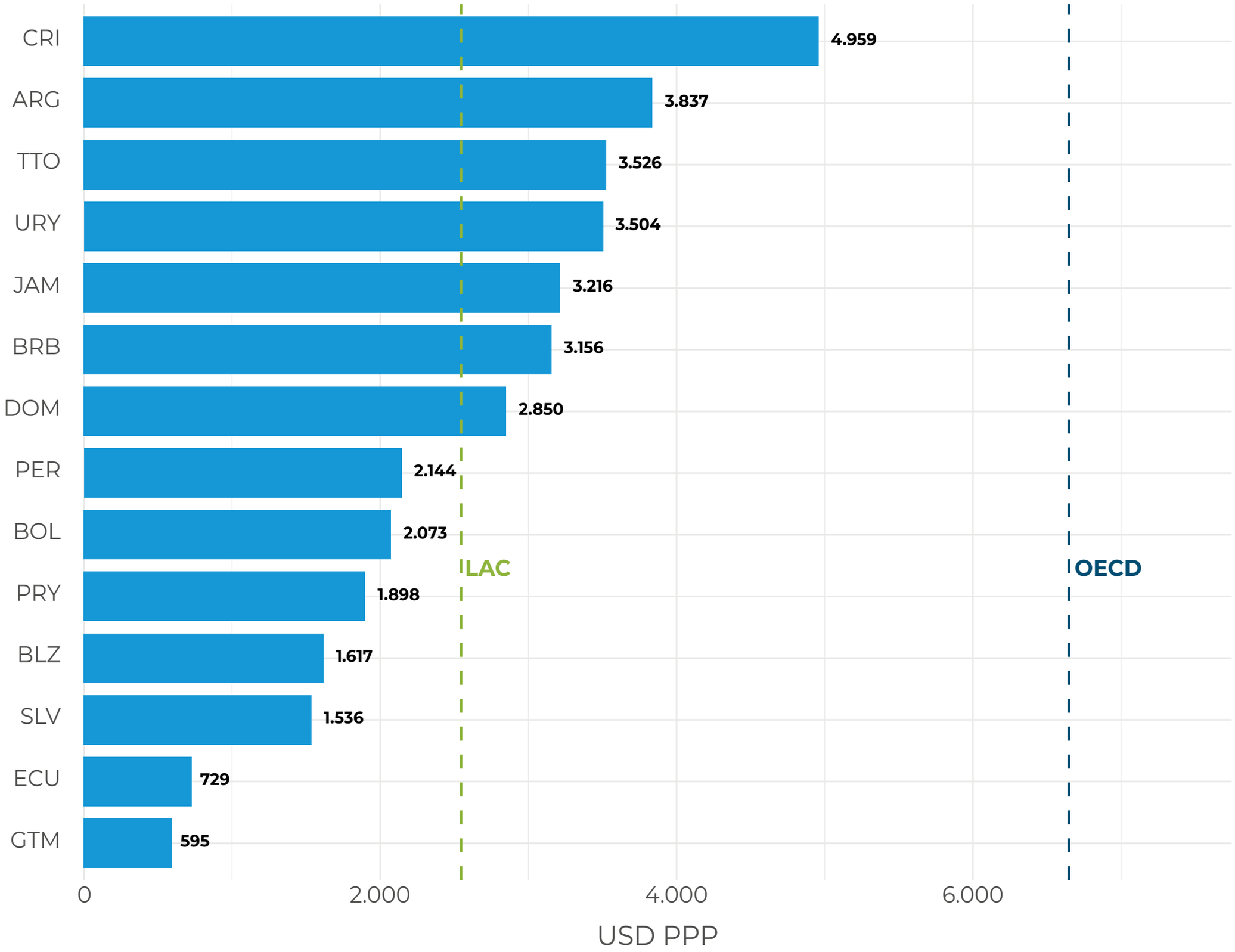


Source: Own elaboration based on CIMA.

The average public investment in secondary education in LAC is 2,600 USD PPP, 62% less than the OECD average. There is also considerable disparity between countries in the region. The highest initial secondary spending is found in Costa Rica (4,959 USD PPP), Argentina (3,837 USD PPP), and Trinidad and Tobago (3,526 USD PPP). In contrast, investment per student is lower in Guatemala (595 USD PPP), Ecuador (729 USD PPP), and El Salvador (1,536 USD PPP) (Figure 3).

7 Data for Trinidad and Tobago (TTO), El Salvador (SLV), Costa Rica (CRI), Uruguay (URY), Argentina (ARG), and the OECD are for 2021. Data for Bolivia (BOL), Ecuador (ECU), Dominican Republic (DOM), Paraguay (PRY), Guatemala (GTM), Belize (BLZ), Peru (PER), Jamaica (JAM), and Barbados (BRB) are for 2022. The value for Latin America and the Caribbean (LAC) is calculated as a simple average of country data.

Figure 3: Initial government expenditure per secondary school student in USD PPP (circa 2022)[8]



Source: Own elaboration based on CIMA.

I.2 Physical resources: access to and use of technology

Digital transformation in education systems can expand access to educational opportunities, promote inclusion, and improve the relevance and quality of learning. Full and equitable access to digital infrastructure is essential for this transformation to be feasible. The PISA 2022 tests are a valuable foundation for analyzing access to digital devices and connectivity in educational institutions.[9]

In addition to access, the forms provided to students allow us to examine the use of digital devices in educational centers in Latin America and the Caribbean[10].

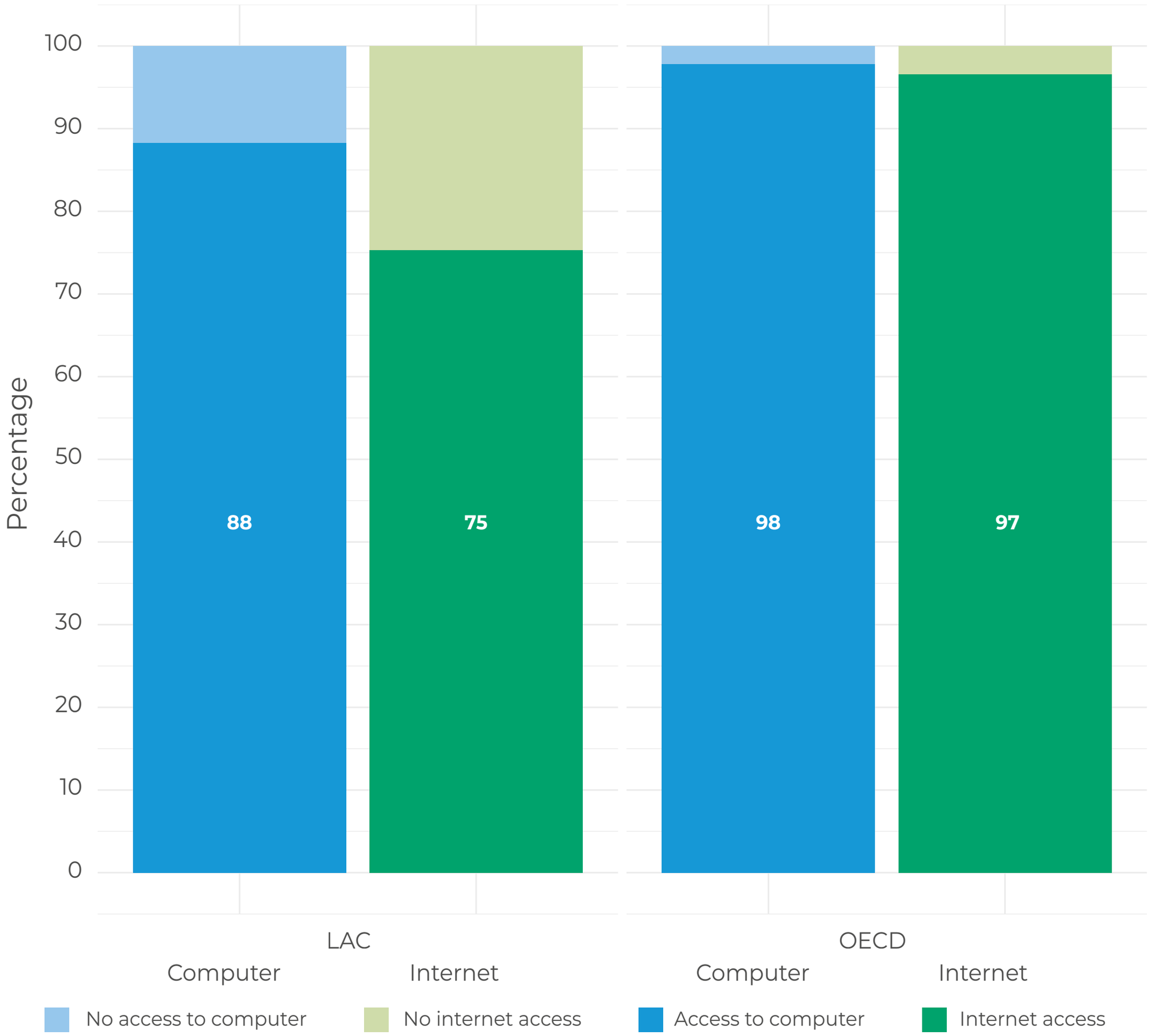
Combining these indicators helps build a fuller picture, which is essential for developing effective regional digital transformation strategies.

8 Data for Trinidad and Tobago (TTO), El Salvador (SLV), Costa Rica (CRI), Uruguay (URY), Argentina (ARG), and the OECD are for 2021. Data for Bolivia (BOL), Ecuador (ECU), Dominican Republic (DOM), Paraguay (PRY), Guatemala (GTM), Belize (BLZ), Peru (PER), Jamaica (JAM), and Barbados (BRB) are for 2022. The value for Latin America and the Caribbean (LAC) is calculated as a simple average of country data.

9 A student is considered to have access to a computer at school if, in the principals' survey, it is indicated that the school has at least one computer available for student use. Similarly, a student is considered to have access to the internet at school if, in the same survey, it is stated that the school has at least one computer with internet access available for student use

10 A subject is considered to use digital devices when the student reports using them in at least half the classes.

Figure 4: Access to digital resources in LAC and the OECD (2022)

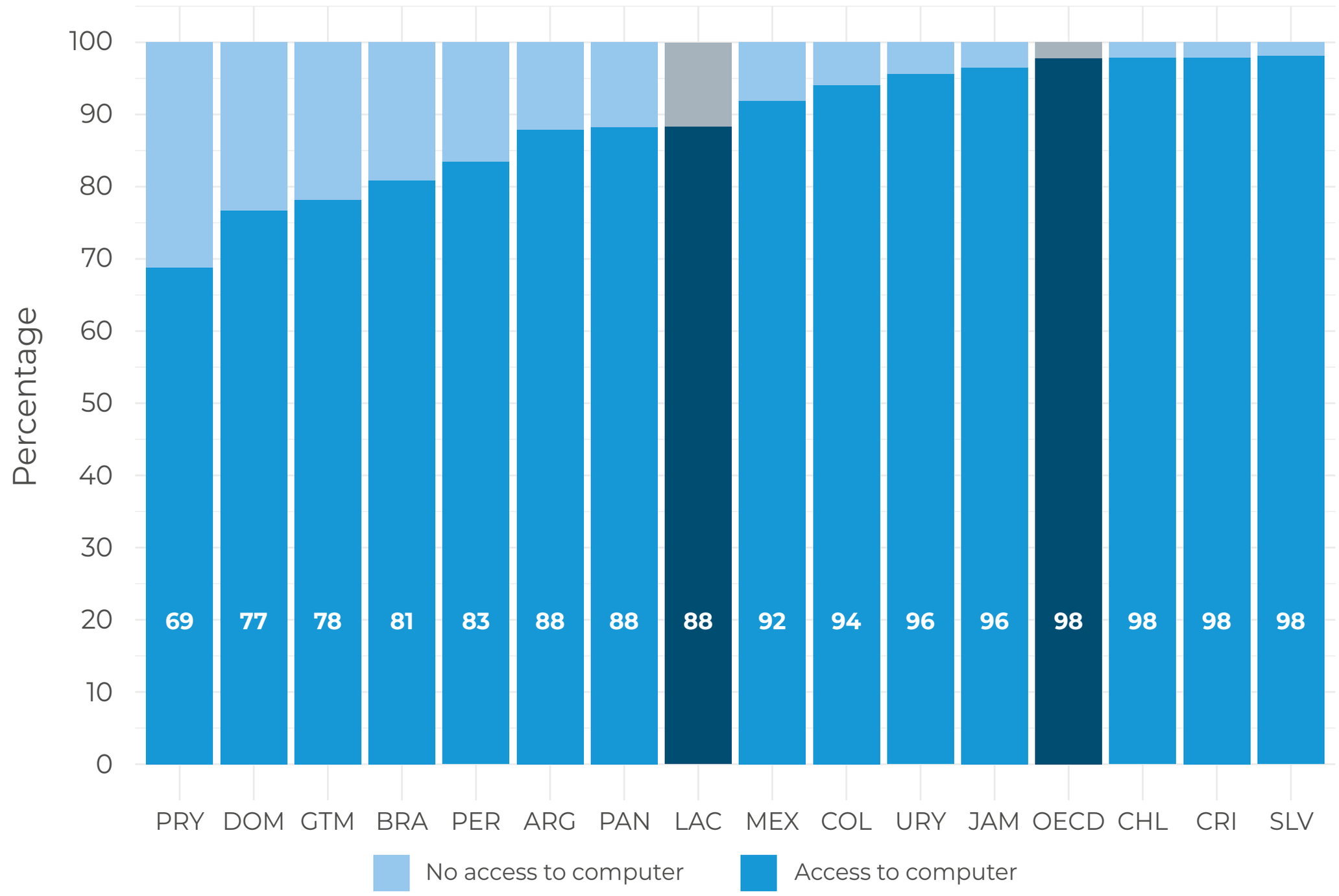


Source: Own elaboration based on CIMA

In Latin America and the Caribbean, 1 in 10 students does not have access to a computer at school. Similarly, 2 in 10 students do not have internet access at those institutions. This contrasts with the OECD average, which reports almost universal access to computers and the internet (Figure 4).

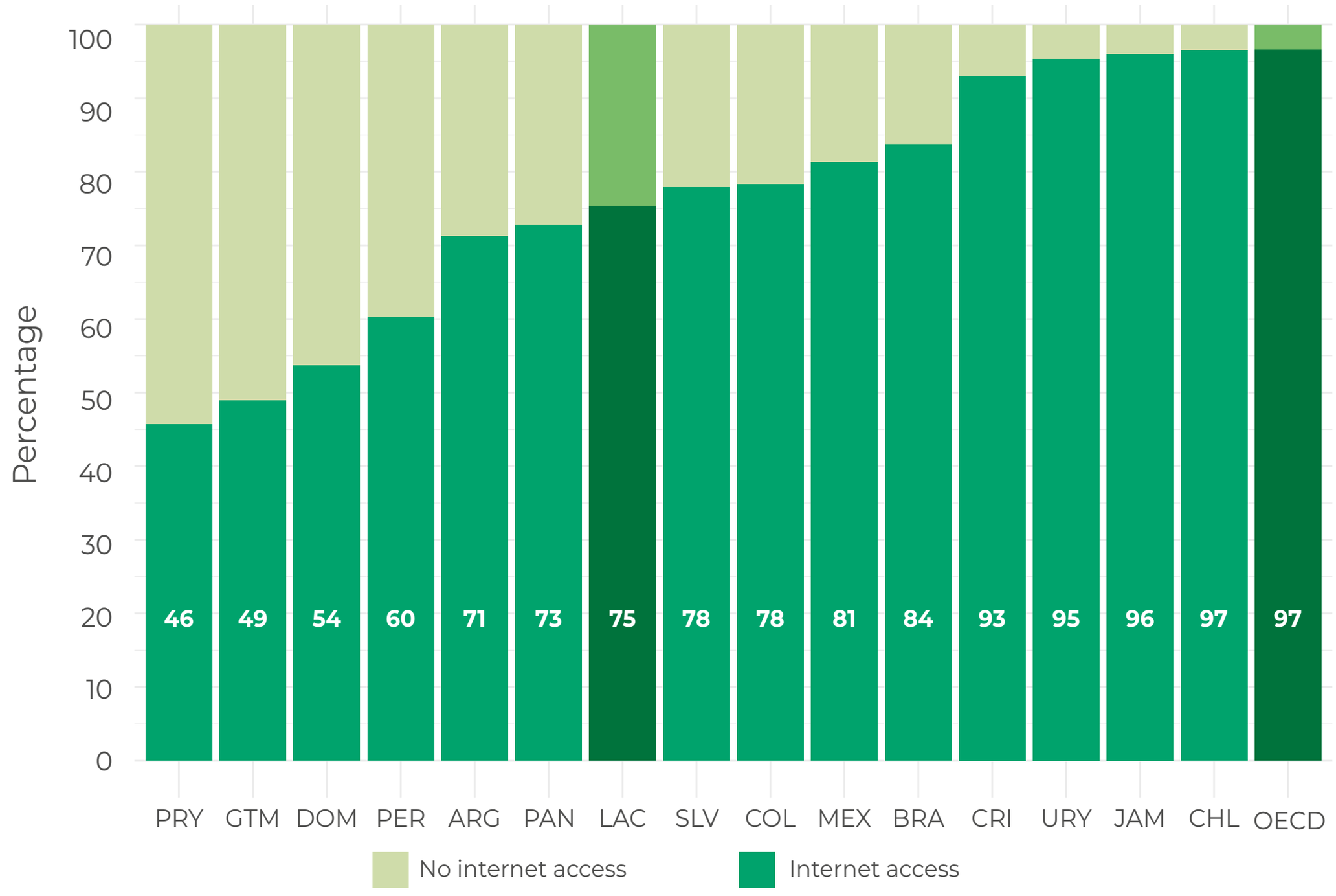
The countries in the region show a considerable disparity in access to the internet and computers. Almost all students have internet access in Chile, Jamaica, and Uruguay, while over half lack internet access in Paraguay and Guatemala. This reflects a significant inequity between countries with more and less access to this digital infrastructure in the region. Similarly, in El Salvador, Costa Rica, and Chile, almost all students have access to computers, while 3 in 10 students in Paraguay do not; in the Dominican Republic and Guatemala, 2 in 10 students lack computer access (Figures 5 and 6).

Figure 5: Access to computers by country (2022)



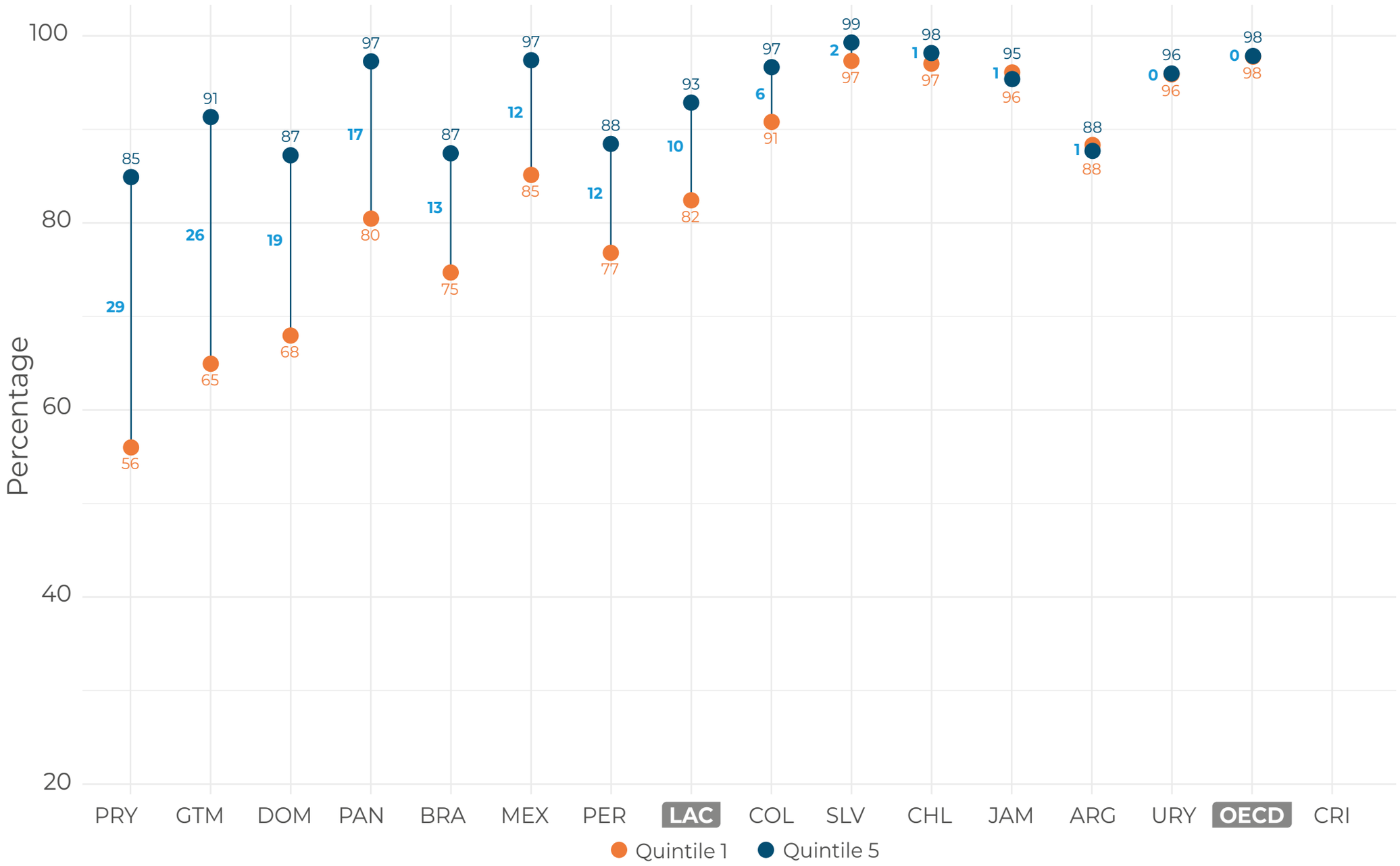
Source: Own elaboration based on CIMA.

Figure 6: Internet access by country (2022)



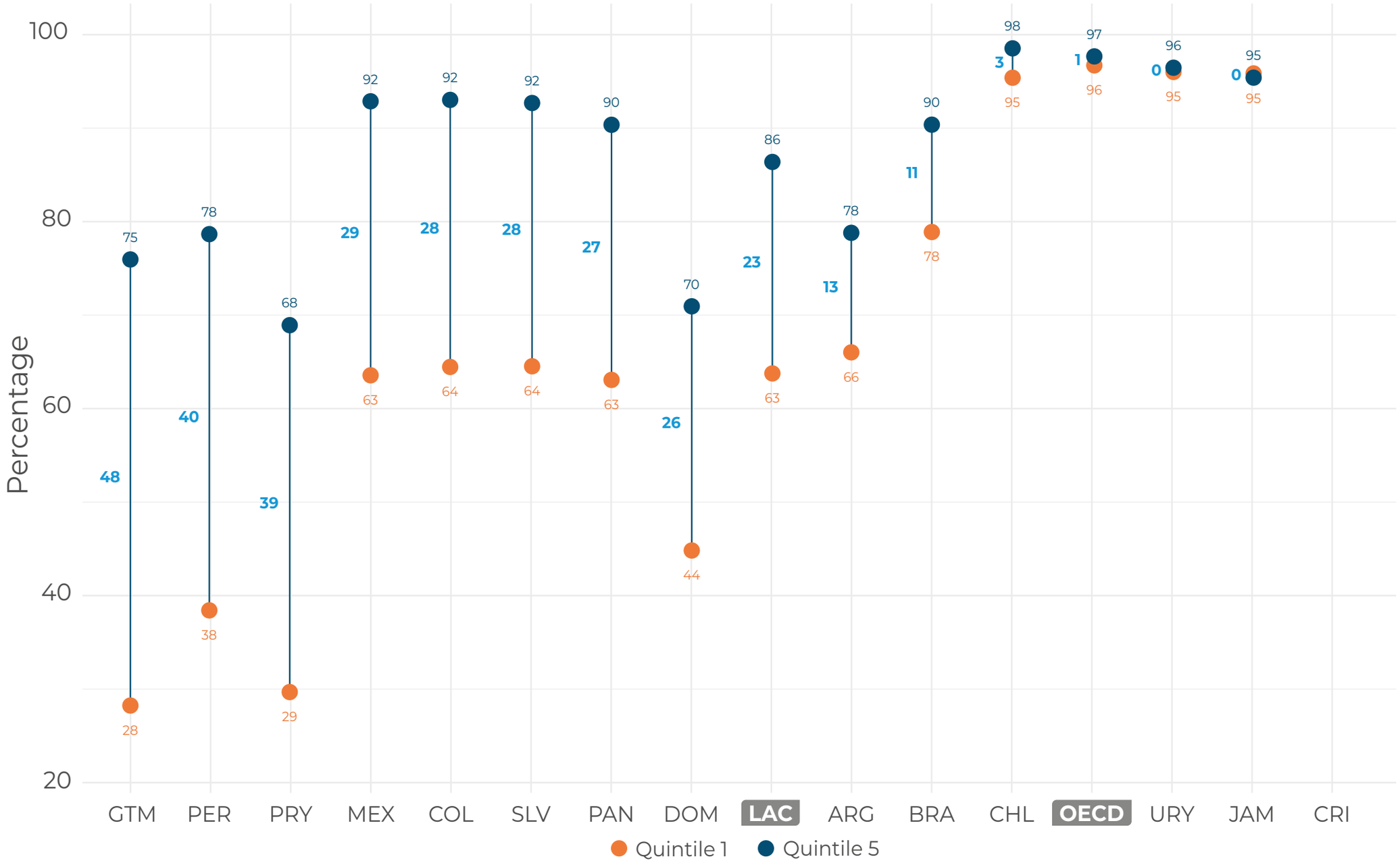
Source: Own elaboration based on CIMA.

Figure 7: Access to computers at school by socioeconomic status (2022)



Source: Own elaboration based on CIMA.

Figure 8: Internet access at school by socioeconomic status (2022)



Source: Own elaboration based on CIMA.

In addition to the difference in access, LAC countries have a significant inequality in access to digital infrastructure by SES. In general, countries with lower average access to the internet and computers have a larger gap by SES.

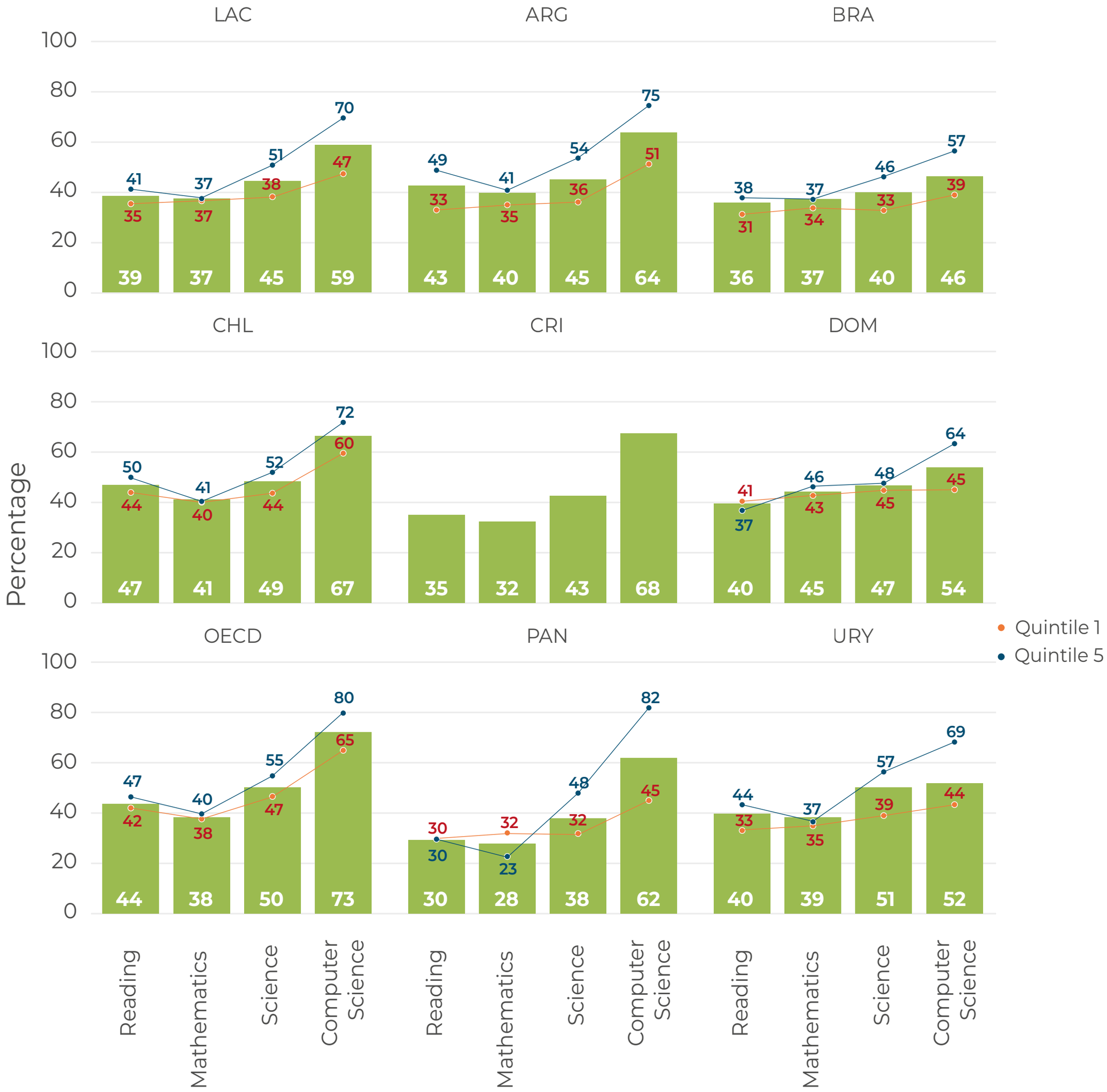
In Guatemala, 2 in 10 students in the highest SES category do not have internet access, compared to 7 in 10 in the lowest SES category. In Peru, 2 in 10 students in the highest SES category lack internet access, compared to 6 in 10 from the lowest SES category. This contrasts with full internet access, regardless of SES, in other countries in the region, like Chile and Jamaica (Figure 8).

A gap by SES is observed in access to computers. In Paraguay, 1 in 10 students in the highest SES category does not have access to computers at school, compared to 4 in 10 students in the lowest SES category. In Guatemala, 1 in 10 students in the highest SES category lack access to computers, while 3 in 10 in the lowest level lack access (Figure 7). Poor access to digital infrastructure, especially for socioeconomically disadvantaged students, has been previously noted (Arias Ortiz et al., 2024).

PISA 2022 also allows us to analyze the use of digital devices by subject, as reported by students. Figure 9 shows the percentage of students who report using digital devices in at least half the classes of the subject analyzed. The region's distribution is similar to the OECD average: the highest use of devices is observed in computer science (59%), then science (45%), reading (39%) and finally, mathematics (37%). The Dominican Republic is the only country that does not follow this pattern, showing a higher use of devices in computer science (54%), followed by science (47%), mathematics (45%), and reading (40%) (Figure 9).

For the regional average, in absolute terms, computer science has the most significant gap between the highest and lowest SES categories (23 percentage points, p. p.), while the smallest gap appears in mathematics (0 p. p.). However, no clear associations are observed between the use of devices and learning outcomes, highlighting the importance of an effective implementation strategy for such devices to improve educational outcomes. Many countries in the region lack appropriate governance structures and guidelines at educational institutions to ensure the efficient use of digital resources for learning and teaching (Arias Ortiz et al., 2024). These shortcomings underscore the need to develop and implement policies and guidelines for school teachers and administrative staff to enhance the use of technology in education.

Figure 9: Percentage of students reporting using digital devices in at least half of the classes, by subject (2022)



Source: Own elaboration based on CIMA.

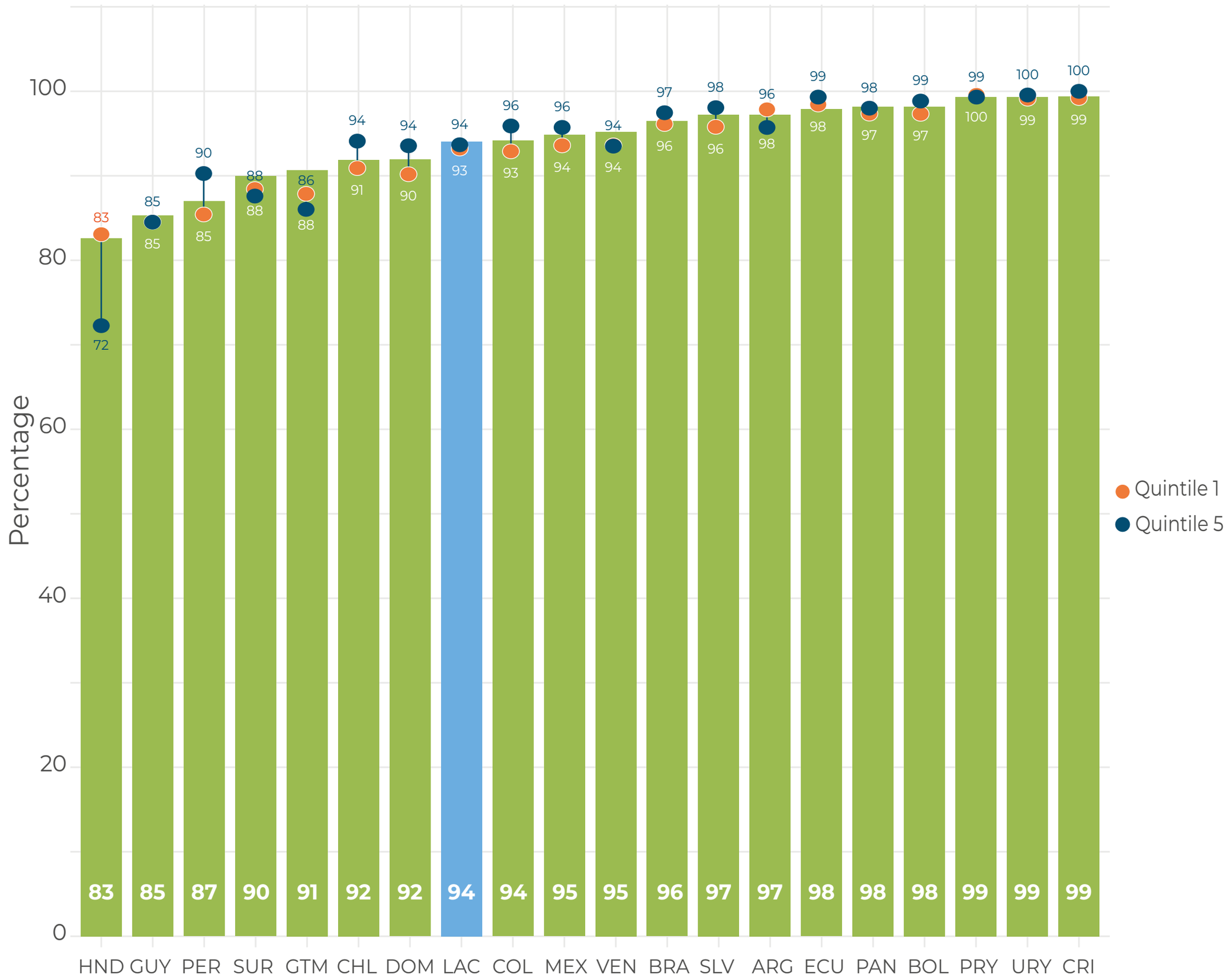
Note: There is no available data on socioeconomic status for Costa Rica in PISA 2022.

I.3 Coverage

The net attendance rate indicator measures the percentage of people effectively attending an education level over the population that should be attending according to their age. Most countries in the region have achieved universal primary coverage. However, deficiencies persist in secondary and tertiary education, especially when coverage is disaggregated by SES.

Although most LAC countries have achieved universal primary coverage, some still need to improve. In Honduras, 2 in 10 people who should be in primary school do not attend. This deficiency is more significant in the lower SES categories, where 3 in 10 people do not attend. In Guyana and Peru, 1 in 10 people of primary school age do not attend primary school. However, the gap in primary school attendance according to SES is small or practically non existent in most countries in the region (Figure 10).

Figure 10: Gap in net primary attendance rate by socioeconomic status (circa 2023)[11]

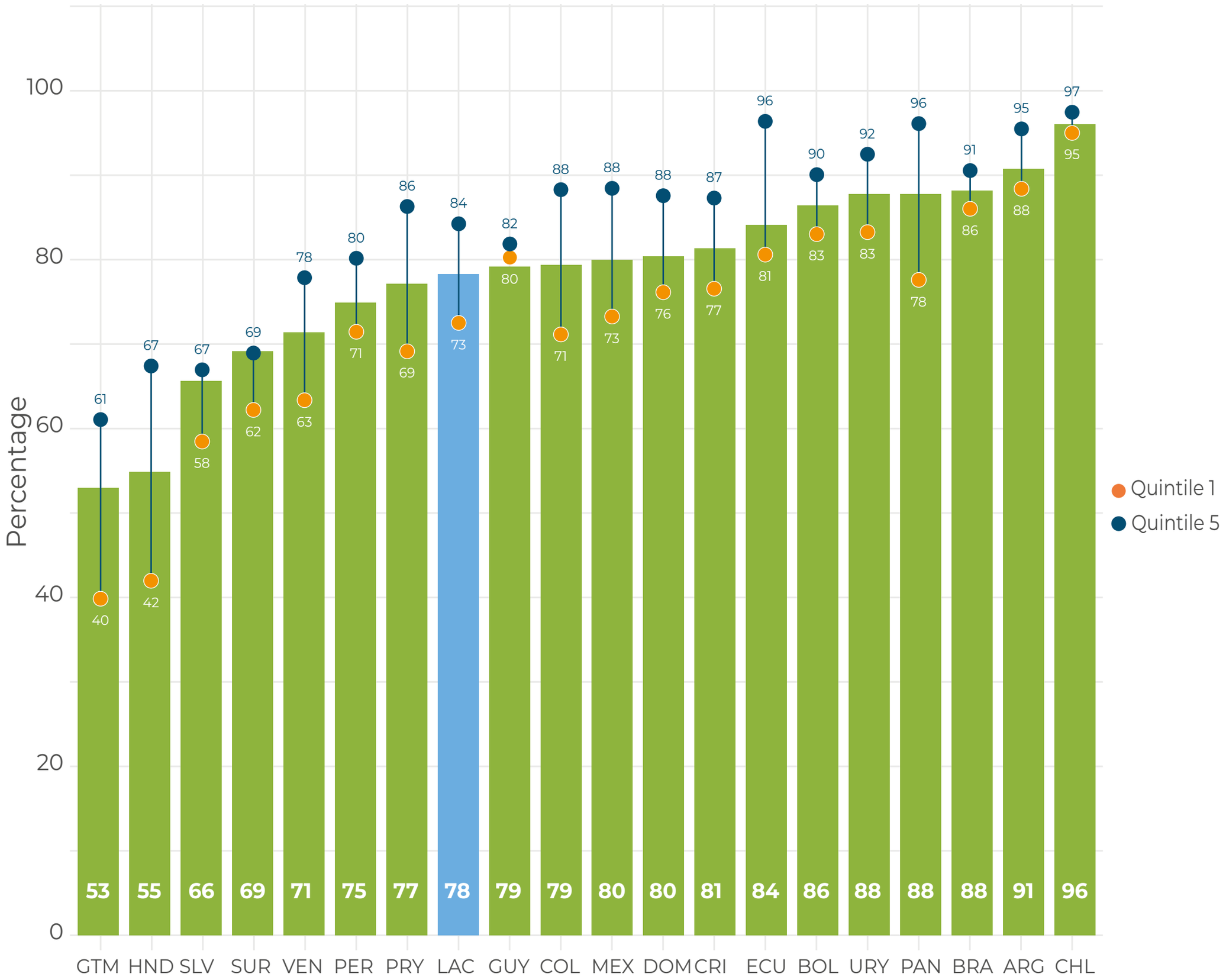


Source: Own elaboration based on CIMA.

In the region, 8 in 10 people of secondary school age attend secondary school. Unlike primary school attendance, gaps in secondary school net attendance persist between people from the highest and lowest SES categories (11 p.p.). The countries with the greatest shortcomings in secondary coverage are Guatemala, where 5 in 10 people attend, and Honduras, with 6 in 10. This increases when analyzing the socioeconomic gap, since only 4 in 10 people in the lowest SES category of secondary school age attend in both countries. This amounts to a 60% difference compared to universal coverage (Figure 11).

11 Data for Bolivia (BOL), Dominican Republic (DOM), Guyana (GUY) and Venezuela (VEN) are for 2021; data for Chile (CHL), Argentina (ARG), Brazil (BRA), Panama (PAN), Uruguay (URY), Costa Rica (CRI), Mexico (MEX), Colombia (COL), Paraguay (PRY), Peru (PER), Suriname (SUR), and Guatemala (GTM) are for 2022; data for Ecuador (ECU), El Salvador (SLV), and Honduras (HND) are for 2023. The value for Latin America and the Caribbean (LAC) is calculated as a simple average of country data.

Figure 11: Gap in net secondary attendance rate by socioeconomic status (circa 2023)[12]



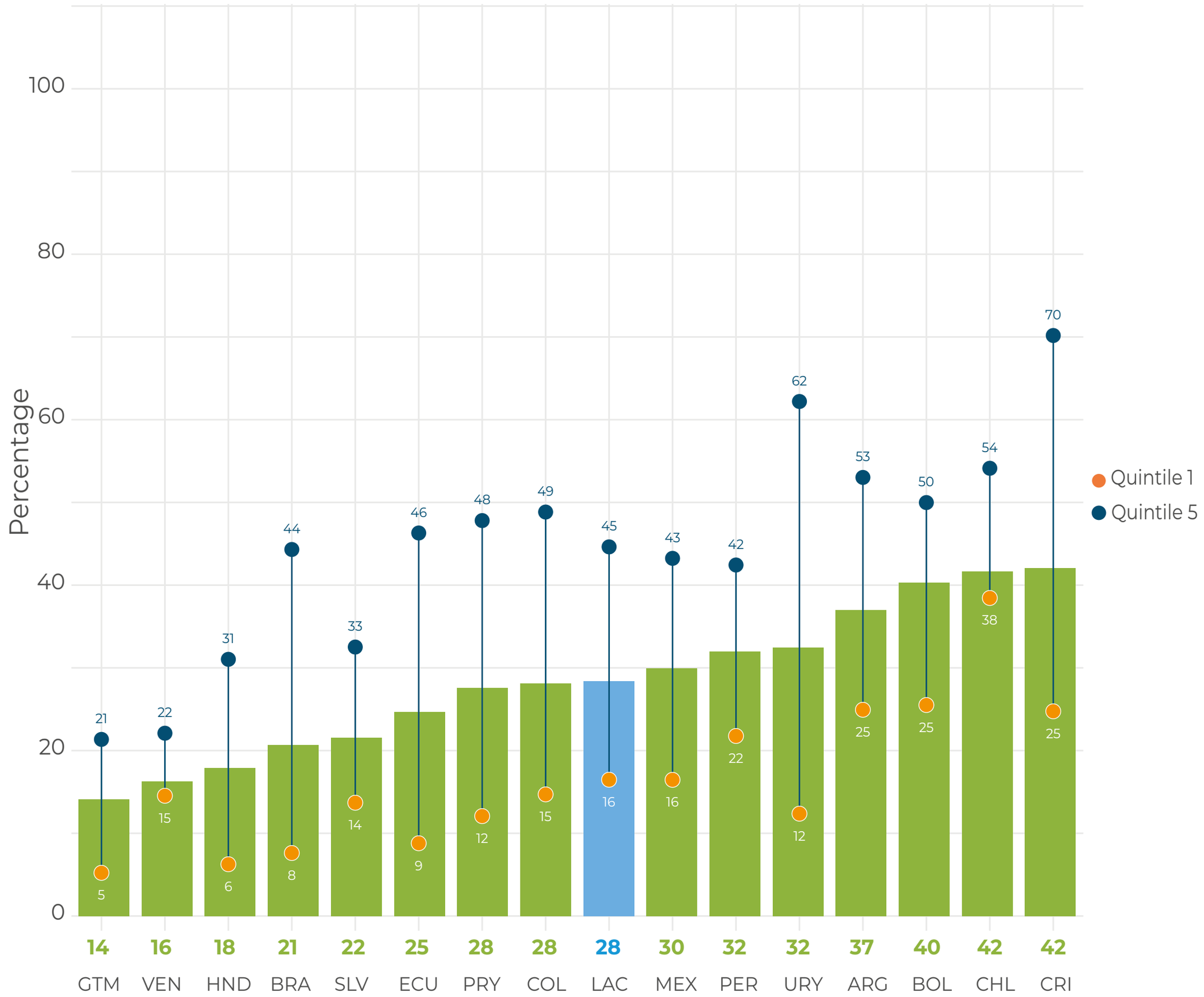
Source: Own elaboration based on CIMA.

Tertiary education coverage is significantly lower than at primary and secondary levels. On average, only 3 in 10 people of tertiary education age attend, and coverage is less than 50% in all countries. The countries with the highest coverage are Costa Rica, Chile, and Bolivia, where 4 in 10 people attend. In contrast, the countries with the lowest tertiary education coverage are Guatemala, with 1 in 10, and Venezuela and Honduras, with 2 in 10 people of tertiary education age attending school.

The gap by SES in tertiary education attendance is greater than primary and secondary education rates. The countries with the largest disparities in tertiary education coverage between the highest and lowest SES categories are Uruguay (50 p.p.), Costa Rica (45 p.p.), and Ecuador (37 p.p.) (Figure 12).

12 Data for Bolivia (BOL), Dominican Republic (DOM), Guyana (GUY) and Venezuela (VEN) are for 2021; data for Chile (CHL), Argentina (ARG), Brazil (BRA), Panama (PAN), Uruguay (URY), Costa Rica (CRI), Mexico (MEX), Colombia (COL), Paraguay (PRY), Peru (PER), Suriname (SUR), and Guatemala (GTM) are for 2022; data for Ecuador (ECU), El Salvador (SLV), and Honduras (HND) are for 2023. The value for Latin America and the Caribbean (LAC) is calculated as a simple average of country data.

Figure 12: Gap in net tertiary attendance rate by socioeconomic status (circa 2023)[13]



Source: Own elaboration based on CIMA.

I.4 Efficiency

The success of an education system can be measured in terms of how efficiently students go through the educational levels until completion and the level of learning achieved. The indicators commonly used to measure efficiency are early dropout rate and completion rate by education level. These indicators allow us to analyze the percentage of students entering the education system and completing their education.

In Latin America and the Caribbean, 27% of students drop out early[14]. The countries with the lowest dropout rates are Chile (7%), Peru (10%) and Bolivia (16%), while the highest dropout rates appear in Guatemala (57%), Honduras (53%), and Uruguay (37%).

13 Data for Bolivia (BOL) and Venezuela (VEN) are for 2021; data for Chile (CHL), Argentina (ARG), Brazil (BRA), Uruguay (URY), Costa Rica (CRI), Mexico (MEX), Colombia (COL), Peru (PER), Paraguay (PRY), and Guatemala (GTM) are for 2022; data for Ecuador (ECU), El Salvador (SLV), and Honduras (HND) are for 2023. The value for Latin America and the Caribbean (LAC) is calculated as a simple average of country data.

14 Early dropout is the percentage of 18–24-year-olds who have not completed secondary education and are not attending any education level.

Significant inequalities are observed in school dropout rates, according to SES. Indeed, in the region, 14% of students in the highest SES category drop out of school early, compared to 39% of their peers in the lowest SES category. The highest dropout rates for students in the lowest SES category appear in Guatemala (80%), Honduras (76%), and Uruguay (60%). These countries also show the most significant gaps in dropout rates between the highest and lowest SES categories (Figure 13).

Figure 13: Early dropout rates by socioeconomic status (circa 2023)[15]

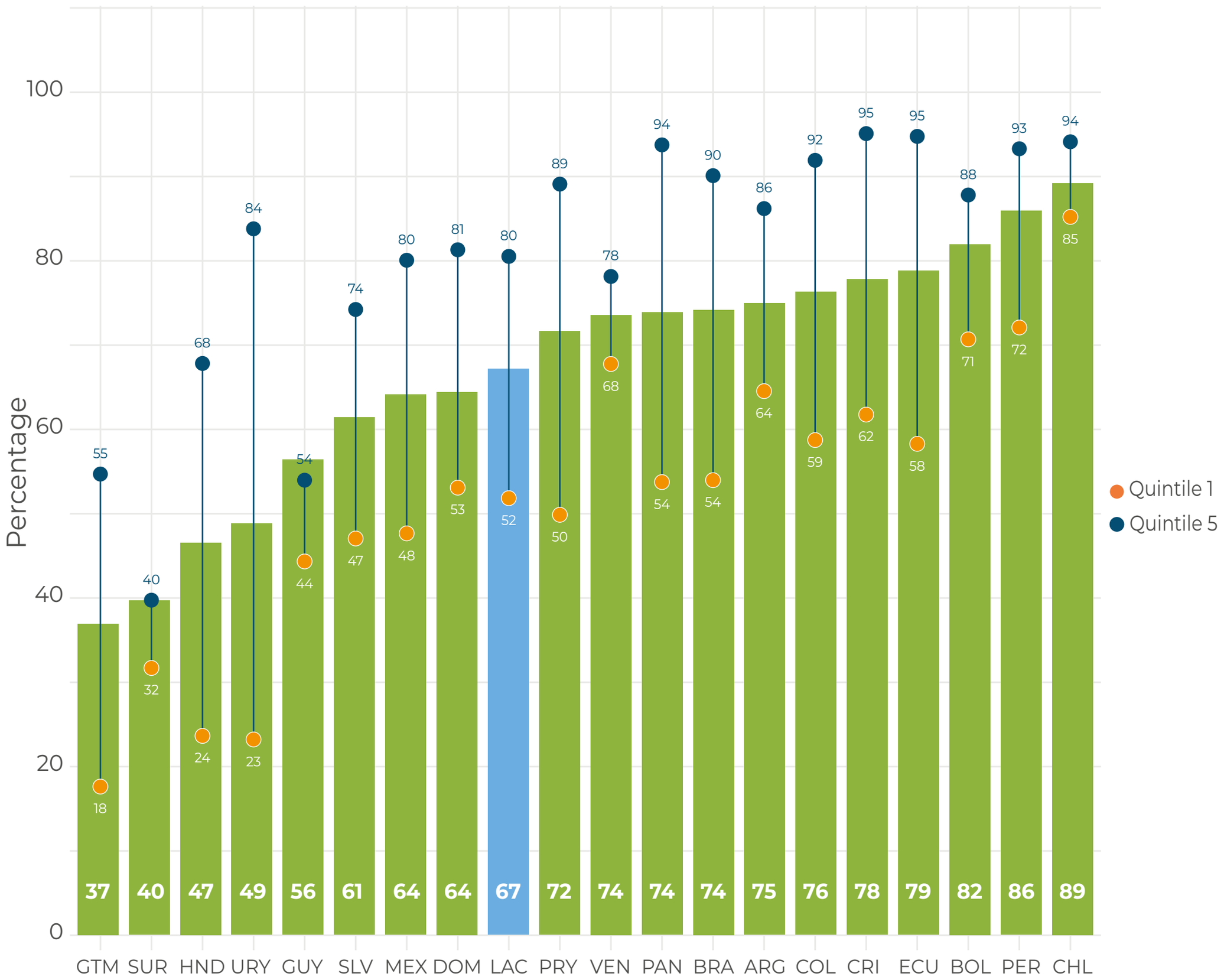
Source: Own elaboration based on CIMA.

15 Data for Bolivia (BOL), Dominican Republic (DOM), Guyana (GUY) and Venezuela (VEN) are for 2021; data for Chile (CHL), Argentina (ARG), Brazil (BRA), Panama (PAN), Uruguay (URY), Costa Rica (CRI), Mexico (MEX), Colombia (COL), Paraguay (PRY), Peru (PER), Suriname (SUR), and Guatemala (GTM) are for 2022; data for Ecuador (ECU), El Salvador (SLV), and Honduras (HND) are for 2023. The value for Latin America and the Caribbean (LAC) is calculated as a simple average of country data.

In the region, 67% of young people complete secondary education[16]. The most successful countries in this regard are Chile (89%), Peru (86%), and Bolivia (83%). In contrast, the countries with the lowest efficiency are Guatemala (37%), Suriname (40%), and Honduras (47%).

The analysis by SES highlights the region's shortcomings in terms of secondary school completion. The lowest completion rates for the lowest SES categories are observed in Guatemala (18%), Uruguay (23%), and Honduras (25%). This implies that, in these countries, over 75% of a cohort of young people from the lowest SES category aged 3 to 5 years older than the theoretical age for the last grade of secondary education have not completed this level (Figure 14).

Figure 14: Secondary education completion rate by socioeconomic status (circa 2023)[17]



Source: Own elaboration based on CIMA.

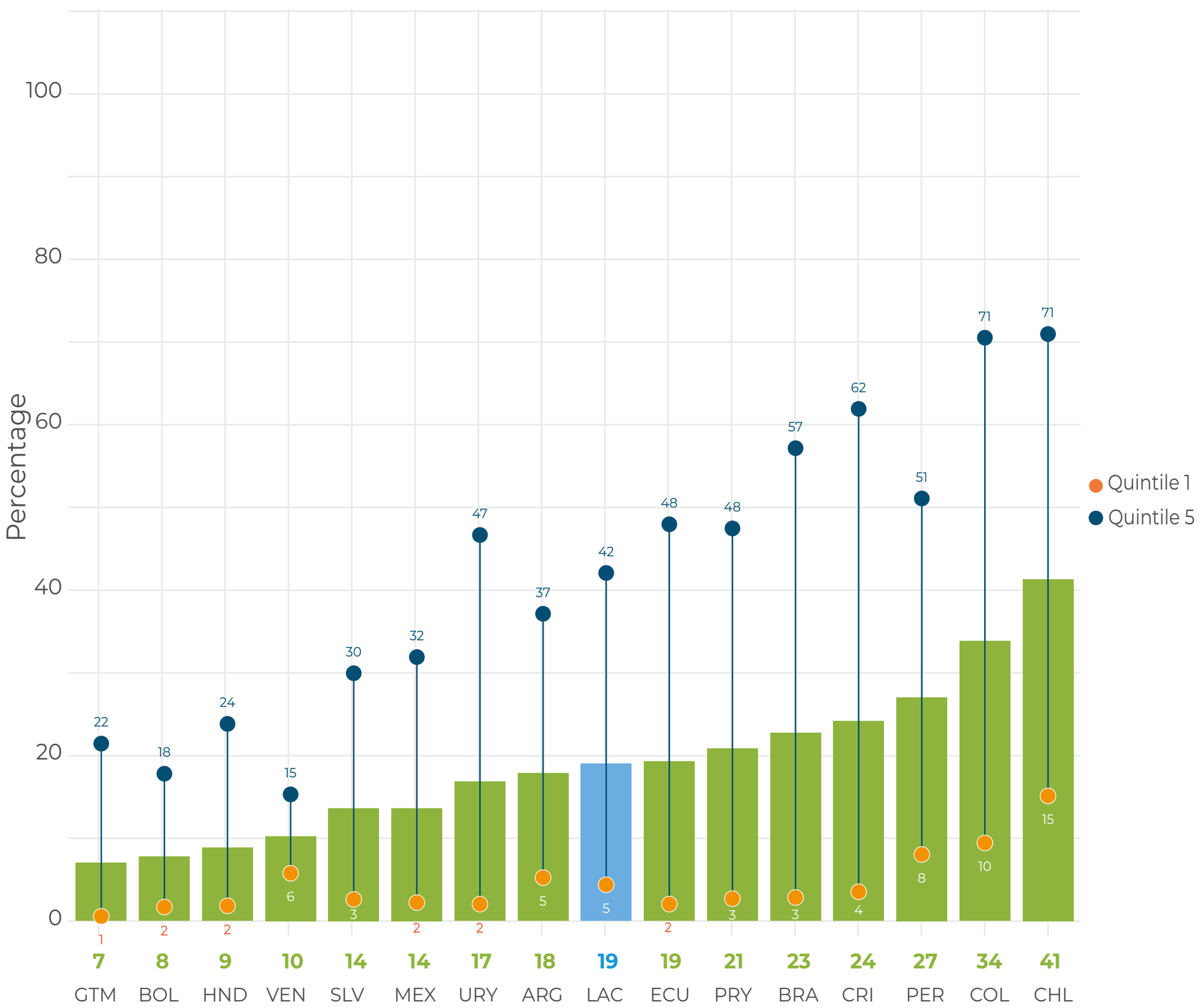
16 The completion rate is the percentage of a cohort of people aged 3 to 5 years older than the theoretical age for the last grade of secondary education who have completed this level.

17 Data for Bolivia (BOL), Dominican Republic (DOM), Guyana (GUY) and Venezuela (VEN) are for 2021; data for Chile (CHL), Argentina (ARG), Brazil (BRA), Panama (PAN), Uruguay (URY), Costa Rica (CRI), Mexico (MEX), Colombia (COL), Paraguay (PRY), Peru (PER), Suriname (SUR), and Guatemala (GTM) are for 2022; data for Ecuador (ECU), El Salvador (SLV), and Honduras (HND) are for 2023. The value for Latin America and the Caribbean (LAC) is calculated as a simple average of country data.

In LAC, 19% of people aged between 25 and 34 have completed tertiary education[18]. This low completion rate is linked to low attendance and completion rates at the secondary level. The countries with the highest efficiency are Chile (41%), Colombia (34%), and Peru (27%). The countries with the lowest efficiency are Guatemala (7%), Bolivia (8%), and Honduras (9%).

An analysis of the gap in the tertiary education completion rate between the highest and lowest SES categories reveals significant regional inequalities. While the completion rate at the highest SES categories is 42%, it is 5% at the lowest level: a gap of 37% percentage points. Furthermore, as the average tertiary education completion rate increases, the gap between the region's highest and lowest SES widens (Figure 15).

Figure 15: Tertiary education completion rates by socioeconomic status (circa 2023)[19]



Source: Own elaboration based on CIMA.

18 The tertiary education completion rate is the percentage of people between the ages of 24 and 35 who have completed tertiary education

19 Data for Bolivia (BOL) and Venezuela (VEN) are for 2021; data for Chile (CHL), Argentina (ARG), Brazil (BRA), Uruguay (URY), Costa Rica (CRI), Mexico (MEX), Colombia (COL), Peru (PER), Paraguay (PRY), and Guatemala (GTM) are for 2022; data for Ecuador (ECU), El Salvador (SLV), and Honduras (HND) are for 2023. The value for Latin America and the Caribbean (LAC) is calculated as a simple average of country data.

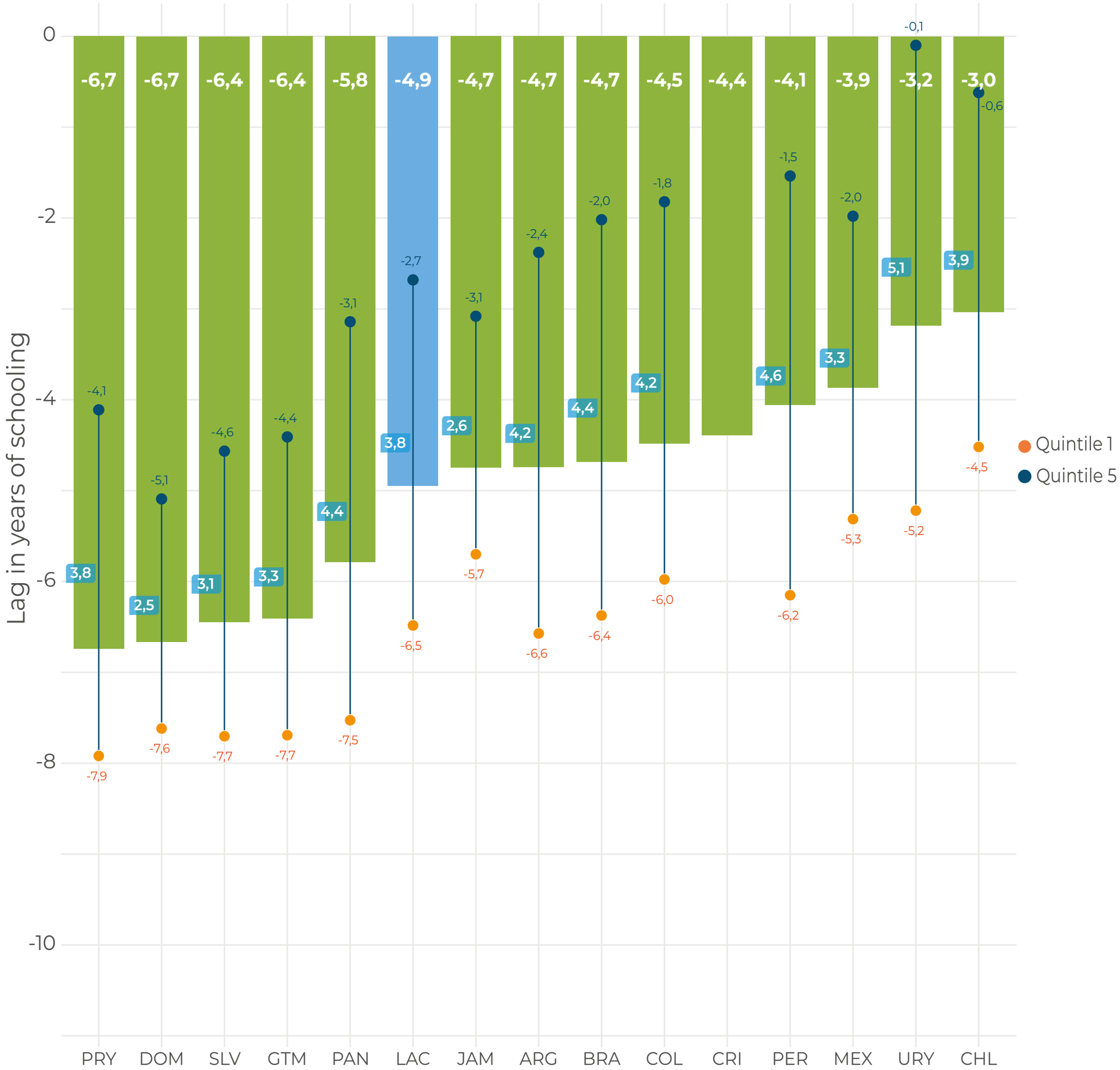
I.5 Learning

We use the PISA 2022 mathematics test results to analyze educational success in terms of learning in LAC. These results have revealed two significant challenges in the region. First, many students fail to develop the fundamental math skills necessary to continue learning and reach their potential[20]. Second, considerable inequality persists, which disadvantages the most socioeconomically vulnerable students (Arias Ortiz et al., 2023b). Since 2018, access to quality education for the poorest students in LAC has dropped or stagnated (Arias Ortiz et al., 2024). Between 2018 and 2022, the share of low-performing students in mathematics among the poorest increased between 3 and 8 percentage points in Brazil, Colombia, Mexico, and Panama. This percentage remained unchanged during the period in the other countries that participated in both rounds. As for the long-term trend, between 2012 and 2022, the percentage of low-performers in mathematics among the poorest increased or remained unchanged in almost every country in the region. Specifically, the prevalence of low mathematics performance among the poorest students in Argentina, Brazil, and Mexico increased between 3 and 10 percentage points (Arias Ortiz et al., 2024).

A 20-point increase in a PISA test roughly reflects the average annual progress of students participating in the assessment (OECD, 2023). In other words, a score 20 points below the OECD average is equivalent to one year behind the OECD average (Arias Ortiz et al., 2023b). Latin America and the Caribbean lag behind the OECD average by an average of five years, and all countries lag behind at least three years. The countries with the highest lag compared to the OECD average are Paraguay (7 years), Dominican Republic (7 years), and El Salvador (6 years). There are also significant gaps within the countries in the region between students from the highest SES categories and students from the lowest SES categories. In over half the countries, students in the lowest SES categories lag at least six years behind the OECD average; the greatest lag is found in Paraguay (8 years), El Salvador (8 years), and Guatemala (8 years). The most significant gaps between the highest and lowest SES categories are found in Uruguay (5.1 years), Peru (4.6 years), Brazil (4.4 years), and Panama (4.4 years) (Figure 16).

20 PISA 2022's main domain of study was mathematics, although reading and science were also assessed.

Figure 16: Lag in years of schooling compared to OECD average (2022)



Source: Own elaboration based on CIMA.

Part II.

Evolution and perspectives on learning assessments

Different aspects of school systems–including teacher competence, school characteristics, and education policies—can be optimized to improve learning opportunities for groups of students in different circumstances. Research and public policy in education aim to understand the factors that make an education system efficient, effective, and equitable in a specific context (Clarke & Lune-Bazaldua, 2021). Implementing large-scale assessments, both nationally and internationally, has opened up numerous research opportunities. These help us better understand how this combination of factors—implemented as public programs and policies—has improved education systems in different countries (Teig & Steinmann, 2023).

Why should we talk about learning assessments? First, large-scale national assessments allow us to measure student performance concerning the learning objectives established in the national curriculum or curricular guidelines. These assessments are essential for monitoring the quality of education systems, focusing on learning outcomes rather than other indicators that measure inputs, access, or completion. As education policy discussion increasingly focuses on quality and equity of learning, the use of these assessments to monitor outcomes increases. This reflects a significant shift from the initial use of assessments, which sought to contrast theoretical knowledge about education processes with empirical evidence, to an approach more focused on assessing the quality of education systems (Wagemaker, 2014). In addition, large-scale assessments can provide information on the relationship between context, resources, and performance. This makes it possible to monitor equity in learning and, ideally, to promote policies that foster a more equitable distribution of educational opportunities (Arregui, 2008). In this sense, policymakers frequently use large-scale assessments to diagnose problems and target interventions for disadvantaged groups.

Learning assessments monitor education quality and equity but are also essential for accountability and policy leverage. These assessments allow policymakers to ensure transparency in reporting results to stakeholders and consolidate trust in education systems. Leverage, on the other hand, refers to the ability to use assessment results to drive policy priorities, raising awareness among various stakeholders about critical aspects of education quality or the relevance of a reform problem that has been identified. Both objectives, accountability and leverage, underscore the importance of assessments as strategic for developing and managing education policies.

Learning assessments should be consolidated as efficient tools for collecting and providing evidence to inform education policy decision making. In recent decades, several LAC countries have strengthened their learning assessment systems, producing tangible benefits in terms of public policy. In this context, it is critical to understand the relationship between large-scale assessments and the policy cycle.

Tobin, Nugroho, and Lietz (2016) provide a comprehensive review showing how, in middle- and low income countries, assessments are most frequently used in the agenda setting, policy implementation, and monitoring and evaluation stages and less frequently in policy formulation. While it is indisputable that national assessments are essential for policy monitoring and evaluation, they must significantly impact education policy making. Assessments can influence education policy in four key areas: curriculum reform and performance standards, assessment policy, resource allocation, and teaching and learning policies (Tobin, Nugroho, & Lietz, 2016).

Several regional examples illustrate how national learning assessments have been essential for policy formulation in these areas. Colombia (2002), Guatemala (2005), and Argentina (2004) implemented reverse engineering processes in the area of curricular reform and performance standards. These assessments fostered the development of more transparent and measurable learning standards by structuring normative tests[21] (Arregui, 2008). Chile is an outstanding example of resource allocation. The SEP Law, enacted in 2008, uses the results of the assessments to classify schools according to their performance, allocate resources and determine their level of support and autonomy according to their classification. This equity-oriented reform was based on publishing persistent learning gaps, crucial for boosting resource redistribution to the poorest schools (Galas, Gutiérrez, & Hamilton, 2022).

In terms of teaching and learning policies, if the results of national assessments are disseminated effectively in schools, they can guide changes in classroom practices and improve teacher training. In Argentina, studies such as Hoyos, Ganimian, and Holland (2019) show that diagnostic feedback on assessment results is linked to better student performance, as teachers spend more time on instruction and use more learning activities in the classroom. Similarly, when principals have access to this feedback, they are more likely to use the results to make management decisions, such as setting school-level learning goals, updating the curriculum, and making staffing decisions (Clarke & Lune-Bazaldua, 2021). A recent study in Haiti shows that sharing information on results with principals in rural areas had a positive impact on school management and student learning (Borger et al., 2024).

However, certain conditions must be met for learning assessments to influence the education policy cycle effectively. According to Tobin, Nugroho, and Lietz (2016), several key barriers can hinder this process: (a) issues with the quality and analysis of assessment programs; (b) financial constraints to conduct high quality assessments or to maintain ongoing programs; (c) weak assessment agencies susceptible to change and with fragile relationships with the government, and (d) insufficient technical capacity of staff to interpret and communicate results effectively. In the region, these barriers are more common than would be desirable.

As in other aspects of education policy, the closing of schools due to COVID-19 had a disruptive effect. In some countries in the region, assessments—especially census-based tests—helped prioritize actions during remote emergency education. PLANEA test results were used in Mexico to develop the curriculum and educational materials. The content that tends to be most challenging for students was analyzed, and pedagogical guides were created during the pandemic. In other countries, they played a central role in identifying the magnitude of learning losses and developing actions to address them. The continuity of Saber 11 tests In Colombia made it possible to measure the impact of COVID-19 on

21 In norm-based assessments, results are obtained by comparing students' performance against a reference group, which precludes comparisons unless equated tests are used. On the other hand, criterion-based assessment is based on the comparison of students' results with predefined criteria, generally derived from the curriculum. This makes it possible to determine the performance level based on previously established external standards (Murillo, 2007).

students' learning in the last grade of secondary education and to determine the characteristics of the schools with the highest dropout rates (Dueñas et al., 2022). In other countries, learning assessments took a back seat and were not prioritized as essential in the emergency.

We must better understand the role of learning tests in formulating education policies in the countries in the region. Therefore, in the following sections, we analyze the evolution of the national assessment systems in the region in terms of their participation in international and regional tests and of the national tests they have implemented.

II.1 International and regional assessments

Although large-scale assessments in education have been used in comparative studies for 60 years, the generalization and global discussion of their results have developed mainly in the last two decades (Bruns et al., 2019). Latin America has seen a significant increase in countries implementing national large-scale learning tests since the second half of the 1990s. We must consider international and regional learning assessments to understand the evolution and consolidation of national learning systems in Latin America and the Caribbean.

Regional and international assessments were initially developed to explore cross-national variation in educational institutions and processes and their relationship with learning outcomes (Keeves, 1995). They allow us to compare performance across countries and monitor the quality of learning globally (Vegas & Petrow, 2008). The Education for All (EFA) initiative was launched in 1990 during the World Conference on Education for All in Thailand. Led by UNESCO, EFA was a significant catalyst for advancing international and regional assessments, placing education quality and the need for ongoing monitoring at the heart of the global education debate. EFA promoted the development and implementation of international comparative assessments globally and the consolidation of regional assessments such as ERCE (for its Spanish Acronym) (Regional Comparative and Explanatory Study) in Latin America. The evidence shows the effect of this type of assessment on national education policy, with attributes that complement the role of national assessment systems. Participation in this type of assessment has significantly impacted the creation or reform of national assessment agencies and programs in middle- and low-income countries (Tobin, Nugroho, & Lietz, 2016). In addition, this participation has helped develop and strengthen technical and methodological capacities locally since the preparation and implementation process provides opportunities for technical exchange and learning about increasingly complex assessment methodologies (Ferrer & Fiszbein, 2015). Participation in such assessments has enriched and developed the capacities of national expert teams (Cox and Meckes, 2016).

Finally, international and regional assessments, especially PISA, stand out for their impact on developing education agendas in Latin America, serving as a powerful mechanism to leverage education policy priorities. According to the literature, these tests help shape the education agenda when international comparisons are used to trigger debates on education quality, mainly when the results are lower than expected (Tobin, Nugroho, & Lietz, 2016).

Regional assessments

In 1994, the Latin American Laboratory for Assessment of the Quality of Education (LLECE) of UNESCO's Regional Bureau for Education in Latin America and the Caribbean (OREALC) was created. It is a milestone for education assessment in Spanish and Portuguese-speaking countries in the region. LLECE appeared when many countries in the region were implementing education reforms without sufficient and relevant design information. Additionally, these reforms lacked a critical mass of resources to assess the quality of education

(Vanni & Valenzuela, 2020). By 1994, only six Latin American countries—Argentina, Brazil, Chile, Colombia, Nicaragua, and the Dominican Republic—had large-scale learning assessment tests, and only three of them (Argentina, Brazil, and Chile) had learning assessment experience in primary education.

Regional assessment programs were created in alignment with EFA to conduct comparative assessments between countries with similar geographic, cultural, linguistic, and economic contexts. This makes them especially valuable for middle- and low-income countries (Tobin, Nugroho, & Lietz, 2016).

The first version of the regional study was called PERCE (1997). It included twelve LAC countries and assessed reading and mathematics in grades 3 and 4[22]. The LLECE learning tests are based on a review of the curricular programs of the countries participating in each cycle. They aim to identify the shared content, and the tests are designed based on this content.

Subsequent rounds—SERCE 2006 and TERCE 2013—consolidated the participation of fifteen countries[23] , evaluated grades 3 and 6 of primary school, and included science and writing in grade 6. Fifteen countries also participated in the fourth Regional Comparative and Explanatory Study (ERCE 2019), but El Salvador joined for the first time, and Chile did not participate[24]. ERCE 2019 assessed student performance in grades 3 and 6 in mathematics, reading, and writing. In addition, performance in natural sciences was evaluated only in grade 6. The questionnaires addressed to students, teachers, families, and school principals were also reinforced. ERCE 2019 also included a new module to assess the socioemotional skills of grade 6 students, specifically focusing on awareness and valuing of others, self participating regulation, and self management (Vanini and Valenzuela, 2020).

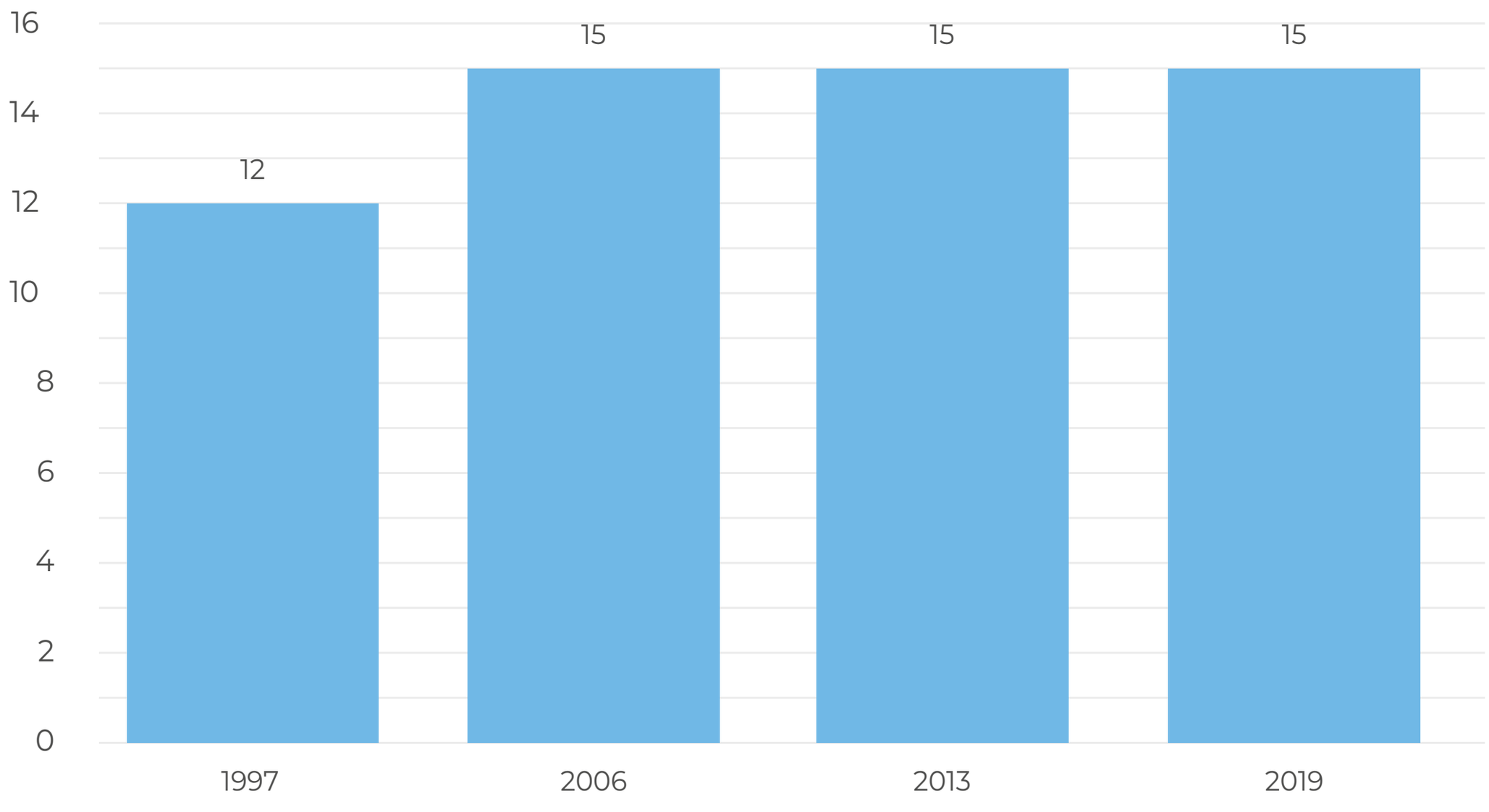
22 Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Honduras, Mexico, Paraguay, Peru, and Venezuela. Although thirteen countries participated, our analysis focuses on the twenty-six IDB member countries, so we do not include data for Cuba. This also applies to the following rounds.

23 Participants in SERCE 2006 included Argentina, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Mexico, Nicaragua, Panama, Paraguay, Peru, and Uruguay. The countries participating in TERCE 2013 were Argentina, Brazil, Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, and Uruguay.

24 Argentina, Brazil, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, and Uruguay.

LLECE assessments have had a high and stable level of participation and remain the most relevant regional assessment for Spanish- and Portuguese-speaking Latin American countries (Brazil), as they do not include English- and French-speaking countries in the region (Vanni and Valenzuela, 2020).

Figure 17: Number of countries participating in ERCE



Source: Own elaboration based on LLECE data.

Box 1:

The role of CXC in learning assessments in the English-speaking Caribbean

The Caribbean Examinations Council (CXC) has established itself as a crucial reference for learning assessments in the English-speaking Caribbean. CXC offers a wide range of large-scale assessments at different education levels, supports curriculum development, and offers services to education institutions in designing, developing, and administering assessments (Caribbean Examinations Council, n.d.).

CXC, established in 1972 by an agreement between governments of fifteen English-speaking Caribbean territories,[25] aimed to develop culturally relevant curricula and examination systems in a post-colonial context that emphasized the importance of cultural independence while considering the need to offer internationally competitive certifications (Gordon, 2019). It was the first transnational initiative in the region competent to develop and implement large-scale assessments. Since then, CXC has evolved to include sixteen Caribbean countries and four external territories (Caribbean Examinations Council, 2023)[26].

The early efforts of the CXC focused on developing the Caribbean Secondary Examination Certificate (CSEC). The CSEC is an assessment created in 1979 that provided a new regionally and culturally relevant roadmap, replacing the O-Level GCE system, administered by the British education authorities, as the standard for assessing students at the end of secondary education (Gordon, 2019). The CSEC encompasses general and technical proficiencies, providing students with certification for further studies and entry to the workplace (Thailinger et al., 2023).

Almost half a century later, the CSEC has consolidated its participation and expanded the skills and subjects evaluated. It currently evaluates thirty-three subjects: twenty-eight general skills and five technical skills. The grading structure is divided into six levels and reports on student performance. Secondary school graduation does not depend on passing the assessment. However, performance in the CSEC is a determining factor for progressing to tertiary education and having future employment opportunities. Admission to tertiary education requires passing at least five subjects—including English and mathematics—and positions in the public sector also require passing at least five CSEC subjects (Caribbean Examinations Council, n.d.).

CXC has developed a range of assessments and certifications to address the region's needs concerning basic education, albeit with less scope and participation. These include the Caribbean Certificate of Secondary Level Competency (CCSLC) in 2007, the Caribbean Vocational Qualification (CVQ) in 2008, and the Caribbean Primary Exit Assessment (CPEA) in 2012[27].

25 The early history of CXC dates back to 1946, long before international and regional assessments became prominent on the international agenda. Nearly two decades later, there was strong support in the Caribbean Conference of Heads of Secondary School for establishing an examination board for the region (Gordon, 2019).

26 Six of the CXS participating countries are IDB members: Barbados, Belize, Guyana, Jamaica, Trinidad and Tobago, and Suriname.

27 The CPEA is a test taken by students in the last year of primary school in some Caribbean countries before moving on to secondary education; no IDB member country participates in it.

The impact of the assessments led by the CXC in the Caribbean has been significant, strengthening the technical capacities of the participating countries. This has been possible thanks to close collaboration with experts and organizations in other regions of the world, which has maintained the quality of the assessments. Unlike the ERCE, the CSEC is a high-stakes assessment[28] as it aims to certify knowledge and is needed for admission to tertiary studies and the public sector. Despite these characteristics, the CSEC remains the primary source of cross-national information from the English-speaking Caribbean countries to understand the quality of student learning at the end of secondary education.

The CXC experience can be a solid foundation for future collaborations and improvements in large-scale assessment in Latin America and the Caribbean. Building bridges between the various regional initiatives, particularly with LLECE, is crucial to strengthen and enrich regional assessment practices.

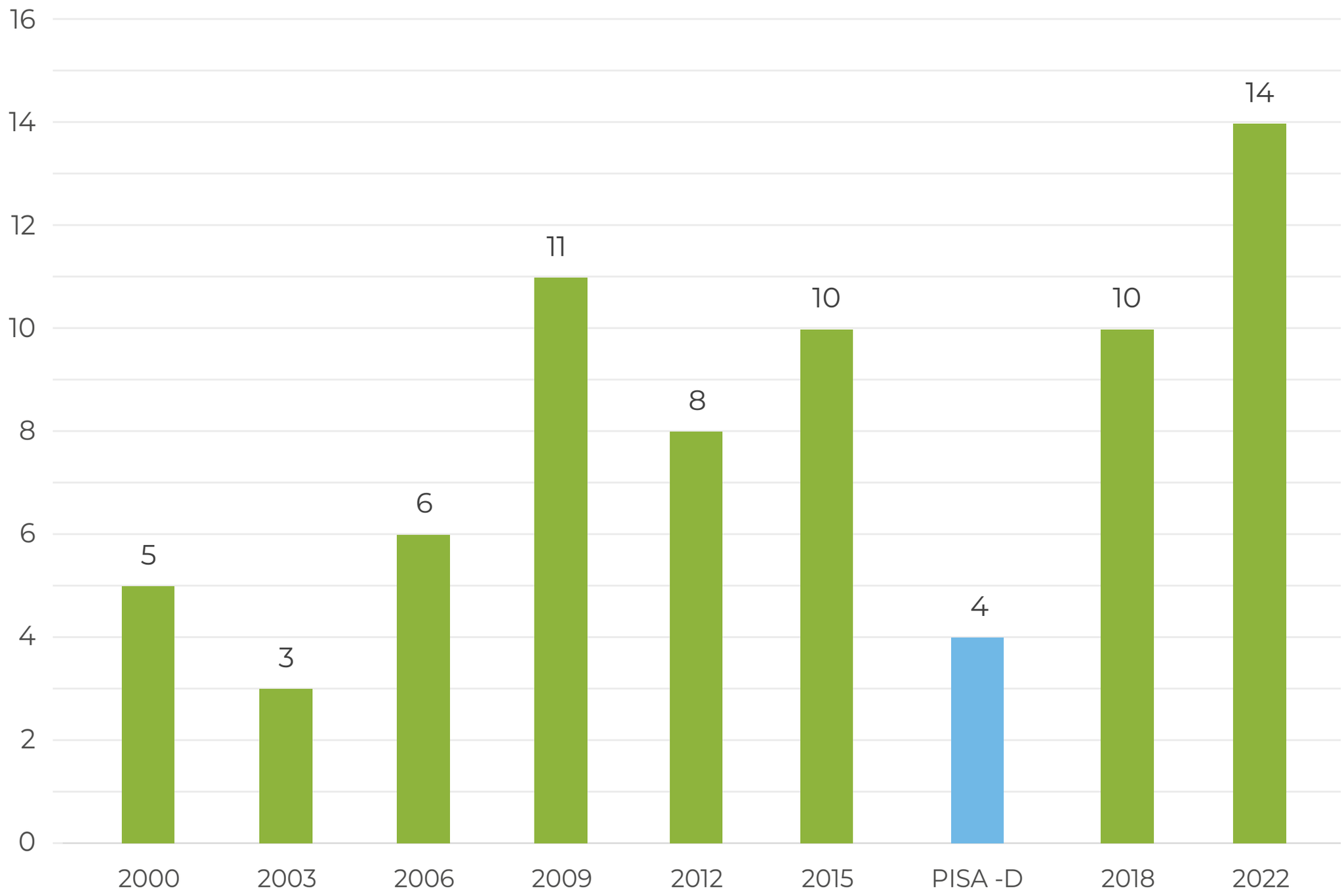
28 Following the definition by Herrero et al. (2022), high-stakes tests include those in which the consequences associated with their results are binding for different stakeholders in the system (students, schools, or teachers). These can be tests to certify attending a course or education level (primary or secondary) or for selection or admission to higher levels; or assessments for monitoring purposes, where the results are used by educational institutions for accountability purposes or there are incentives associated with the results.

International assessments

The creation of the Programme for International Student Assessment (PISA) in 2000 outlined a common policy agenda to assess educational quality and foster evidence-based policies. The PISA assessment tests 15-year-old students in at least seventh grade and measures not just whether students can reproduce what they have learned but also how well students can apply their knowledge in unfamiliar settings, inside and outside of school. (Arias Ortiz et al., 2024). PISA assesses knowledge and skills in mathematics, science, and reading and is conducted every three years. However, the eighth round, scheduled for 2021, was postponed to 2022 due to the COVID-19 pandemic.

Forty-three countries participated in the first round of PISA (2000), including five Latin American countries: Argentina, Brazil, Chile, Mexico, and Peru. The participation of countries from the region has been growing steadily; in 2022, there was a record of 14 LAC countries participating. Since the 2010s, international assessments such as PISA are no longer limited to middle- and high-income countries as in their beginnings but are starting to cover more middle- and low-income countries (Ramírez et al., 2018).

Figure 18: Number of countries participating in PISA



Source: Own elaboration based on OECD data.

The PISA for Development (PISA-D) assessment, conducted in 2017, deserves special attention given its relevance to the region. PISA-D was created to adapt the assessment to the contexts of low- and middle income countries, introducing items designed to assess fundamental skills. Seven countries participated, including four from the region: Ecuador, Guatemala, Honduras, and Paraguay. The main objective was to better distinguish and show differences in the performance of students who would have been classified at the "Below 1" level in the regular version of PISA (OECD, 2018). In addition, PISA-D allowed countries with less developed assessment units to strengthen their technical capacities to potentially join the regular global assessment program (Ferrer & Fiszbein, 2015). Paraguay and Guatemala participated in PISA 2022 following their participation in PISA-D.

Very few countries in the region participate in other international assessments. The International Association for the Evaluation of Educational Achievement[29] (IEA) has regularly conducted large-scale assessments since 1995. The IEA assessments include the assessment of student achievement in mathematics and science (TIMSS), reading comprehension (PIRLS), civic and citizenship education (ICCS), and student digital literacy (ICILS). The participation of LAC countries in these four tests is meager and even lower in the most recent rounds. Chile is the only country in the region participating in the four international assessments mentioned above. However, it has not participated in the latest rounds of TIMSS and ICCS. Colombia, Argentina, and Brazil have participated in three of the four evaluations, although only Brazil has been active in the latest rounds.

29 International organization of national research centers and institutions, government research agencies, academics and analysts working to research, understand, and improve education worldwide.

Table 1: Participation of LAC countries in IEA International tests

Countries	PIRLS					TIMSS								ICCS			ICILS		
	2001	2006	2011	2016	2021	1995	1999	2003	2007	2011	2015	2019	2023	2009	2016	2022	2013	2018	2023
Chile				✓	✓		✓	✓		✓	✓			✓	✓		✓	✓	✓
Colombia	✓		✓			✓			✓					✓	✓	✓			
Brazil					✓	✓						✓	✓			✓			
Argentina	✓			✓				✓			✓						✓		
Trinidad and Tobago		✓	✓	✓															
Honduras			✓							✓									
Uruguay																		✓	✓
Belize	✓																		
Paraguay														✓					
El Salvador									✓										
Mexico														✓	✓				
Peru															✓				
Guatemala														✓					
Dominican Republic														✓	✓				
LAC	3	1	3	3	2	2	1	2	2	2	2	1	1	6	5	2	2	2	2

Source: Own elaboration based on IEA data.

Why are these assessments important for national assessment systems?

Regional and international large-scale assessments, especially ERCE and PISA, have significantly driven the development and evolution of national education systems in the region. Participation in international tests allows us to compare education systems between countries. Also, it provides the opportunity to share techniques, organizational structures, and policies that have been efficient and successful in other contexts. The availability of detailed and comparable information helps identify specific areas of high performance and flags the areas of greatest weakness (World Bank, 2011).

Indeed, participation in international and regional tests has been a key opportunity to strengthen the institutional framework and capacities of national education assessment systems. These experiences have driven improvements in assessment design and implementation, allowing countries to consolidate their technical capacities in this area (UNESCO, 2014).

In addition, participation in international assessments can lead to the creation and reform of national assessment agencies and programs (Tobin, Nugroho, & Lietz, 2016). Participation in assessments such as PISA has been essential to strengthening large-scale assessments in Brazil, contributing to technical and political convergence and consolidating the commitment to monitoring education policies (Fernandes & Gremaud, 2020). Specifically, participation in PISA in 2000 and 2003 prompted reforms in the Sistema Nacional de Avaliação do Ensino Básico (SAEB), transforming it into a census-based assessment that can analyze subgroups and performance factors more thoroughly (Saracho Martínez, 2007).

The LLECE has established itself as a benchmark for education assessment in Latin American countries. It provides a framework for cooperation and technical support that has helped train teams responsible for national assessment and evaluation systems (Vanni & Valenzuela, 2020). The horizontal collaboration promoted by LLECE has been essential. Balarin (2021) stresses that Latin American countries were vital in defining ERCE learning domains and developing test items, helping strengthen national assessment systems.

Preparing and implementing the PISA tests has also served as a mechanism to develop and strengthen local technical and methodological capacities, facilitating knowledge exchange and the acquisition of more advanced assessment methodologies (Ferrer & Fiszbein, 2015). It has also provided exposure to innovative ideas and best practices in large-scale assessment (Rivas, 2015). The joint work with PISA test experts has allowed local agencies to acquire new capacities, especially in item response theory, defining performance levels, establishing cut-off lines, and optimizing reliability when grading open-ended questions (Ferrer, 2009). In addition, instruments have been included to measure contextual factors that help explain student performance (Altinok et al., 2018).

The countries in the region are not leaders in the design and development of PISA tests. However, national assessment units consider them a valuable platform for developing and strengthening their technical and methodological capacities. Indeed, there is a significant learning process in various technical aspects of large-scale assessments, such as test design, working with open-ended questions, and developing instruments that enable time comparisons and quality control in the sampling, administration, and grading processes (Ferrer & Fiszbein, 2015).

Participation in international tests is also essential to promote national education policies. The tests can leverage education policy priorities and, in some cases, create a favorable environment for education reforms, standing out as essential when analyzing their international impact. Globally, the literature shows that the impact of testing on leveraging can go in both directions. For example, Germany's PISA 2000 results shocked public opinion and the political arena, opening a window for reforms (Heyneman & Lee, 2014). However, in England, the same results served to justify and legitimize already existing policies (Cox & Meckes, 2016). The impact of international assessments on agenda setting in middle- and low income countries often appears through international comparisons that spark debates about lower than expected education quality (Tobin, Nugroho, & Lietz, 2016). Participation in PISA in Latin America has highlighted education equity by showcasing the achievement gaps between population segments (Rivas, 2015). Moreover, in the region, international assessments are more often used to leverage policy priorities than other assessment types, such as regional or national tests (Tobin, Nugroho, & Lietz, 2016). For example, Brazil's poor performance in PISA 2000 and 2003 helped catalyze and support the education assessment and reform movement that had begun in the 1990s (Saracho Martínez, 2007). In Peru, PISA results influenced education policy in the first decade of the century, as the shock caused by the country's PISA results helped nurture a debate on education reform needs (Cueto, 2007).

Some Latin American countries are interesting examples where participation in large-scale regional and international assessments has influenced specific education policy reforms. The literature suggests that education systems often incorporate new knowledge and skills into curricula based on their participation in international assessment programs. This usually translates into curricular reforms and defining or adjusting performance standards (Cox & Meckes, 2016). In 2009, Chile aligned its mathematics and science curricula with TIMSS standards. It used the IEA Civic Education Study (CIVED) framework to reform citizen education, integrating it across the curriculum and emphasizing civic skills and knowledge (Cariola et al., 2011). Ferrer and Fiszbein (2015) emphasize that reading comprehension skills have been standardized into three distinct but complementary cognitive processes with PISA: locating information in a text, making inferences, and assessing the quality of the text to convey meaning. Before PISA, there were many definitions of "reading," depending on the tests or curricula implemented. In other cases, participation in

this type of assessment impacts teacher training reforms. PISA results informed the development of a federal professional training program, the Open University Brazil (UAB), which provides distance education and training for teachers and school leaders across various areas (Tobin, Nugroho, & Lietz, 2016).

Ferrer and Fiszbein (2015) reflect on the value of PISA for countries in the region. One of the most frequent objections is the perception that investing time and resources in an assessment in which the region tends to rank in the last places makes no sense. However, the authors argue that this overlooks the value of PISA, which provides a regular benchmark with reliable instruments every three years. In addition, the results may lead to national debates of interest, focusing on education quality, which may result in more significant support for reforms. As stated above, these tests can strengthen the technical capacities of national assessment teams, act as a lobbying tool to leverage political priorities and serve as catalysts for education reform and improvement. However, we must acknowledge that they are not appropriate for designing specific policies or programs at the school or classroom level or for addressing the needs of vulnerable or lagging sectors. This requires national and local assessments that provide a more detailed and contextualized diagnosis of the needs and challenges of each school and education community. This makes complementarity between international and national assessments essential.

Participation in international and regional assessments is particularly relevant for monitoring the targets established in SDG 4. The objectives related to the quality of learning in foundational skills (mathematics and reading) are framed in goal 4.1: “By 2030, ensure that all girls and boys complete free, equitable and quality primary and secondary education leading to relevant and effective learning outcomes” (United Nations General Assembly, 2015). A group of indicators has been defined to ensure adequate monitoring of improvements in the quality of basic education in foundational skills (4.1.1). They measure the proportion of children and adolescents who have reached at least a minimum level of proficiency in reading and mathematics at times points in the school cycle: (a) at the end of grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary.

It is challenging to monitor the indicators of target 4.1, given the scarcity of accurate and comparable data to assess learning outcomes and their global evolution over time (Montoya, 2023). Over 60% of countries globally still need data on learning quality despite taking recent actions to obtain reliable and comparable learning data (see Box 2). Of the twenty six LAC IDB member countries, sixteen countries have tracked proficiency levels in mathematics and reading at the end of grades 2 and 3 and at the end of lower secondary education during the 2016-2022 period. Additionally, seventeen countries have done so at the end of primary school, all based on their participation in regional and/or international tests.

ERCE has established itself as the primary tool for monitoring learning in the 2030 agenda for Latin America and the Caribbean, enabling monitoring in sixteen countries in the region[30]. ERCE monitors reading and mathematics indicators at the end of second or third grade and at the end of elementary school. Trinidad and Tobago’s participation in PIRLS brings the number of countries reporting on this indicator to seventeen for reading at the end of primary school (considering the 2016-2022 period). On the other hand, PISA and PISA-D are the primary sources of information for monitoring both subjects at the end of lower secondary school. They enabled monitoring in fourteen countries in the region by 2022 and sixteen countries in 2017-2022. Except for Trinidad and Tobago and Jamaica, Caribbean countries do not currently participate in regional or international assessments that meet the UIS criteria (Box 2). This makes it difficult to follow up and monitor learning in line with SDG 4.1.

30 In Bolivia, information is reported for 2017 with data from the National Diagnosis conducted after implementing TERCE instruments, though they are not strictly comparable.

Box 2:

Learning assessments and monitoring of SDG 4

Given the lack of data, the Global Alliance for Monitoring Learning (GAML), with technical support from the Australian Council for Educational Research (ACER) and the Global Education Monitoring Report (GEMR), has defined global Minimum Proficiency Levels (MPL) and their alignment to the results of various international and regional assessments. This makes it possible to use multiple sources of information to monitor the quality of learning to promote better coverage (UNESCO Institute for Statistics, 2024). The international assessment programs used to report on group 4.1.1 are the Programme for International Student Assessment (PISA), the International Reading and Literacy Study (PIRLS), and the Trends in International Mathematics and Science Study (TIMSS). The protocol also considers five regional assessments applied globally, including Regional Comparative and Explanatory Studies (ERCE) as regional evidence for Latin America and the Caribbean.

	Grade 2 or 3		End of primary		End of lower secondary	
	Reading	Mathematics	Reading	Mathematics	Reading	Mathematics
PIRLS (I)	✓		✓			
TIMSS (I)		✓		✓		✓
PISA/PISA-D (I)					✓	✓
SEA-PLM (R-Southeast Asia)			✓	✓		
ERCE (R-LAC)	✓	✓	✓	✓		
PASEC (R-Africa)	✓	✓	✓	✓		
SACMEQ (R-Africa)			✓	✓		
PILNA (R-Pacific)	✓	✓	✓	✓		

Fuente: GAML (2023). I=evaluación internacional, R=evaluación regional.

The potential use of national learning assessments as a reporting source for SDG 4.1.1 indicators is still in its pilot stage and subject to compliance with the six criteria established by the GAML (UNESCO Institute for Statistics, 2024):

1. The assessment uses item response theory (IRT);
2. Results are available as a percentage of students by proficiency level;
3. A proficiency level descriptor is aligned to the minimum proficiency level (MPL);
4. The content sufficiently covers the overall content framework in the relevant domain, either in **reading** and/or **mathematics**. The content alignment tool enables this mapping, and guidance is provided in the report **(CAT)**;
5. It complies with the minimum good practices. The procedure alignment tool allows countries to assess the level of compliance with a set of good practices and, most importantly, achieve a minimum level **(PAT)** and
6. Characteristics are added to the data points (name of the national assessment, minimum proficiency level and grade).

Finally, the GAML has also developed the Assessment for Minimum Proficiency Level **(AMPL)**, which determines whether a student has reached the minimum proficiency level in mathematics or reading for their specific education level. These tools produce comparable results across countries and can be used to report on indicators 4.1.1. a and b. The assessments are part of a set of accessible methodological tools to measure learning, build capacity to generate, analyze and periodically report learning data, and monitor progress towards SDG 4.

II.2 National learning assessments in the region: evolution and current status

National learning assessments in Latin America[31] evolved in three major stages: a first stage in the 1990s of building assessment capacities; a second stage in the 2010s where, generally, the systems were expanded; and a third stage from the 2010s until the pandemic disruption, featuring a consolidation and revision of assessment objectives and concepts (Galas, Gutiérrez, & Hamilton, 2022). Although these stages shape the main discussions and technical transformations in learning assessment in the region, assessment systems and large-scale national learning tests have evolved in several ways. Below, we delve into the main characteristics of learning tests regarding scope, quality, and governance.

Building assessment capacities (1990-2000)

The first stage features the development of assessment capacities. The first assessments were driven by international partnerships and administered by universities, ministries of education, and assessment institutions. According to classical test theory, mathematics and language learning were measured with a normative and psychometric approach.

In most countries in the region, experience in large-scale educational assessment was slowly acquired by administering tests with specific purposes such as regulating admission to higher education (Colombia), measuring the impact of an education improvement project funded by an international organization (Guatemala, Ecuador, and Honduras), or regulating the private sector and helping families make informed decisions in a context of greater school choice (Chile). In the 1990s, several countries adopted assessment through standardized tests as part of the decade's more or less comprehensive education reform programs. As seen below, these assessments were conducted within institutional and organizational frameworks explicitly created for assessment purposes (Ferrer, 2006). In exceptional cases, such as Chile, the focus of the tests quickly shifted to using the results to target resources and interventions in low-performing schools more effectively (Elacqua & Fabrega, 2004). In 1997, the results of the assessments were essential for the Chilean Congress to approve an extension of the annual school time by 40% to address the education quality deficit detected in 1994 when almost 40% of grade 4 students did not understand the texts they read (Schiefelbein & Schiefelbein, 2003). In Uruguay, the results of the first national assessment conducted in 1996 showed the relationship between the sociocultural context of students and their learning outcomes, highlighting the need to include these factors in education reforms (Benveniste, 2000).

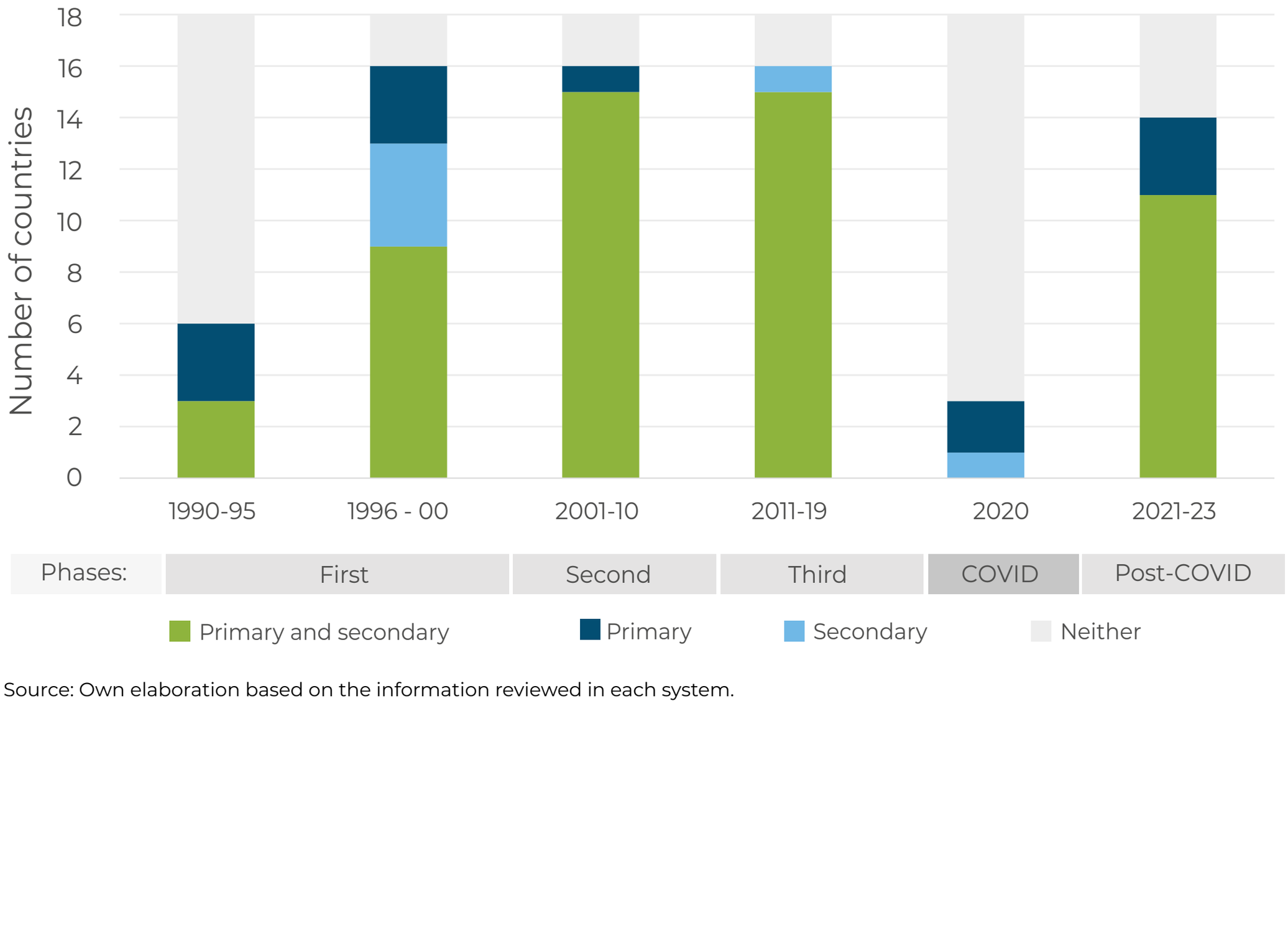
At this stage, the funding and institutional framework of the assessment processes were diverse and changing. Most countries in the region began the learning assessment process with different funding sources, either international[32] or national. The new assessment responsibilities were attributed to specialized departments in the ministries of education (Argentina, Bolivia, Ecuador, El Salvador, Mexico, Paraguay, Peru, Dominican Republic, Uruguay, and Venezuela); existing institutions that expanded their roles (INEP in Brazil and ICFES in Colombia); and universities, based on their technical experience and in partnership with the ministries of education (Chile, Costa Rica, Guatemala, and Honduras).

31 The early creation of LLECE led to intense exchange and collaboration within Latin America, which did not occur in the Caribbean to the same degree. Therefore, here we focus on the history of the large-scale national assessment systems in the eighteen Latin American countries, Caribbean countries excluded.

32 The first national learning tests in countries such as Guatemala, Ecuador, Honduras, and Uruguay were funded and promoted by international organizations.

The assessments focused on mathematics and language and were normative in their methodological approach. In this period, the assessments focused on measuring academic achievement in language and mathematics; only Chile, Brazil, and Argentina assessed natural and social sciences, and Uruguay measured citizen attitudes. Most countries implemented a conventional testing methodology and a normative approach, where results are interpreted regarding a particular group usually referred to as the "standard" or "normal" group. These tests summarized the results in scores or number of correct answers, known as classical test theory (CTT). Only Brazil, Chile, and Mexico included an item response theory (IRT) approach early on. They used a criterion-referenced approach: the results were interpreted based on a given skill or what the student knows how to do (Harris, 1991). Item response theory provides a more detailed description of what students know and can do, allowing for a more accurate assessment of knowledge level and offering greater potential for pedagogical interventions than reports based only on the mean percentage of correct responses (Ferrer & Fiszbein, 2015).

Figure 19: Countries with large-scale learning tests by education level; 18 countries (1990-2023)



Source: Own elaboration based on the information reviewed in each system.

During this period, countries also made decisions on the education level to be evaluated and the coverage of the assessments (sample- or census-based). Until the first half of the 1990s, only one third of the countries in the region had any large-scale national learning tests (Figure 19). In primary education, only three countries (Argentina, Brazil, and Chile) conducted tests, assessing performance in mathematics and language. Chile administered census-based tests in primary school in mathematics and language and measured natural and social sciences with census-based tests. Only Colombia and Costa Rica conducted national census-based tests for students in the last year of secondary education. Argentina began to evaluate students in the first and last year of secondary education in mathematics and language.

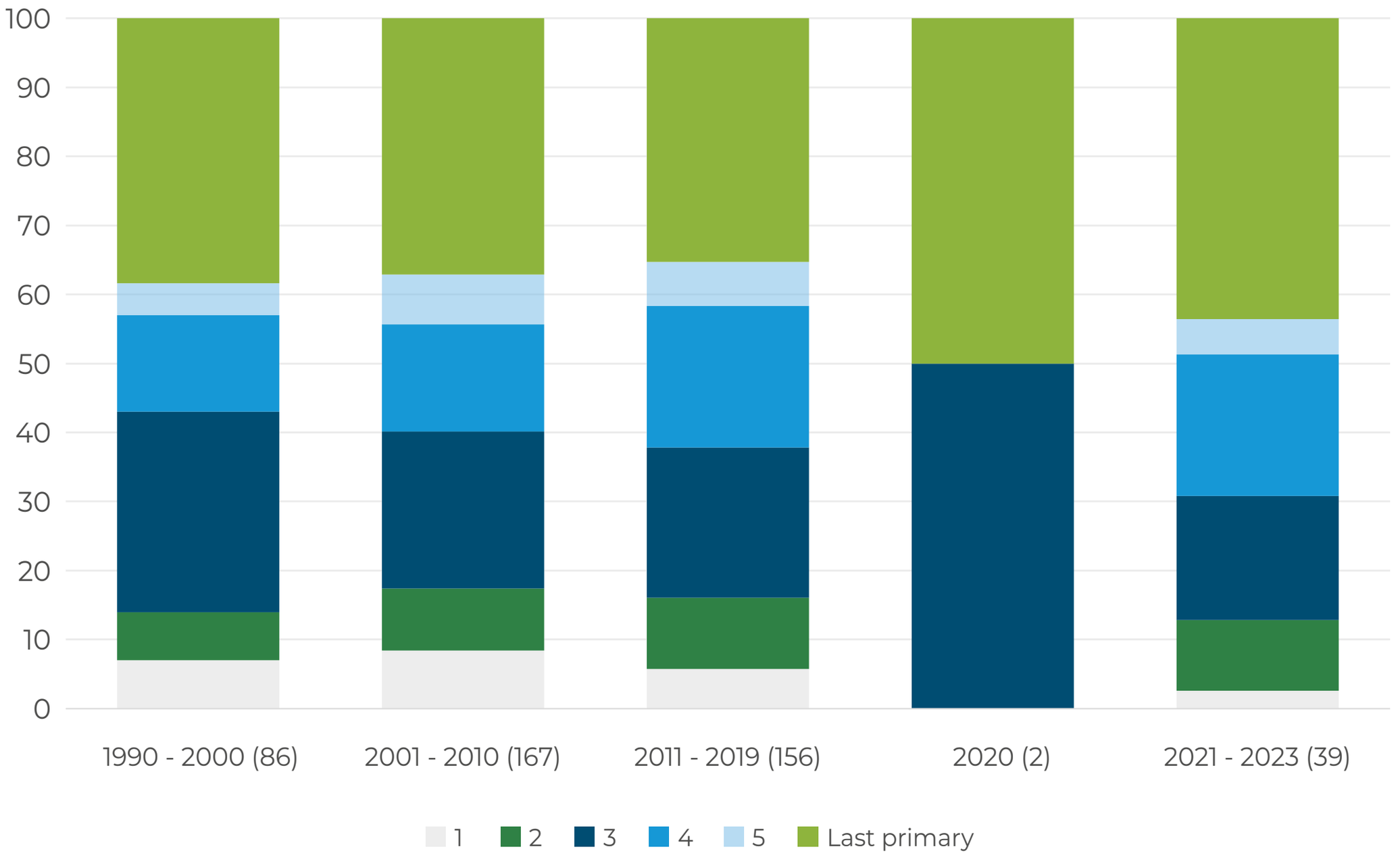
By the end of the decade, sixteen of the eighteen countries analyzed had conducted at least a first assessment at primary (Bolivia, Honduras, Guatemala, and Mexico), secondary (Colombia, Dominican Republic, and El Salvador), or both levels (Argentina, Brazil, Chile, Costa Rica, Ecuador, Peru, Paraguay, Uruguay, and Venezuela). The tests in primary education were generally administered to students in the last year and in intermediate grades, such as grades 3 and 4. In contrast, the secondary education tests were mainly applied to students in the last year of lower secondary school (Figure 20). The tests lacked significant impact and applicability in education policy, and there was even resistance to disseminating their results[33]. However, census-based tests were used early on in the Chilean system to create rankings according to school scores. The first ranking that the System for Measuring the Quality of Education (SIMCE, for its Spanish acronym) (Sistema de Medición de la Calidad de la Educación) published did not provide information about the socioeconomic characteristics of the institutions. This gave rise to misinterpretations and criticism because the results were not comparable[34]. On the other hand, census-based tests in the last year of secondary school in the Dominican Republic, El Salvador, and Costa Rica were high-stakes tests and aimed to set a standard for passing the level. In Colombia, the tests were required for admission to higher education. Conversely, sample-based intermediate or last secondary school tests had more formative objectives. For example, in Argentina, Brazil, and Ecuador, reports were published with methodological recommendations for teachers. The reports discussed some of the test items and explained the difficulties faced by students in the exams. However, the results of the assessments were not part of the public debate or the education system (Ferrer, 2006).

33 In 2000, Mexico and Peru censored the publication of results obtained in international and national tests (Ferrer & Fiszbein, 2015).

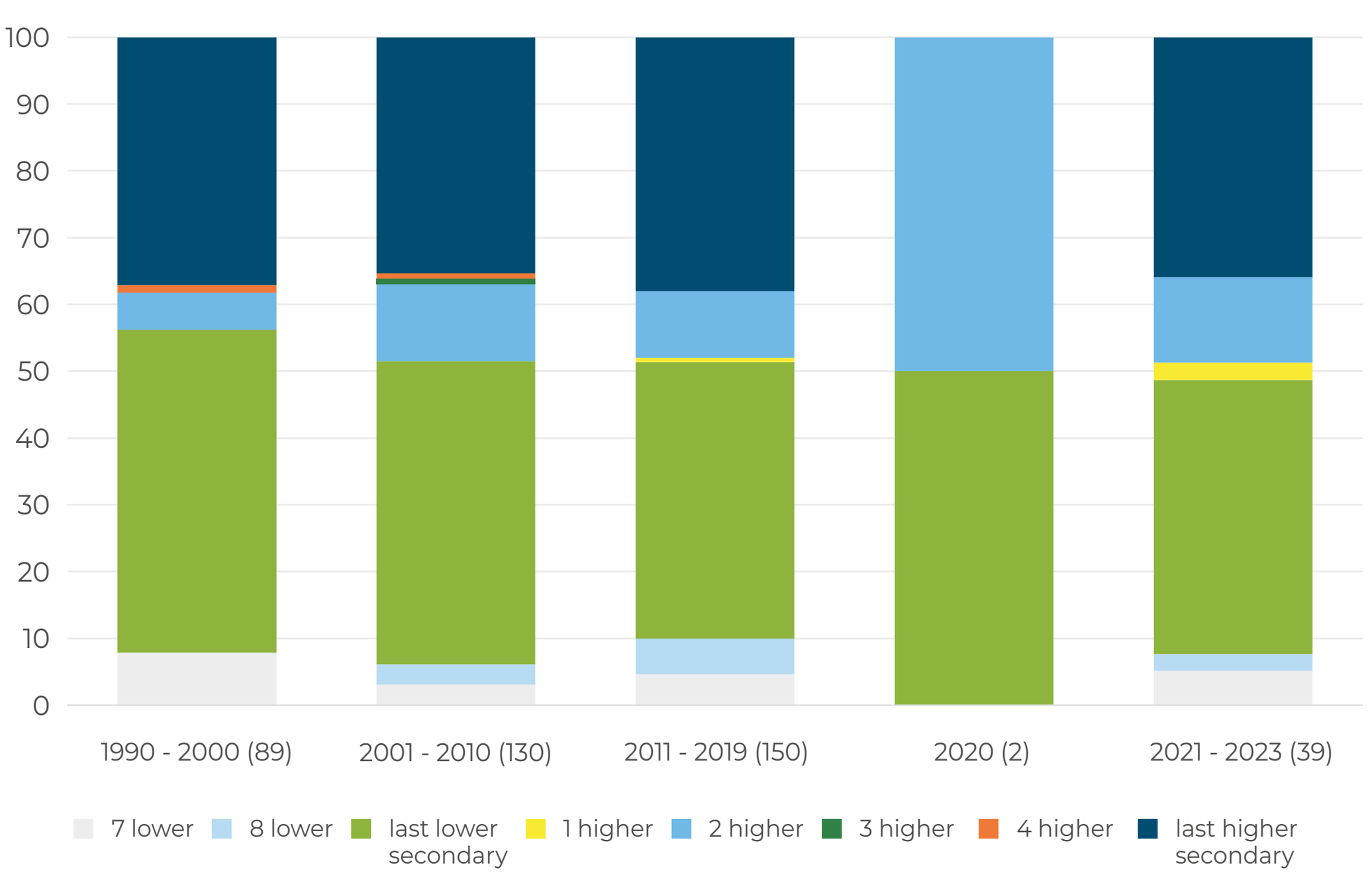
34 At this time, the Chilean press prepared school rankings using the information published by the government for each school. These rankings did not consider contextual factors and directly compared public and private schools. It was not until 2008, with the implementation of the SEP Law, that the government began to publish more robust rankings which were adjusted for external factors.

Figure 20: Grades assessed in large-scale learning tests by education level; 18 countries (1990-2023)

Primary – ISCED 1[35]



Secondary – ISCED 2 and 3



Source: Own elaboration based on the information reviewed in each system.

Note: The number of tests for each year is shown in brackets.

35 In terms of education achievement classification, the International Standard Classification of Education (ISCED) homogeneously classifies the education levels of primary education (ISCED 1), lower secondary education (ISCED 2), and upper secondary education (ISCED 3). For more details, see UNESCO-UIS (2013).

Institutional changes and expansion of educational assessment systems (2001-2010)

The first decade of the 21st century brought changes in the institutional framework of national assessments, a trend toward primary and secondary test collection using contextual forms, presented normative results, and a more significant move toward IRT approaches. Assessments were more widely disseminated and used.

The institutional framework of education assessment systems underwent several changes during this decade. In Costa Rica and Chile, the ministries of education assumed responsibility for the assessments previously shared with universities. In Panama, the ministry appointed the National Directorate of Education Assessment (DNEE, for its Spanish acronym) (Dirección Nacional de Evaluación Educativa) to design and apply assessments to produce information for decision making. In Mexico, the National System for the Evaluation of Education (INEE, for its Spanish acronym) (Instituto Nacional para la Evaluación de la Educación) was created in 2002 as a dependent public agency. In 2012, it gained autonomy by law in the education reform. In 2000, Argentina created the Institute of Education Quality (Instituto de Calidad Educativa), a semi-autonomous agency of the Ministry of Education, which was soon replaced by the National Directorate for Information and Evaluation of Education Quality (DiNIECE, for its Spanish acronym) (Dirección Nacional de Información y Evaluación de la Calidad Educativa), under the Ministry. Analyses of the time indicate that institutional arrangements were more stable when responsibilities were managed outside the structure of the ministries of education through institutes with greater administrative and technical autonomy (Ferrer, 2006). The following section returns to this discussion and includes other regional cases and the evolution of institutional arrangements.

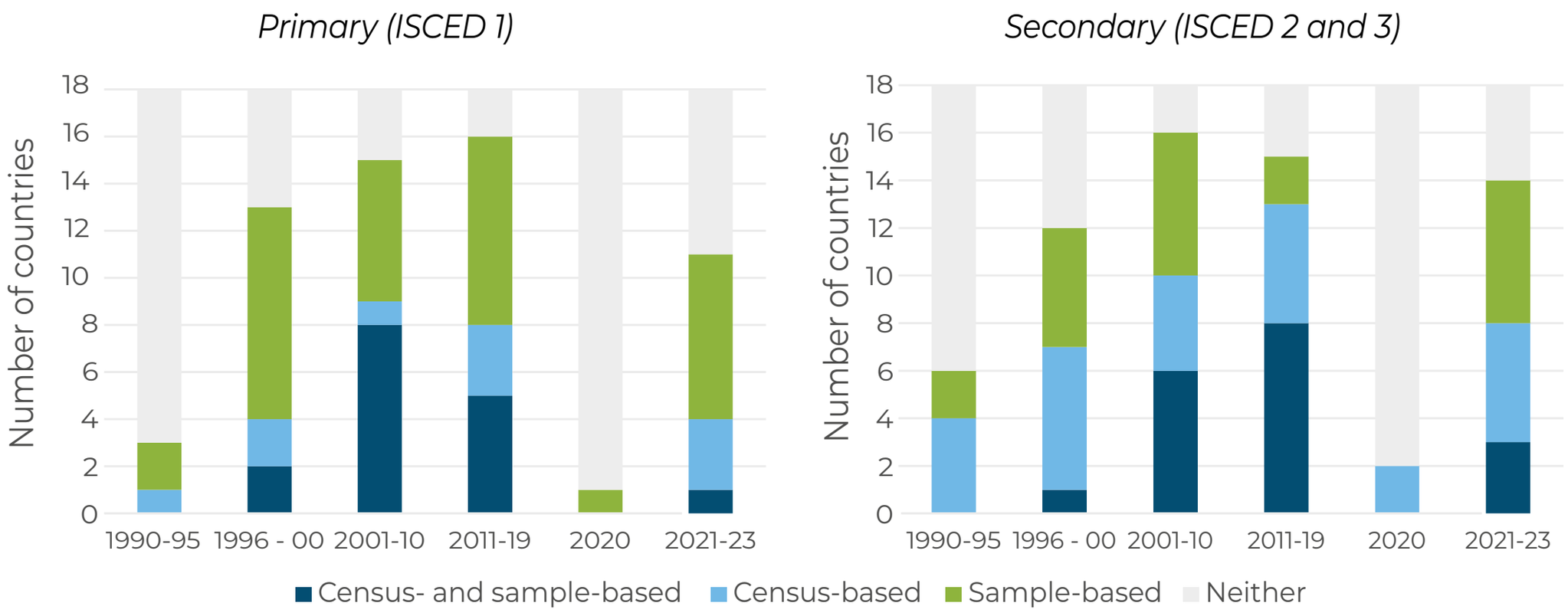
As for data collection, the number of countries with at least one test was maintained, and the number of countries collecting information for both levels increased. Sixteen of the eighteen countries in the region conducted an assessment. Colombia, Costa Rica, and Panama implemented large-scale tests in primary education for the last grade and intermediate grades (in the case of Colombia) for diagnostic purposes.

In addition, the thematic coverage and contextual information collected in the tests increased. Assessments during this period continued to focus on mathematics and language, and more countries measured results in natural and social sciences in primary education (Costa Rica, Ecuador, Panama, Dominican Republic, Colombia, and Mexico). Mexico included civic education in its assessments, and Uruguay stopped measuring citizen attitudes and included natural sciences. At the beginning of this decade, context questionnaires began to be widely used to improve the interpretation and dissemination of learning outcomes. As the assessment systems consolidated gradually, there was a growing interest in delving into the impact of individual, institutional, and teaching factors on learning outcomes. Although Uruguay, Brazil, Argentina, Peru, Chile, and Colombia already included complementary forms for different types of education stakeholders (parents, teachers, and principals), during this decade, the analyses of results conducted by Chile and Colombia began to include socioeconomic indices of students and schools and the study of associated factors that made result comparison more "fair." Additionally, to some extent, they were able to measure the added value of the school given that the results could consider the differences in the students' environment (ICFES, 2017; Rodríguez-Garcés, Padilla-Fuentes & Suazo Ruiz, 2020).

Methodologically speaking, the psychometric theories used in the assessments were updated. In fact, by 2008, most countries in the region, including Bolivia, Argentina, Brazil, Chile, Guatemala, and Ecuador, reported results by achievement levels rather than scores (Galas, Gutiérrez, & Hamilton, 2022); eleven of the eighteen Latin American countries implemented IRT approaches in some of their assessments.

Practices for publishing results were also consolidated, and, as the technical aspects, recurrence and types of tests strengthened, the countries began to place greater emphasis on the use of their information. One of the first uses of the results and dissemination of the assessments in the region is reflecting on the learning goals and standards to be assessed. According to Arregui (2008), the countries in the region were not clear about how they would use the results at the beginning of the large-scale assessments. In a reverse engineering process, the countries began to assign different uses to the information collected once the collection process had begun. In this decade, considering the technical advances of the tests, measurable education standards were created in Colombia, Peru, and Brazil in search of the design and implementation of education policies based on their results. Additionally, resources were allocated to close learning gaps, pedagogical changes were encouraged, and accountability and the introduction of incentives for various stakeholders and education institutions were promoted by assigning specific implications of the test results to different stakeholders (Galas, Gutiérrez & Hamilton, 2022; Ferrer & Fiszbein, 2015). Making the results of the assessments available to researchers was another significant step that became more relevant at this time in Colombia and Brazil, and which will be discussed in more detail in the following section. In 1999, the Committee of Education Information Producers (COMEP) was consolidated in Brazil to promote coordination among educational information producers and to create an integrated system to disseminate specialized information. However, the dissemination and use of the results were still key weaknesses in the region's learning assessment systems, with low engagement of civil society members, which did not encourage the demand for better quality education (Ferrer, 2006).

Figure 21. Countries with large-scale national learning tests by type of test and education level; 18 countries (1990-2023)



Source: Own elaboration based on the information reviewed in each system.

In this regard, choosing between a census-based and sample-based approach began to depend on factors such as the intended use of the results and the available budget. Census-based assessments were essential in systems where testing was part of a school accountability system to provide formative feedback to schools through newsletters or where the aim was to identify low-performing schools and allocate resources or evaluate education policies. Sample-based assessments were considered more efficient in systems aiming to monitor the overall performance of the education system over time or to understand the contextual factors that affect learning. Both approaches can inform continuous improvement and education policy and practice. Selecting a sample of students for assessment is less costly than assessing all students. Still, it limits the ability to develop detailed school-level reports, which restricts the use of assessments to improve teaching practices in all schools. Some countries, such as Chile and Brazil, combine census- and sample-based approaches to design their large-scale national assessments to meet stakeholders' information needs more effectively. The combination helps reduce overall costs while providing a detailed description of the education system (Clarke & Lune-Bazaldua, 2021). Between 2001 and 2010, of the eighteen countries analyzed, at the primary level, six implemented only sample-based assessments, eight implemented sample- and census-based assessments, one implemented census-based assessments, and three did not conduct any assessments at this level. In secondary education, four countries implemented only census-based assessments, six sample- and census-based assessments, and six sample-based assessments, and only two (Venezuela and Bolivia) did not conduct any assessments (Figure 21).

Different practices and unintended consequences became the subject of controversy and analysis in systems with census-based high-stakes assessments[36]. In Mexico, for example, ENLACE test results were linked to teachers' salary increases in 2008. However, no policy was designed to monitor the impact on teaching in the classroom; that is, the connection could lead to curricular narrowing or test driven teaching. This encouraged unethical practices such as selling exams or over preparing certain types of students (Martínez, 2015). In Brazil, after the emergence of Prova Brasil (2005), with a census based version for public education and a sample based version for private education, the Basic Education Development Index (IDEB) was created in 2007. This set a significant standard in using learning tests to create synthetic indicators of education quality and allocate incentives within the system. The IDEB combines learning outcomes with pass rates, providing a comprehensive assessment of education quality beyond academic achievement. IDEB introduced an approach to accountability that was unusual in Latin America, establishing medium- and long-term education goals at different levels. This promotes progress monitoring and accountability (Ferrer & Fiszbein, 2015). Based on this first index, several Brazilian states designed their own systems to monitor quality per school and assign teacher salary incentives, while addressing some unintended practices (Box 3).

36 As previously mentioned, the definition from Herrero et al. (2022) is used, which states that high-stakes assessments are those with binding consequences associated with their results for various stakeholders in the system (students, schools, or teachers). These assessments may include tests for certification purposes at a particular educational level (primary or secondary) or for selection or admission into higher levels; or evaluations for monitoring purposes, where the results are used for institutional accountability or where incentives are tied to the outcomes.

Box 3:

Balancing the unintended consequences of learning assessments

High-stakes learning tests are of great value in several countries worldwide when combined with good teachers and a strategic vision (World Bank, 2018). Additionally, the tests have potential adverse effects that have been debated and studied, including the following:

- **Curriculum reduction and unethical practices.** High-stakes testing can narrow the scope of education by encouraging practices such as teaching to the test, excluding learning in non-tested areas, or limiting innovation in the classroom, as teachers may lack incentives to try new methods (Blazer, 2011; Ascorra et al., 2019). Strong accountability policies can also encourage unethical practices to achieve better test scores. The literature includes examples of students with low performance or special educational needs excluded from assessments in Chile and fraudulent actions such as the sale of exams in Mexico (Cáceres, Muñoz, & Valenzuela, 2021).
- **Distortions in teachers' and schools' perceptions of education quality.** Education assessment systems with high stakes tests in the region, such as Chile's, have published the results from very early on. This enables comparisons with schools of similar socioeconomic characteristics. Some stakeholders sometimes take this information out of context (e.g., the media or local authorities). This has been the subject of extensive debate due to the perceived stigmatization of teachers and schools from vulnerable contexts (MINEDUCA, 2014). Although INEP explicitly avoids ranking schools by average performance in Brazil, the media and other stakeholders (some ministers included) still use the results to create their own league tables. Although these practices may seek to apply internal and external pressure for schools to improve, they risk distorting perceptions of education quality because they usually overlook the context in which the school operates or the progress made over time (OECD, 2021). Publishing and using the results to promote debate, discussion, and pressure within the system to improve student learning is one of the main benefits of the tests. The essential thing to do is mitigate some of these risks by offering guidance on constructive ways to compare systems and schools, thus avoiding some of the unintended and potentially negative consequences of having such data. In Chile, evidence shows that accountability has had a positive impact: low performing schools have implemented policies to improve in the short term; chronically underperforming schools have invested in technical support and teacher training to avoid sanctions (Elacqua et al., 2016; Elacqua, 2019)..
- **Over-assessment and poor coordination among territorial authorities.** Federal countries like Brazil, with a strong student assessment culture and an established national assessment system, may be inefficient when coordinating assessments locally. In Brazil, about twenty-three of the twenty-seven federal units (states and the Federal District) have administered standardized assessments in recent years, and municipalities may also conduct assessments in their school networks. This arrangement is necessary in certain respects, as local assessments in Brazil often provide timely data that can be used for more formative purposes than those of SAEB. However, the situation also poses challenges regarding policy coordination, as students in various parts of the country may be subjected to multiple external assessments measuring the same subjects in the same school year (OECD, 2021).

Brazil and Chile have taken actions and made specific reforms to mitigate some of the negative impacts of the tests. For example, in Brazil, the states of São Paulo and Espírito Santo sought to promote equity in the design of education quality indices by favoring the performance of less privileged students. In São Paulo, four levels of student performance—from basic to advanced—were created, and the proportion of students in each level was calculated to frame the performance index in terms of the gap between each student's average and ideal position. The goal was to prevent teachers from focusing on the students most likely to show improvement in test scores, thus giving less attention to those at the higher and lower ends of achievement. In Espírito Santo, an additional level of student performance entitled "the excluded" was created. This level comprised all the students who dropped out of school or did not attend end-of-year assessments and were assigned a weight of zero. Here, they sought to promote efforts to prevent attrition and discourage the hiding of lower-performing students before the tests (Brooke, 2016). While these practices may be helpful to counteract the potential consequences of linking teacher salaries to test performance, several discussions remain about their consequences in narrowing the curriculum and limiting classroom practices to test preparation. In Chile, investment in improving the monitoring system has effectively countered the exclusion of disadvantaged students in assessments (Pérez & Maldonado, 2015).

While these efforts are essential, policymakers must balance the weight of testing in a country's assessment system. In this sense, promoting complementarity in the purpose and design of national and subnational standardized assessments is essential. This should be done to efficiently use resources, not overburden students, not send decontextualized information on test results, and to discourage unethical practices or those affecting specific population groups.

Consolidation of national assessment systems (2011-2019)

The second decade of the 21st century presented more consolidated assessment systems and new assessment institutions that applied census-based assessments at both levels. A critical and analytical approach was adopted on the conceptualization of education quality, the use of learning test results and the application of school environment, actions and attitudes, and socioemotional questionnaires.

Institutionally, this decade marked the emergence of several independent assessment agencies with specific impacts on each system. Independent agencies in the region share the mission of assessing learning but also differ in several dimensions, including their funding models, autonomy, grades assessed, governance models, scope regarding result delivery and dissemination, and influence on education policies. In 2011, Chile decentralized the Education Quality Assurance Agency (ACE, for its Spanish Acronym) (Agencia de la Calidad Educativa); in 2012, Ecuador created the National Institute for Education Evaluation (INEVAL, for its Spanish Acronym) (Instituto Nacional de Evaluación Educativa); in 2017, Paraguay created the National Institute for Education Evaluation (INEE, for its Spanish Acronym) (Instituto Nacional de Evaluación Educativa); and Bolivia created the Plurinational Observatory of Education Quality (OPCE, for its Spanish acronym) (Observatorio Plurinacional de la Calidad Educativa), an autonomous institution in charge of education assessment.

Each agency has specific functions and develops different types of tests. ICFES in Colombia and INEP in Brazil are in charge of all assessments, including SABER 11 and SABER Pro in Colombia and SAEB and ENEM in Brazil. In Chile, the DEMRE is in charge of the assessment for entry into higher education (formerly PSU, now PAES: Higher Education Access Test). In contrast, teacher assessment is still the responsibility of the CPEIP. The Education Quality Agency in Chile and INEEd (National Education Evaluation Institute) in Uruguay focus on standardized tests, while Uruguay's ANEP (National Administration of Public Education) oversees formative assessments and international examinations.

The funding models of independent assessment institutes vary. INEVAL (Ecuador), the Education Quality Agency (Chile) and INEEd (Uruguay) depend on an allocation from the national general budget. ICFES (Colombia) was created as a state-owned social enterprise that must generate resources. This is done from the revenues accrued with Saber 11, Saber TyT, and Saber Pro; other entities (such as the Ministry of Education) may hire ICFES to design and apply different tests. The 2009 reforms led the ICFES to assume more formalized responsibilities to provide assessment services, generate revenues, and use the surpluses to reinvest in technical and programmatic development (Ferrer & Fiszbein, 2015). The INEP (Brazil) has a mixed model: it receives income directly from the people who take the ENEM and a budget allocation for the school census and SAEB tests. In all cases, the authorities decide to administer the tests and allocate the necessary resources to do so, except those paid for by citizens (Saber 11, Saber TyT, Saber Pro, and ENEM). The institutional framework of each country determines the structure of the independent assessment institute if deemed necessary, as it is not the only alternative.

Regardless of the different models, few assessment institutes in the region have shielded themselves from changes in education policy decisions. For example, after the country's political context changed, Bolivia's OPCE has faced technical and political difficulties when implementing learning assessment activities[37]. In Mexico, the INEE, which had been the technical agency in charge of education achievement assessments, disappeared after the 2013 education reform and its implementation in 2018. This was a major setback in test administration[38].

Despite these institutional changes, the degree of application of the test was the same as in the previous decade. In this decade, the number of countries with tests in primary education increased with the implementation of sample-based diagnostic tests in the Dominican Republic; the number of tests in secondary education decreased given the lack of tests in Panama (Figure 19). By type of assessment, Nicaragua and Uruguay began to administer census-based tests at the secondary level, Panama at the primary level, and Argentina and Paraguay at both levels. Therefore, eight of the eighteen countries had census-based tests in primary and thirteen in secondary education (Figure 21).

On the other hand, there was a broad discussion on the conceptualization of education quality, the uses of tests, and the direction of public policy. This decade introduced a debate on education quality as a process of obtaining results in large-scale tests that only accounted for results in cognitive processes (Martínez-Iñiguez, Tobón & López-Ramírez, 2020; Vázquez Olivera, 2015). In this context, several countries focused on assessing other skills in addition to basic areas. Therefore, education quality indices were created and used to a greater extent to synthetically and accurately measure the quality of service provided by schools and, at the same time, guide schools, districts, managers, and teachers (Elacqua, Martinez, & Westh, 2019). Most countries in the region had dimensions and skills related to socioemotional development in their learning standards, either transversally or in a specific subject or time in the curriculum. Therefore, several countries began to include questions to measure different socioemotional skills in forms attached to their large-scale assessments. This was done to improve these skills or as an input to analyze performance results (Arias, Hincapié & Paredes, 2020).

37 According to Zea (2020), the implementation of Avelino Siñani-Elizardo Pérez Education Law No. 070 (ASEP) led education policy in Bolivia be characterized by a discourse against global trends, lower enrollment rates in private schools, and a move away from large-scale international assessments.

38 According to Flamand, Arriaga and Santizo (2020), after the 2013 reform, the INEE became the main authority in education policy assessment, with functions that overlapped those of the Ministry of Public Education–SEP. In its new role, the INEE was in charge of assessment and proposing policies to promote education and teaching quality. Teachers were deeply dissatisfied with the reform and the weight that large-scale assessments had on their performance rating. This, together with communication errors by the INEE, and discrepancies by the SEP, led to a major political revolt that ended with the dissolution of the institute.

Box 4:

Measuring socioemotional skills in Latin America

In the second decade of the century, the region started a broad debate on conceptualizing and operationalizing socioemotional skill measurement.

Until 2019, only Chile, Colombia, Mexico, and Uruguay included additional self-reports in the large-scale assessments. Although these were significant advances, there has been some lack of conceptual clarity in naming and operationalizing socioemotional skills. Countries have adopted different frameworks and concepts such as socioemotional learning, non-cognitive skills, 21st century skills, or soft skills, among others (Arias, Hincapié, & Paredes, 2020; INEEd, 2019).

In the wake of the COVID-19 pandemic, Colombia implemented changes in the socioemotional skills measurement framework to provide a general structure for its development process and to evaluate the results of the schools' efforts to train their students in socioemotional skills (ICFES, 2021). Ecuador and Peru began to measure this type of skill systematically through forms administered to students.

	Chile	Colombia	Mexico	Uruguay	Ecuador	Peru
First measurement	2013	2002*	2017	2017	2021	2021
Latest measurement	2023	2022	2019	2023	2023	2022
Framework/reform		2021		2017		
Content of the socioemotional form						
Coping with adverse situations			✓		✓	
Confidence in yourself					✓	
Self-perception	✓				✓	
Academic self-assessment and school motivation	✓			✓		
Self-efficacy				✓		✓
Self-confidence		✓				
Satisfaction with life						
Behavioral self-regulation				✓		✓
Autonomy and relationships						✓
Empathy			✓	✓		✓
Responsibility						✓
Decision making			✓			✓
Emotion recognition		✓				
Expression of emotions		✓				
Teamwork		✓				
Sense of agency			✓			
Perseverance	✓		✓	✓		
Emotional regulation			✓	✓		
Relationship skills						
Social media skills			✓			
Growth mindset			✓			
Assessment of the task				✓		
(inter or externalizing) Risky behaviors				✓		
Grades	4 and 11 (annual) 6 and 8 (every 2 years	3, 5, 9, from 7	6, 9, 12	6, 9	4 primary, 3 secondary	6, 2 secondary
Type of test	Census-based	Census-based until 2018; sample-based from 2022	Census-based / sample-based	Sample-based	Sample/ census-based	Census-based / sample-based

Assessing socioemotional skills became more frequent in the region. Colombia started using forms to collect information on students' actions and attitudes, school environment and socioemotional aspects in the Citizen Skills test. This has been administered since 2002[39] to students in grade 5 of primary and grade 9 of secondary school. The objective was to measure the students' emotional self-regulation, self motivation, school self-efficacy, and teamwork. The PLANEA learning test in Mexico assesses language and mathematics and has included a questionnaire to measure socioemotional skills since 2017. In grades 6 and 9, the questionnaire measured empathy, sense of agency, and emotional regulation in line with the socioemotional skills developed in the Plan and Curricula in Basic Education (grades 5 to 9). In grade 12, the questionnaire measured empathy, perseverance, stress management, and decision making (Arias, Hincapié, & Paredes, 2020). In Uruguay, the Aristas Assessment of Educational Achievements measures performance in reading and mathematics and has included a questionnaire to measure students' socioemotional skills since 2018.

Several countries developed and used synthetic quality indices based on test results and other system characteristics. In Colombia at the end of 2014, the Synthetic Index of Educational Quality (ISCE) was designed to provide the education community with an indicator of school quality that was easy to interpret and that would serve as input for reflecting and defining specific goals for each institution according to the education level offered (primary, lower secondary, and upper secondary). It was developed following the IDEB (2007) from Brazil. It included the learning levels of students in school, the pass and retention rates, improvements in learning, and indicators of the school environment collected in the context forms accompanying the tests. Since 2013, Chile has been measuring indicators of personal and social development (IDP) that are part of the different dimensions in "other indicators of educational quality": academic self-esteem and school motivation, participation and citizen education, and school coexistence. In Chile, school rankings have been created since 1995 following the assessment criteria that have varied over time. In 2011, the country adopted a new methodology—described in the Quality Assurance Law—for ranking schools according to their quality. The main objective of this index is to hold schools and their managers accountable for meeting learning standards and the quality of the service they provide, as well as to identify needs for targeted support. This is done by weighing student learning in the schools (tests), the pass and retention rate, improvements in learning, measures of school equity, and indicators of a school environment. While each index in Brazil, Colombia, and Chile has distinct ranking objectives, between 50 and 80% of the weight of the school quality indicators corresponds to the level and improvements in learning measured with these assessments (Elacqua, Martinez, & Westh, 2019).

In this decade, it has also become more common to publish microdata, forms, and technical test reports on the platforms of ministries or quality agencies. This practice has sought to encourage transparency and has enhanced the use of the results for research and public policy purposes. Microdata are published in ten of the eighteen countries (Argentina, Brazil, Colombia, Ecuador, Guatemala, Mexico[40], Paraguay, and Uruguay) or can be requested from the institutions' platforms (Peru and Chile).

39 Until 2006, except for 2004, and then a single edition in 2012.

40 Before the dissolution of the INEE.

Box 5:

National Learning Assessments in the Caribbean

In addition to the assessments conducted within the Caribbean Examinations Council (CXC), several Caribbean countries conduct national assessments to evaluate student learning in various areas and education levels. The following is a summary of the national assessments in the eight IDB member countries in the Caribbean.

Country	Assessment	Education level	Grade	Description
Bahamas	Grade level assessment test (GLAT)	Primary	Grades 3 and 6	It assesses mathematics and language arts. Grade 6 students are also tested in science and social studies. GLAT is administered to all government primary schools and most independent primary schools.
	Bahamas Junior Certificate (BJC)	Lower secondary	Grade 9	It assesses mastery of core subjects and is required to advance to the next cycle (upper secondary school). The exams lead to the Bahamas Junior Certificate (BJC), which is designed to measure mastery of the curriculum in subjects such as art, crafts, English language, English literature, general science, health science, home economics, mathematics, religious education, social studies, and technical drawing.
	Bahamas General Certificate of Secondary Education (BGCSE)	Upper secondary	Grade 12 (last year)	Exit exam required for university or professional studies. It is based on the UK General Certificate of Secondary Education (GCSE) and International General Certificate of Secondary Education (IGCSE) models, but it has been tailored to meet the specific needs of The Bahamas. Twenty-five subjects are currently offered.
Barbados	Barbados Secondary School Entrance Examination (BSSEE)	Primary	Grade 6	It is used to assign students to secondary schools and assess their proficiency level in various subject areas.
Belize	ATLIB	Upper secondary	Grade 12 (last year)	Exam at the end of secondary school, required for access to tertiary education.
	Belize Junior Achievement Test (BJAT)	Primary	Grade 6	It assesses skills in mathematics, English, social and natural sciences. The assessment was interrupted in 2019/2020 due to the pandemic; it has not been implemented since.
	Primary School Examinations (PSE)	Lower secondary	Grade 8	It assesses mathematics, English, social and natural sciences. The assessment was interrupted in 2019/2020 due to the pandemic; it has not been implemented since.
Jamaica	Grade One Individual Learning Profile (GOILP)	Primary	Grade 1	The GOILP is administered individually before students enter grade 1. The assessment provides initial information on each student, including their skills and proficiency level in six subjects: general knowledge, numerical concepts, spoken language, reading and writing, drawing, work habits, and classroom behavior.
	Grade 3 Diagnostic Tests (G3DT)	Primary	Grade 3	It is curriculum-based and is administered to students who have completed grade 3. It assesses language and mathematics at the end of the school year (June). It mainly aims to identify weak areas in student learning.
	Grade Four Literacy Test (GFLT) Grade Six Achievement Test (GSAT)	Primary	Grade 4	The literacy component includes the assessment of word recognition, reading comprehension, and writing. As for numeracy, students are assessed on operations and numerical representation, measurement and geometry, and algebra and statistics. One of three levels is assigned: mastery, almost-mastery, and non-mastery. Students who do not achieve proficiency in both literacy and numeracy cannot take the PEP and must retake the exams.
	Primary Exit Profile (PEP)	Primary	Grades 4, 5, and 6	It is a series of assessments that are administered throughout grades 4, 5, and 6. The first ones are GFLT and GFNT in grade 4. It aims to measure student readiness for grade 7 and to place students in secondary schools. It assesses skills in language arts and mathematics (grades 4, 5, and 6) and social studies and science (grade 5 only). The series consists of three components: performance tasks (at all three levels), ability tests, and curriculum-based tests (only for grade 6).

Country	Assessment	Education level	Grade	Description
Guyana	National Grade Two Assessment (NGTA)	Primary	Grade 2	Diagnostic assessment based on grades 1 and 2 language arts and mathematics curriculum. Five percent of the grade contributes to NGSA (National Grade Six Assessment) results.
	National Grade Four Assessment (NGFA)	Primary	Grade 4	Diagnostic assessment based on the grades 1 to 4 curriculum including literacy and numeracy standards. Ten percent of the grade contributes to NGSA results.
	National Grade Six Assessment (NGSA)	Primary	Grade 6	It is used to assign students to secondary schools. It assesses mathematics, language arts, science, and social studies. Students must have previously completed NGTA and NGFA.
	National Grade Nine Assessment (NGNA)	Lower secondary	Grade 9	Diagnostic assessment based on the grades 7 to 9 curriculum.
Trinidad and Tobago	Secondary Entrance Assessment (SEA)	Primary	Grade 5	It is administered at the end of grade 5, approximately at age 11; it is used as part of the admission process for all secondary schools.
	National Certificate of Secondary Education (NCSE)	Lower secondary	Grade 8	It is necessary for admission to upper secondary education. It is administered at the end of the third cycle in public and private schools. It covers eight subjects (mathematics, English, Spanish, science, social studies, arts, physical education, and technical studies).
Suriname	GLO 6	Primary	Grade 6	It assesses knowledge and skills at the end of primary education. Mathematics, mother tongue, social sciences, and natural sciences are evaluated. The students who get passing grades can enroll in the first academic cycle of secondary education; the others may choose technical or vocational secondary education.
Haiti	Evaluation Nationale d'apprentissages (ENA)	Primary	Grades 4 (2016-2017) and 6 (2022)	Following a pre-pilot version in 2015 in 60 schools, the first two rounds of national assessments were conducted in grade 4 in 2016 and 2017, replacing the previous grade-6 tests. These tests seek to provide information on proficiency in mathematics, Creole, and French before the end of primary school. The 2016 application was sample-based, with 526 schools nationwide; the 2017 application was census-based for public schools (1,771) and sample based for private schools (204). The 2022 assessment was conducted for grade 6 on the same subjects and a nationwide sample-based design.

Disruption and current state of learning assessments (2020-2023)

The disruption caused by the COVID-19 pandemic affected testing continuity in most countries. The gradual recovery of testing allows us to reflect on the importance of its objectives and institutional framework, the consolidation of the assessment and funding system, and the stability of its administration and use. Disruption has accelerated the debate on trends and challenges towards the future of assessment in education systems.

The COVID-19 pandemic disruption impacted test administration. Only three countries in the region continued testing in 2020. During the pandemic, the education of children and young people in the region was strongly affected; the most prolonged school closures in the world occurred in the region (Herrero et al., 2022). This disruption also affected the administration of large-scale national assessments, which were collected face-to-face on paper and pencil in almost all countries. The exception is Uruguay, which since 2013 had accumulated experience in implementing digital tests such as the 2013 National Learning test and the 2017 and 2019 Aristas digital pilots. In 2020, only Colombia and El Salvador administered census based tests in secondary schools, and Uruguay ad-

ministered the Aristas (sample-based) test in primary. Of the eighteen countries analyzed, only three conducted some type of large-scale national assessment.

Testing continuity was affected during the pandemic, but a swift recovery is evident. Globally, at least 70% of countries continued with their large-scale national assessment programs in the 2021/2022 school year (OECD, UNESCO, UNICEF, and World Bank, 2022). In Latin America, 78% of countries resumed assessment between 2021 and 2023, surpassing the global average. In primary education, eleven of the eighteen countries analyzed conducted at least one test between 2021 and 2023. However, four of the seven countries discontinued testing long before the pandemic: in Bolivia and Venezuela, the last large scale primary assessment was conducted over twenty years ago, while in Nicaragua and El Salvador, it was over eight years ago. In Uruguay, the primary Aristas tests are still within the application schedule, considering that they were last run in 2020, and the established frequency is every three years. In secondary school, fourteen of the eighteen countries analyzed conducted large-scale assessments after 2020. In the remaining four countries—Bolivia, Panama, Nicaragua, and Venezuela—these assessments have not been implemented for at least eight years (Table 2).

After the pandemic, test coverage and the levels assessed changed. Of the countries that conducted primary assessments, after the pandemic, Colombia and Paraguay switched from census- to sample based tests, while Costa Rica began to implement census-based tests in primary education. As we will discuss below, budgetary and political factors have driven these changes. Argentina went from running tests for grades 3 and 6 to only grade 6, and Chile from grades 4 and 6 to only grade 6. In secondary school, Peru's test became sample-based; Ecuador went from running the test in grades 3 and 10 in middle school to only doing so in grade 10; and the systems in El Salvador, Guatemala, and Mexico also reduced the number of grades tested, conducting assessments only in the final grade of the cycle. On the contrary, Colombia began testing for the first time in grade 7 of secondary school.

Table 2: Implementation of assessment(s) by year (1990-2023); 18 countries

Primary (ISCED 1)

	ARG	BOL	BRA	CHL	COL	CRI	DOM	ECU	GTM	HND	MEX	NIC	PAN	PRY	PER	SLV	URY	VEN
90			✓															
91																		
92				✓														
93			✓	✓														
94				✓														
95	✓		✓	✓														
96	✓			✓				✓						✓	✓		✓	
97	✓	✓	✓					✓		✓				✓				
98	✓							✓	✓	✓	✓				✓		✓	✓
99	✓		✓	✓		✓			✓	✓	✓			✓			✓	
00	✓					✓		✓	✓	✓	✓			✓				
01			✓			✓				✓	✓				✓	✓	✓	
02				✓		✓				✓	✓	✓					✓	
03	✓		✓		✓	✓				✓	✓							
04						✓			✓	✓				✓	✓			
05	✓		✓	✓	✓	✓				✓	✓		✓			✓	✓	
06				✓		✓			✓	✓	✓	✓						
07	✓		✓	✓				✓	✓	✓	✓				✓			
08				✓		✓		✓	✓	✓					✓	✓		
09			✓	✓	✓			✓	✓	✓	✓	✓			✓		✓	
10	✓			✓				✓	✓	✓	✓	✓		✓	✓			
11			✓	✓				✓		✓	✓				✓			
12				✓	✓	✓		✓		✓	✓				✓	✓		
13	✓		✓	✓	✓			✓	✓	✓	✓				✓		✓	
14			✓	✓	✓			✓	✓	✓	✓				✓			
15			✓	✓	✓			✓		✓	✓	✓		✓	✓			
16	✓		✓	✓	✓			✓		✓			✓		✓			
17	✓		✓	✓	✓		✓	✓		✓			✓				✓	
18	✓			✓			✓	✓		✓	✓		✓	✓	✓			
19			✓	✓		✓		✓	✓		✓				✓			
20																	✓	
21	✓		✓		✓	✓		✓										
22	✓			✓	✓		✓	✓	✓						✓			
23	✓		✓	✓	✓	✓	✓	✓	✓	✓				✓	✓			

Secondary (ISCED 2 and 3)

	ARG	BOL	BRA	CHL	COL	CRI	DOM	ECU	GTM	HND	MEX	NIC	PAN	PRY	PER	SLV	URY	VEN
90					✓													
91					✓		✓											
92				✓	✓		✓											
93	✓			✓	✓		✓											
94	✓			✓	✓		✓											
95	✓		✓	✓	✓	✓	✓											
96	✓			✓	✓	✓	✓	✓										
97	✓		✓	✓	✓	✓	✓	✓								✓		
98	✓			✓	✓	✓	✓	✓						✓	✓	✓		✓
99	✓		✓		✓	✓	✓									✓	✓	
00	✓			✓	✓	✓	✓	✓								✓		
01			✓	✓	✓	✓	✓								✓	✓		
02					✓	✓	✓									✓		
03	✓		✓	✓	✓	✓	✓									✓	✓	
04				✓	✓	✓	✓								✓	✓		
05	✓		✓		✓	✓	✓				✓		✓			✓		
06				✓	✓	✓	✓		✓					✓		✓		
07	✓		✓	✓	✓	✓	✓	✓	✓	✓				✓		✓		
08				✓	✓	✓	✓	✓	✓	✓	✓					✓		
09			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓		
10	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓		
11			✓	✓	✓	✓	✓	✓	✓	✓	✓					✓		
12				✓	✓	✓	✓	✓	✓	✓	✓					✓		
13	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓					✓		
14				✓	✓	✓	✓	✓	✓	✓	✓					✓		
15			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		
16	✓			✓	✓	✓	✓	✓	✓	✓					✓	✓		
17	✓		✓	✓	✓	✓		✓	✓	✓	✓					✓		
18				✓	✓	✓		✓	✓	✓				✓	✓	✓	✓	
19	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓		
20					✓											✓		
21			✓		✓	✓		✓	✓							✓		
22	✓			✓	✓		✓	✓	✓		✓				✓	✓	✓	
23			✓	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓		

Source: Own elaboration based on the information reviewed in each system.

Test objectives and stakes

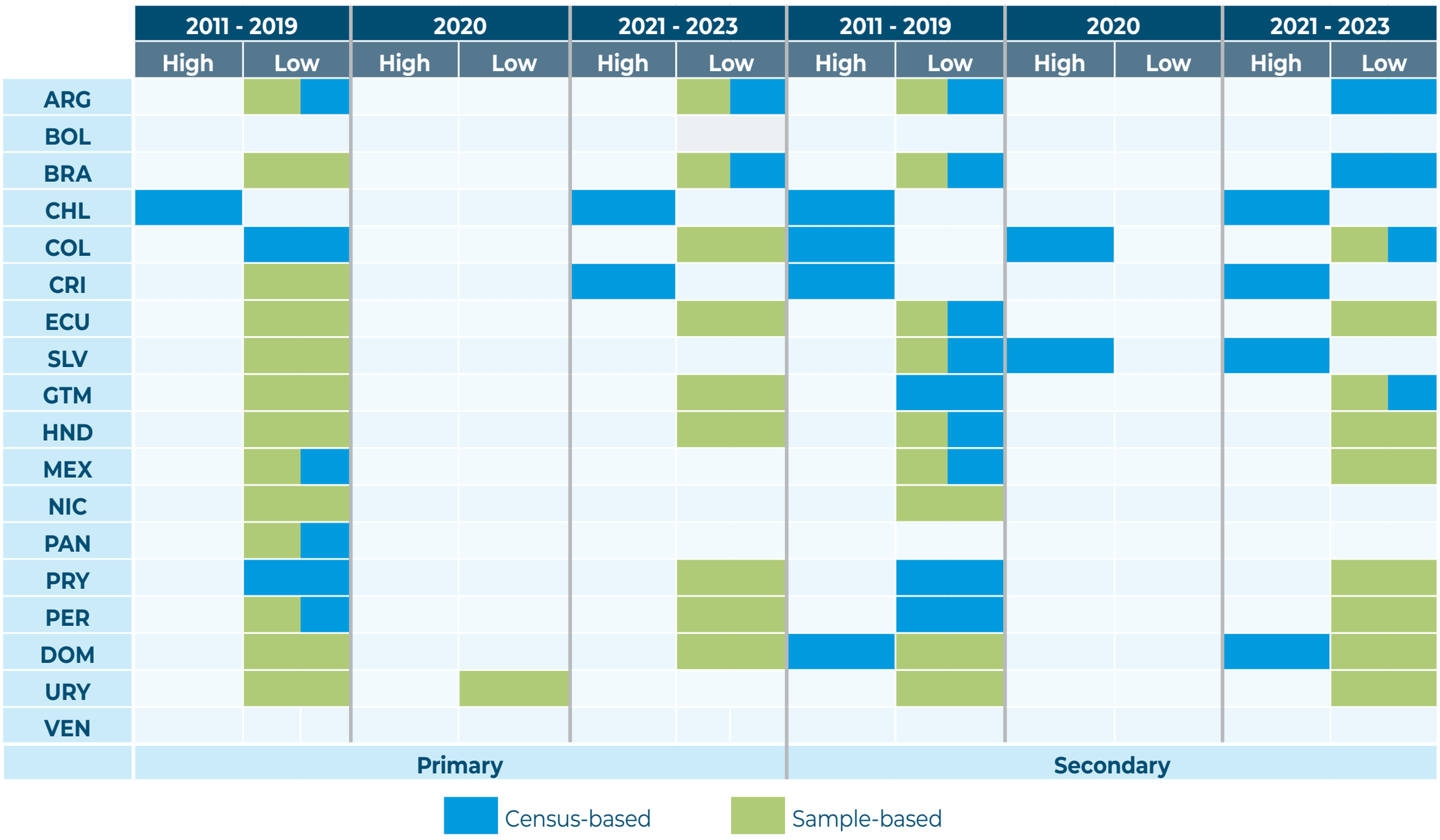
As stated above, tests can be classified according to the implications of their results. Low-stakes tests produce information for different formative purposes; they may be census or sample-based and do not have direct consequences. On the other hand, high-stakes test results are used for various purposes. For example, they can be used to set incentives for teachers and schools (Chile, Brazil, Peru, and formerly Mexico), for promoting students (Costa Rica, El Salvador, Dominican Republic, and Colombia for access to higher education) or to inform the population about student performance in education centers (Brazil, Chile, Colombia). High-stakes systems often use census-based assessment frameworks and are usually part of a system of education accountability (Cueto, 2007).

Before the pandemic, sixteen of the eighteen countries administered tests in secondary education. Of these, seven ran census-based high-stakes tests for schools or teachers (SIMCE in Chile, Prova in Brazil, and census based tests in Peru) or assessments used as certification tests for secondary students (PAES in El Salvador, Bachiller in Costa Rica, National Tests in the Dominican Republic) or as an admission test for higher education (Saber 11 in Colombia). Colombia's higher education admission exams and El Salvador's certification exams were conducted during the pandemic. Both countries adjusted their calendars, and El Salvador switched the test from paper and pencil to digital format.89[41]. In contrast, the Bachillerato test had been in effect in Costa Rica from 1988 until 2019, when the Higher Council of Education (CSE, for its Spanish acronym) (Consejo Superior de Educación) approved the National Tests for Strengthening Learning for Renewal of Opportunities (FARO, for its Spanish acronym) (Pruebas Nacionales para el Fortalecimiento de Aprendizajes para la Renovación de Oportunidades). This was done to determine the achievement level expected of students (primary, technical, and secondary) and as a requirement for promotion. Due to the pandemic and political factors, test implementation was postponed until 2021, temporarily canceled in 2022 (MEP, 2022) and resumed in 2023.

In general, one of the main characteristics of high-stakes tests is their census-based nature and stability over time. High-stakes tests are usually implemented according to a rigorous schedule, and their results are delivered to schools and students on previously announced dates in the academic calendar. Some are part of the students' academic progress, so their continuity is essential (Ferrer & Fiszbein, 2015). In El Salvador, the Avanzo tests (2021-2023) replaced the PAES tests, which had been conducted annually since 1997. Furthermore, the Saber 11 tests in Colombia have been running uninterruptedly since 1968. SIMCE tests in primary education in Chile are not high-stakes tests for students, but they have significant implications for schools and have been stable (with different timing and characteristics) since 1992.

41 In 2020, the PAES test (1997-2019) was replaced with the Avanzo test (2020-2023). Their differences included a socioemotional component in Avanzo and a digital collection form (Campos Morán, Navarrete Gálvez & Campos Solórzano 2023).

Table 3: Type of tests and stakes before, during, and after the pandemic



Source: Own elaboration based on the information reviewed in each system.

Institutional framework of learning assessment systems

As mentioned above, the institutional framework supporting the assessments (the dependence or independence of the assessment institution regarding the Ministry of Education) affects the continuity of their application, coverage, and use. This includes, among other things, the stability of technical teams, the allocation and management of human and budgetary resources, communication between assessors and information users, and the political management of the narrative surrounding assessments (Galas, Gutiérrez, & Hamilton, 2020; Ferrer & Fiszbein, 2015).

A more or less consistent difference is that independent institutions can design and implement medium and long-term assessment plans and are autonomous to disseminate their results. Non independent agencies might find it challenging to maintain long-term assessment plans due to the demands of political authorities, especially when there are administration changes. In Costa Rica, the last two administration changes have impacted the continuity of standardized tests. Between 2019 and 2020, the Bachillerato Tests, which had been in place since 1988, were eliminated and replaced with the FARO tests. After the change of administration in 2022, the FARO tests were eliminated and replaced with the National Standardized Test in 2023 (Consejo Nacional de Rectores, 2023). Political factors have affected the consolidation of national assessment systems in Venezuela and Bolivia. Bolivia only surveyed primary education in 1997, and Venezuela only surveyed both levels in 1998.

Systems with independent institutions have strengths such as a strong background testing continuity due to greater financial autonomy, which promotes sustainability and technical specialization. On the other hand, these institutions find difficulties in determining degrees of autonomy and coordination mechanisms with the Ministry of Education and other institutions, aligning tests to the curriculum, and connecting guidance and improvement strategies. In this sense, some countries have units within the ministries of

education that have been steadily consistent. This is the case in Argentina, Peru, and Guatemala, which no longer depend on budgets due to cooperation agreements with international organizations. Their assessments have become critical to the national education budget (Ferrer & Fiszbein, 2015).

Table 4: Inclusion of assessment by year and type of assessment institution (2000-2023); 18 countries

Primary (ISCED 1)

	Year	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Autonomous	BRA		✓		✓		✓		✓		✓		✓		✓	✓ |	✓	✓	✓		✓		✓		✓
	BOL																								
	COL				✓		✓				✓			✓	✓	✓	✓	✓	✓				✓	✓	✓
	CHL			✓			✓	✓	✓	✓	✓	✓	✓ |	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
	ECU	✓							✓	✓	✓	✓	✓ |	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
	URY		✓	✓			✓				✓				✓				✓ |			✓			
	PAR	✓				✓						✓					✓			✓ |					✓
Non-autonomous	ARG	✓			✓		✓		✓			✓			✓			✓	✓	✓			✓	✓	✓
	CRI	✓	✓	✓	✓	✓	✓	✓		✓				✓							✓		✓		✓
	DOM																		✓	✓				✓	✓
	HND	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓
	NIC			✓				✓			✓	✓					✓								
	VEN																								
	SLV		✓				✓			✓				✓											
	GTM	✓				✓		✓	✓	✓	✓	✓			✓	✓					✓			✓	✓
	PAN						✓											✓	✓	✓			✓ |		
	MEX	✓	✓	✓ |	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓			✓	✓				
	PER		✓			✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓

Secondary (ISCED 2 and 3)

Autonomous	BRA		✓		✓		✓		✓		✓		✓		✓		✓ |		✓		✓		✓		✓
	BOL																								
	COL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓ |	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	CHL	✓	✓		✓	✓		✓	✓	✓	✓	✓	✓ |	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
	ECU	✓							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
	URY				✓													✓ |		✓				✓	
	PAR							✓	✓			✓					✓			✓					✓
Non-autonomous	ARG	✓			✓		✓		✓			✓			✓			✓	✓		✓			✓	
	CRI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓
	DOM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓			✓	✓
	HND								✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓
	NIC										✓	✓					✓								
	VEN																								
	SLV	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	GTM							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
	PAN						✓																		
	MEX				✓ |		✓			✓	✓	✓	✓	✓	✓	✓	✓		✓		✓ |			✓	
	PER		✓			✓											✓	✓		✓	✓			✓	✓

✓ Learning Assessment | Institutional changes

Source: Own elaboration based on the information reviewed in each system.

Institutional and financial clarity of the assessments is essential to ensure assessment continuity. In Honduras, the USAID international cooperation project through the Improving Student Achievement Project in Honduras (MIDEH) ensured test continuity between 2005 and 2017. The Ministry of Education allocated a specific budget item for standardized learning assessments as of 2018. However, after the end of the USAID project in 2018, the item bank (FEREMA, 2022) was no longer updated, and national assessments resumed in 2023, but their results were no longer published. No tests were administered between 2021 and 2023 in El Salvador, Mexico, or Panama. In El Salvador, the most recent test was applied in primary education in 2012 with international funds. In Mexico, the dissolution of the INEE and the subsequent takeover of assessment functions by the National Commission for the Continuous Improvement of Education (Mejoredu) has caused delays in test continuity. In Panama, the ministry has postponed the CRECER tests for over three years. In Paraguay, funds were no longer available by the Fund for Excellence in Education and Research (FEEI, for its Spanish acronym) as of 2023; it had been funding the INEE since 2015. This has jeopardized the continuity of large-scale assessments.

The willingness and incentives to use assessment information for dissemination and research are also significant factors in promoting continuity and improvement. In Brazil, the INEP created the Committee of Educational Information Products–COMPED (1998), encompassing fifteen education institutions, to develop an articulated system to disseminate education information. The Colombian ICFES has been launching research calls since 2010 to encourage analyses of the quality of education in the country and promote the use and improvement of the information produced. All the assessment systems with autonomous institutions in the countries analyzed (except Bolivia) make the microdata collected freely available or available upon request (Chile). This fosters transparency and the use of the information for research. Argentina, Paraguay, Peru, and Mexico (on the platform of former INEE) are some of the countries with nonautonomous institutions that make this information available.

Use of digital technologies

One of the main characteristics of the tests conducted during the disruption of the pandemic in Uruguay, El Salvador, and Colombia is how these countries had been adapting to the use of digital technologies for testing or how they had adopted policies of access to digital devices. In Uruguay, the only country in the region to maintain a low-stakes test in primary education, factors specific to the test and external factors influenced its implementation. The Aristas tests are sample-based and conducted in digital format, which simplifies administration logistics and fits perfectly with the face-to-face format to create favorable conditions for implementation. In addition, external factors such as the early opening of schools (in 2020) and the target population's high rate of access to a computer during the pandemic (90%) also facilitated the implementation (Herrero et al., 2022).

In El Salvador, the AVANZO test, which replaced the PAES test, was created in digital format. By 2020, the test could be taken flexibly in a given time range when students had access to an electronic device (MINEDUCYT, 2022). After the first round of 2020, a policy was extended in 2021 to provide all high school seniors with non-returnable laptops to take the test and continue their higher education. Finally, in Colombia, changes were introduced in applying the Saber 11 exam in 2020. For example, fewer questions were included in each test (except for the socioeconomic questionnaire), and the test takers took the exam in a single session (usually there were two sessions) of five hours and thirty minutes in classrooms specially prepared to protect their health and prevent a massive spread event (ICFES, 2022). In 2020, the test was not conducted in digital format (unlike other tests conducted by the entity, such as Saber Pro for university education and Saber T&T for technical and technological vocational education). However, the institute has conducted several digital pilots for their potential digital presentation since 2021.

Formative and diagnostic assessments

Following the pandemic, there was renewed interest in the region to disseminate, promote, and institutionalize formative and diagnostic assessment. Unlike summative assessment (or assessment of learning), which focuses on the end of an education stage, formative assessment (or assessment for learning) is conducted throughout the process, providing continuous feedback to students and teachers so that the learning and teaching strategies can be adjusted and improved (UNESCO, 2021). Diagnostic assessments, which have a formative approach, are administered at the beginning of a school cycle or year to identify the students' starting point and guide the teaching and learning processes according to the students' needs. Since 2020, several countries have implemented new formative and diagnostic assessments to monitor learning (UNESCO, 2021). Honduras implemented diagnostic learning assessments only once in 2020. Since 2021, Chile has used the Comprehensive Learning Diagnosis (DIA, for its Spanish acronym) (Diagnóstico Integral de Aprendizajes) platform, which is voluntary and available to all schools countrywide, to monitor academic and socioemotional learning at three points during the school year. In 2021, Brazil launched formative diagnostic assessments for all primary school grades in public schools, and from 2022, also for secondary schools. In 2021, Guatemala implemented a diagnostic evaluation (DE), and in 2022 and 2023, the formative tests, Aprender +, were conducted in all primary and secondary grades. As of 2022, Mexico's assessment model has included annual formative diagnostic assessments from grades 2 to 6 of primary school and grades 1 to 3 of secondary school in the areas of reading, mathematics, civics and ethics through different means (online, digital system, or paper and pencil). El Salvador administered the Assessment of Productive and Citizenship Capacities (ECPYC, for its Spanish acronym) (Evaluación de Capacidades Productivas y Ciudadanas) between 2015 and 2018. From 2021, the country conducted the Knowing my Achievements assessments (Conociendo mis logros) from grade 3 of primary school to grade 2 of secondary school. These assessments aimed to encourage reflection and the implementation of reinforcement and continuous improvement plans. Finally, Uruguay has provided formative assessments annually since 2011 for teachers to use with their students in primary and secondary education through the SEA platform. Teachers can even create their own tests based on the released activities from the item bank, allowing them to administer formative assessments to their students whenever they deem it appropriate.

While most countries have adopted diagnostic and formative tests to complement national standardized tests, El Salvador, Mexico, and Guatemala have yet to resume large-scale summative national assessments for primary school after 2020, as they focus only on diagnostic and formative tests as an assessment tool for learning.

II.3 Concluding Remarks: towards an education assessment culture

As seen above, the evolution of learning assessment systems in Latin America and the Caribbean has varied, but the systems share many challenges. Having solid tools for monitoring learning continues to be a clear objective of public education policy in the region. The future debate focuses on the content, characteristics, and uses of learning tests to address the quality challenges that persist in the region. The following describes the main issues shaping these discussions based on analysis and interviews with Latin American learning assessment leaders[42].

Capacity building. National assessment systems in the region are heterogeneous in their maturity and the technical capacity of their teams, the periodicity, reliability, traceability, and comparability of their tests and the use of results. In this regard, regional assessments like ERCE not only help countries minimally meet the monitoring objective of SDG 4 but also provide an opportunity for those interested in building technical capacities and train their technical teams, developing more robust learning assessment systems. Indeed, participating in international and regional tests allows countries to take advantage of favorable spaces for technical capacity building, even if they require additional investment. If there are no national assessments or they are interrupted, these types of assessments allow countries to observe the progress of education systems, for example, by monitoring SDG 4. Therefore, regional and international assessments are essential in the region, and their continuity should be consolidated. It is crucial to highlight the lack of academic centers that measure the quality of education in the region. In this context, Chile and Brazil stand out with centers such as the Centro de Medición MIDE UC and the Centro de Políticas Públicas e Avaliação da Educação (CAEd); both have been key to developing and strengthening educational assessment systems, both nationally and regionally. The shortage of experts in psychometrics in Latin America underscores the need to invest in specialized centers and PhD programs in this area. This would significantly improve the region's technical capacity and quality of assessment systems.

Efficiency in the delivery and use of results is one of the significant challenges facing education systems. Processing test results, delivering and disseminating results reports and using them in schools—and even more so, doing this for teachers to make improvements in the classroom—remain hugely challenging and almost utopian. Some education systems have managed to minimize result processing times to publish them at least in the same school year, but this happens in few cases. One of the main objectives of learning assessments should be appropriating and using the results. In this sense, the leading entities guide their efforts to produce relevant result reports quickly.

Assessing the use of different types of tests (diagnostic, formative, and summative) to improve learning. It is healthy for systems to include different types of assessments to establish a system of checks and balances that provides information about the performance of students to educational policymakers, authorities, educational institutions, teachers in the classroom, and families and caregivers. This information comes mainly from learning assessments and associated factor questionnaires. Diagnostic and formative tests are typically conducted by teachers and speedy delivery of comparative results. In contrast, large-scale summative assessments have more rigorous technical characteristics to ensure reliability, comparability, and traceability. Formative assessments require teachers to own and deeply understand the learning objectives. A high degree of maturity in assessment skills is needed to reach this point. This may be why education systems have followed a natural path that begins with large-scale standar-

42 The interviewees were, in alphabetical order, the following: Gina Garcés, Assessment Director, Ministry of Education of Panama; Carlos Henríquez, LLECE Coordinator and former Director of the Chilean Education Quality Agency; Harvey Sánchez, CEO of the Latin American Agency for Evaluation and Public Policy and Executive Director of INEVAL in Ecuador; and Ancell Schecker, Vice Minister of Technical and Pedagogical Affairs of the Ministry of Education of the Dominican Republic.

dized assessments and undergoes result dissemination processes, including training on item development, interpretation of results, and pedagogical recommendations. Teachers appropriate the use of results, allowing them to use formative assessments administered directly with a shorter delivery time and quicker classroom interventions. This is a specific example of how both types of tests can coexist with certain precautions: regarding the testing schedule, to avoid simultaneous applications that lead to constant interruptions to the school calendar, in different grades to avoid having over-assessed populations and others without any assessment, and above all by focusing on delivering results reports to guide classroom work so teachers can use them to implement improvements as soon as possible.

Funding for national assessments depends mainly on a significant annual budget allocation as a percentage of the education budget. Therefore, when resources for education dwindle, the authorities decide to postpone or not conduct the assessments, affecting the assessment schedule and the monitoring of student performance indicators in the education system. Education authorities make information-based decisions. Therefore, it is essential to ensure compliance with the assessment implementation schedule, since evaluations are the primary input for decision making.

Quality and validity Consolidating national assessment systems requires a sustainable testing plan with technically reliable and comparable tests. When this plan is implemented and the quality and validity of the tests are confirmed, we can state that the assessment system has reached a high degree of maturity. Unfortunately, few education assessment programs have accumulated sufficient evidence to support their interpretations and uses. The lack of evidence on validity seriously threatens assessments, compromising their value and political and technical feasibility. The consolidation of assessment systems in the region has been affected by the lack of compliance with assessment plans, changes in baselines or assessment criteria, lack of funding and technically trained human resources, or lack of political will. Commitment and adherence to assessment plans can be supported by the education community when people trust the assessment system and when the community realizes that the results are helpful. In addition, multilateral organizations that sometimes fund tests or international agencies that promote capacity building can advocate for assessment systems.

Assessment of socioemotional skills. After the prolonged closure during the pandemic, the reopening of schools brought a deep discussion on how to welcome children, youth, adolescents, and teachers. It was urgent to identify the learning achievement gap. However, the socioemotional effects should have been a priority to help restore the school environment. In this sense, assessing the socioemotional components to identify remedial actions that will help restore teaching and learning environments is also a priority. In addition, the pandemic prompted other types of changes in testing components. Ecuador implemented a module to obtain information on the situation of students during the pandemic, and a socioemotional component was included. In a longitudinal study promoted by civil society and the IDB in Brazil, nine rounds of surveys were conducted with families during the pandemic and during school reopening to study the perception of families in this process. These studies were replicated in Colombia and Uruguay. Chile and Peru also focused on monitoring socioemotional components during and after school closures.

The use of technology has been widely discussed as a cost-effective alternative for countries. The review of computer-based testing experiences and their potential implementation in the future made it clear that the region needs more infrastructure and connectivity to deploy computer- and census-based tests. This implementation requires 100 percent coverage, which has yet to be achieved. Covering the entire territory would require purchasing or renting computers and transporting them around the country, which increases costs and

is inefficient. As long as this barrier is not overcome, the possibility of moving in this direction remains distant. This should not be compared to computer- and sample-based tests, which have a different structure given the size of the sample and the logistics needed. This makes them viable and merits comparative cost-effectiveness studies addressing both modalities, as processing and results delivery times could be significantly reduced.

We still have a long way to go to establish a genuine assessment culture in the region. In the post pandemic context, assessments are essential for identifying how education systems have been affected and using the results to design strategies for learning to recover. However, the disruption in the assessment schedules and their slow recovery to 2019 levels might indicate the resistance that still persists in the face of assessment rounds that show evidence of lower performance compared to previous measurements. Education systems lost several valuable months (or years) in that process, but we still have time to recover and continue building our national learning assessment systems.

Annex

Report Methodology

The data used for this study were collected between September 2023 and March 2024. The methodology was primarily based on searching and reviewing public documents and official communications from education authorities and assessment agencies and on focused interviews with experts who have led education assessment departments or institutions in Ecuador, Panama, the Dominican Republic, and Chile.

The assessments considered for each country in the analysis include large-scale summative learning tests conducted by eighteen Latin American countries between 1990 and 2023 in primary and secondary education. The report focuses on assessment for monitoring purposes of high-or low stakes tests, and examinations to certify attending a course or education level (primary or secondary) or for selection or admission to higher education, provided that this test is part of the primary or secondary curriculum.

The assessments analyzed are detailed in the following table.

Country	Name(s) of test(s)
Argentina	Operativo Nacional de Evaluación ONE, Aprender
Bolivia	Sistema de Medición y Evaluación de la Calidad de la Educación SIMECAL
Brazil	SAEB, Aneb, Anresc, ANA
Chile	Sistema nacional de evaluación de resultados de aprendizaje SIMCE
Colombia	Pruebas SABER 11, 9, 7, 5, 3
Costa Rica	Pruebas Nacionales Diagnósticas MEP, Bachillerato, III Ciclo, FARO
Ecuador	Aprendo, SER
El Salvador	Prueba de Acceso a la Educación Superior PAES, Avanzo
Guatemala	Evaluación del Rendimiento Escolar PRONERE, Evaluación a Tercero Básico TER
Honduras	Evaluaciones Rendimiento Académico
Mexico	Estándares Nacionales, Enlace, Exámenes de la Calidad y el Logro Educativos (EXCALE), PLANEA
Nicaragua	Sistema Nacional de Evaluación de la Educación Básica y Media, Evaluación Nacional
Panama	Sistema Nacional de Evaluación de la Calidad SINECA, Crecer
Paraguay	SNEPE, Sistema Nacional de Evaluación del Proceso Educativo, ENLACE
Peru	Evaluación Muestral de Estudiantes, Evaluación Censal de Estudiantes
Dominican Republic	Sistema de Pruebas Nacionales, pruebas diagnósticas
Uruguay	Evaluación Nacional de Aprendizajes, Aristas
Venezuela	SINEA, Sistema Nacional de Medición y Evaluación del Aprendizaje

References

Altinok, N., Angrist, N. y Patrinos, H. A. (2018). Global data set on education quality (1965-2015). *World Bank Policy Research Working Paper,* (8314).

Arias Ortiz, E., Bos, M. S., Giambruno, C. y Zoido, P. (2023a). *América Latina y el Caribe en PISA 2022: ¿cuántos tienen bajo desempeño?* CIMA, Banco Interamericano de Desarrollo. https://doi.org/10.18235/0005316

Arias Ortiz, E., Bos, M. M. S., Giambruno, C. y Zoido, P. (2023b). *América Latina y el Caribe en PISA 2022:* ¿cómo le fue a la región? CIMA, Banco Interamericano de Desarrollo. https://doi.org/10.18235/0005318

Arias Ortiz, E., Giambruno, C., Morduchowicz, A. y Pineda, B. (2023c). *El estado de la educación en América Latina y el Caribe 2023*. Banco Interamericano de Desarrollo. https://doi.org/10.18235/0005515

Arias Ortiz, E., Bos, M. S., Chen Peraza, J., Giambruno, C., Levin, V., Oubiña, V., Pineda, J. A. y Zoido, P. (2024). *El aprendizaje no puede esperar: Lecciones para América Latina y el Caribe a partir de PISA 2022*. Banco Interamericano de Desarrollo y Banco Mundial. https://doi.org/10.18235/0005671

Arias, E., Hincapié, D. y Paredes, D. (2020). *Educar para la vida: El desarrollo de las habilidades socioemocionales y el rol de los docentes*. Banco Interamericano de Desarrollo.

Arregui, P. (2008). *Uso de los resultados de evaluaciones educativas a gran escala en América Latina*. PREAL. Grupo de Trabajo sobre Estándares y Evaluación.

Asamblea General de las Naciones Unidas. (2015). *Resolución aprobada por la Asamblea General el 25 de septiembre de 2015 [sin remisión previa a una Comisión Principal (A/70/L.1)] (Documento oficial No. 70/1). Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible*. Recuperado de https://unctad.org/system/files/official-document/ares70d1_es.pdf

Ascorra, P., Carrasco, C., López, V. y Morales, M. (2019). *Políticas de convivencia escolar en tiempos de rendición de cuentas*. Education Policy Analysis Archives, 1-24.

Balarin, M. (2021). *The Political Economy of Education Reforms in Peru 1995-2020*. GRADE.

Banco Mundial. (2018). *Using student assessments to improve instruction and learning*. World Bank.

Blazer, C. (2011). *Unintended Consequences of High-stakes Testing*. Information Capsule.

Borger, M., Elacqua, G., Jacas, I., Neilson, C. y Olsen, A. S. W. (2024). Report cards: Parental preferences, information and school choice in Haiti. *Economics of Education Review, 102,* 102560.

Brooke, N. (2016). High-stakes accountability using teacher salary incentives in Brazil: An update. *Revista de Currículum y Formación de Profesorado*, 207-250.

Bruns, B., Akmal, M. y Birdsall, N. (2019). *The Political Economy of Testing in Latin America and Sub-Saharan Africa*. RISE.

Cáceres, C., Muñoz, C. y Valenzuela, J. (2021). *Responsabilidad personal docente y motivación escolar. Revista Electrónica Interuniversitaria de Formación del Profesorado*, 175-188.

Campos Morán, S., Navarrete Gálvez, P. M. y Campos Solórzano, M. R. (2023). *Evaluación estandarizada de competencias en El Salvador: Legado del modelo PAES y perspectivas de la prueba AVANZO*. Entorno, 50-59.

Caribbean Examinations Council. (s.f.). Caribbean Examinations Council. https://www.cxc.org/about/about-cxc/

Caribbean Examinations Council (2023). Caribbean Examinations Council: Annual Report 2022. Disponible en: https://www.cxc.org/cxc-annual-report-2022/

Cariola, L., Covacevich, C., Grubler, J., Lagos, E. y Oritz, M. (2011). *Chilean participation in IEA studies. In IEA 1958-2008: 50 years of experiences and memories* (Vol. 2008, p. 50).

Centro de Información para la Mejora de los Aprendizajes - CIMA. (n.d.). *Banco Interamericano de Desarrollo*. https://cima.iadb.org/es

Clarke, M. y Lune-Bazaldua, D. (2021). P*rimer on Large-Scale Assessments of Educational Achievement*. World Bank.

Consejo Nacional de Rectores. (2023). *Noveno Estado de la Educación 2023*. Programa Estado de la Nación.

Cox, C. y Meckes, L. (2016). International large-scale assessment studies and educational policy-making in Chile: contexts and dimensions of influence. *Research Papers in Education, 31*(5), 502-515.

Cueto, S. (2007). Las evaluaciones nacionales e internacionales de rendimiento escolar en el Perú: balance y perspectivas. En *Investigación, políticas y desarrollo en el Perú* (pp. 405-455). GRADE.

De Hoyos, R., Ganimian, A. y Holland, P. (2019). *Teaching with the Test: Experimental Evidence on Diagnostic Feedback and Capacity-Building for Schools in Argentina*. World Bank, 123-140.

Dueñas, X., Elacqua, G., Jaramillo, L. y Margitic, J. F. (2022). *Pérdidas de aprendizaje debido al COVID-19. Desafíos y hallazgos en Colombia*. BID.

Elacqua, G., Martinez, M. y Westh, A. S. (2019). *Diseño de índices de calidad escolar: Lecciones de la experiencia internacional*. BID.

Elacqua, G. y Fabrega, R. (2004). *El consumidor de la educación: El actor olvidado de la libre elección de escuelas en Chile*. Santiago.

FEREMA. (2022). *Progreso Educativo de Honduras* (IPEH) 2022. FEREMA.

Fernandes, R. y Gremaud, A. P. (2020). Large-scale educational assessment and accountability: a brief analysis of the Brazilian experience.

Ferrer, G. (2006). *Educational Assessment Systems in Latin America: Current Practice and Future Challenges*. Partnership for Educational Revitalization in the Americas (PREAL).

Ferrer, G. y Fiszbein, A. (2015). *¿Qué ha sucedido con los sistemas de evaluación de aprendizajes en América Latina?* Washington DC: El Diálogo.

Flamand, L., Arriaga, R. y Santizo, C. (2020). *Reforma educativa y políticas de evaluación en México. ¿Instrumentos para abatir el rezago escolar y promover la igualdad de oportunidades? Revista Foro Internacional (Foro Internacional FI)*, 717-753.

Galas, C., Gutiérrez, M. T. y Hamilton, M. E. (2020). *Uso de resultados de las pruebas de aprendizaje en el diseño de las políticas educativas en América Latina*. UNESCO, Buenos Aires.

Gordon, E. D. (2019). P*roblems, pressures and policies affecting the progress of the Caribbean Examinations Council examinations: A postcolonial response to secondary education in Jamaica* (Doctoral dissertation, UCL (University College London)).

Harris, A. (1991). *Effective Assessment of Educational Progress: A Review of Strategies for Measuring Learning Achievement*. World Bank.

Herrero Tejada, A., López Leavy, M., Finoli, M., Hevia de la Jara, F., Stanton, S. y Fiszbein, A. (2022). *Evaluación en pandemia: ¿Cómo diagnosticamos las pérdidas de aprendizajes para informar los esfuerzos de recuperación?* Diálogo Interamericano

Heyneman, S. P. y Lee, B. (2014). The impact of international studies of academic achievement on policy and research. *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis*, 37-72.

Instituto Colombiano para la Evaluación de la Educación - ICFES. (2017). *Experiencia acumulada en el levantamiento de datos de factores asociados, 2012-2015 SABER 3, 5 y 9*. ICFES, Bogotá.

Instituto Colombiano para la Evaluación de la Educación - ICFES. (2021). *Habilidades Socioemocionales. Marco de referencia para la evaluación ICFES*. ICFES, Bogotá.

Instituto Colombiano para la Evaluación de la Educación - ICFES. (2022). *Informe nacional de resultados del examen Saber 11° 2020*. ICFES, Bogotá.

Keeves, J. P. (1995). *The world of school learning: selected key findings from 35 years of IEA research*. International association for the evaluation of educational achievement (IEA).

Martínez, F. (2015). *Las pruebas ENLACE y EXCALE, un estudio de validación*. INEE, Ciudad de México.

Martínez-Iñiguez, J., Tobón, S., López-Ramírez, E. y Manzanilla-Granados, H. M. (2020). *Calidad educativa: un estudio documental desde una perspectiva socioformativa. Revista Latinoamericana de Estudios Educativos*, 233-258.

Ministerio de Educación - MINEDUCA. (2014). *Hacia un sistema completo y equilibrado de evaluación de los aprendizajes en Chile. Informe Equipo de Tarea para la Revisión del SIMCE*. MINEDUCA, Santiago.

Ministerio de Educación Pública - MEP. (2022). *Informe Nacional Pruebas Nacionales FARO: Primaria – Secundaria, II semestre-2021*. Ministerio de Educación Pública, San José.

Ministerio de Educación, Ciencia y Tecnología - MINEDUCYT. (2022). *Informe de Resultados Avanzo*. MINEDUCYT, San Salvador.

Montoya, S. (2023). Reporting learning outcomes in basic education: Country's options for indicator 4.1. 1. *UNESCO Institute for Statistics. Montreal, Canada*. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2022/08/Countrys-reporting-option-_Zambia_AAEA.Final_.pdf

Murillo, F. J. (2007). Resultados de aprendizaje en América Latina a partir de las evaluaciones nacionales. UNESCO/OREALC.

OCDE, UNESCO, UNICEF y Banco Mundial. (2022). *From Learning Recovery to Education Transformation: Insights and Reflections from the 4th Survey on National Education Responses to COVID-19 School Closures*. OCDE, UNESCO, UNICEF, Banco Mundial, París, Montreal, New York, Washington D. C.

OCDE. (2018). PISA for development assessment and analytical framework: Reading, mathematics and science. *OECD Publishing*.

OCDE. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

OCDE. (2021). *National Assessment Reform: Core Considerations for Brazil*. OECD Education Policy Perspectives, París.

Pérez, G. y Maldonado, C. (2015). *Panorama de los sistemas de monitoreo y evaluación en América Latina*. CIDE, México.

Pérez, G. y Serrano, S. (2016). *Impacto de la evaluación estandarizada en los sistemas educativos en América Latina*. Banco Interamericano de Desarrollo.

Ramirez, F. O., Schofer, E. y Meyer, J.W. (2018). "International Tests, National Assessments, and Educational Development (1970–2012)". Comparative Education Review, vol. 62, núm. 3, págs. 344 a 364.

Rivas, A. (2015). *América Latina después de PISA: Lecciones aprendidas de la educación en siete países (2000-2015)*. Fundación Cippec.

Rodríguez-Garcés, C., Padilla-Fuentes, G. y Suazo Ruiz, C. (2020). Medición de calidad educativa en Chile: Lo que reportan los indicadores de desarrollo cognitivo, personal y social en la escuela. *Revista Pilquen - Sección Psicopedagogía*, 34-48.

Sánchez, M. y del Sagrario, F. (2013). Las evaluaciones estandarizadas: Sus efectos en tres países latinoamericanos. *Revista Latinoamericana de Estudios Educativos (México)*, 97-124.

Saracho Martínez, A. (2007). *Casos de Estudio de Reacciones a PISA 2000 y 2003: Alemania, Polonia, Brasil y Otros*. [Case Study of Reactions to PISA 2000 and 2003: Germany, Poland, Brazil and Others]. México, DF: Fundación IDEA.

Teig, N. y Steinmann, I. (2023). Leveraging large-scale assessments for effective and equitable school practices: The case of the Nordic countries. *Large-scale Assessments in Education*, 21.

Thailinger, A., Pecha, C., Beuermann, D., Ortiz, E. A., Hobbs, C. y Piras, C. (2023). Gender Gaps in the English-speaking Caribbean: Education, Skills, and Wages. InterAmerican Development Bank.

Tobin, M., Nugroho, D. y Lietz, P. (2016). Large-scale assessments of students' learning and education policy: synthesising evidence across world regions. R*esearch Papers in Education, 31(*5), 578-594.

UNESCO. (2019). La promesa de las evaluaciones de aprendizaje a gran escala. Reconocer los límites para generar oportunidades. UNESCO.

UNESCO. (2021). *Evaluación formativa: Una oportunidad para transformar la educación en tiempos de pandemia*. UNESCO.

UNESCO Institute for Statistics. (2024). GAML/TCG criteria for use of an assessment to report on SDG 4.1. 1. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2024/02/GAML-Criteria-for-reporting.pdf

UNESCO Institute for Statistics. (2013). *Clasificación Internacional Normalizada de la Educación CINE 2011*. UNESCO-UIS.

UNESCO Institute for Statistics. (2022). Metadata for Indicator 1.a.1 - Education expenditures per capita as a percentage of GDP. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2021/09/Metadata-FFA-1.a.gdp_.pdf

Vanni, X. y Valenzuela, J. P. B. (2020). Evaluación del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. *Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura*.

Vázquez Olivera, M. G. (2015). La calidad de la educación. Reformas educativas y control social en América Latina. *Revista de Estudios Latinoamericanos*, 93-124.

Vegas, E. y Petrow, J. (2008). *Learning in Latin America: The challenge for the 21st century*. Banco Mundial.

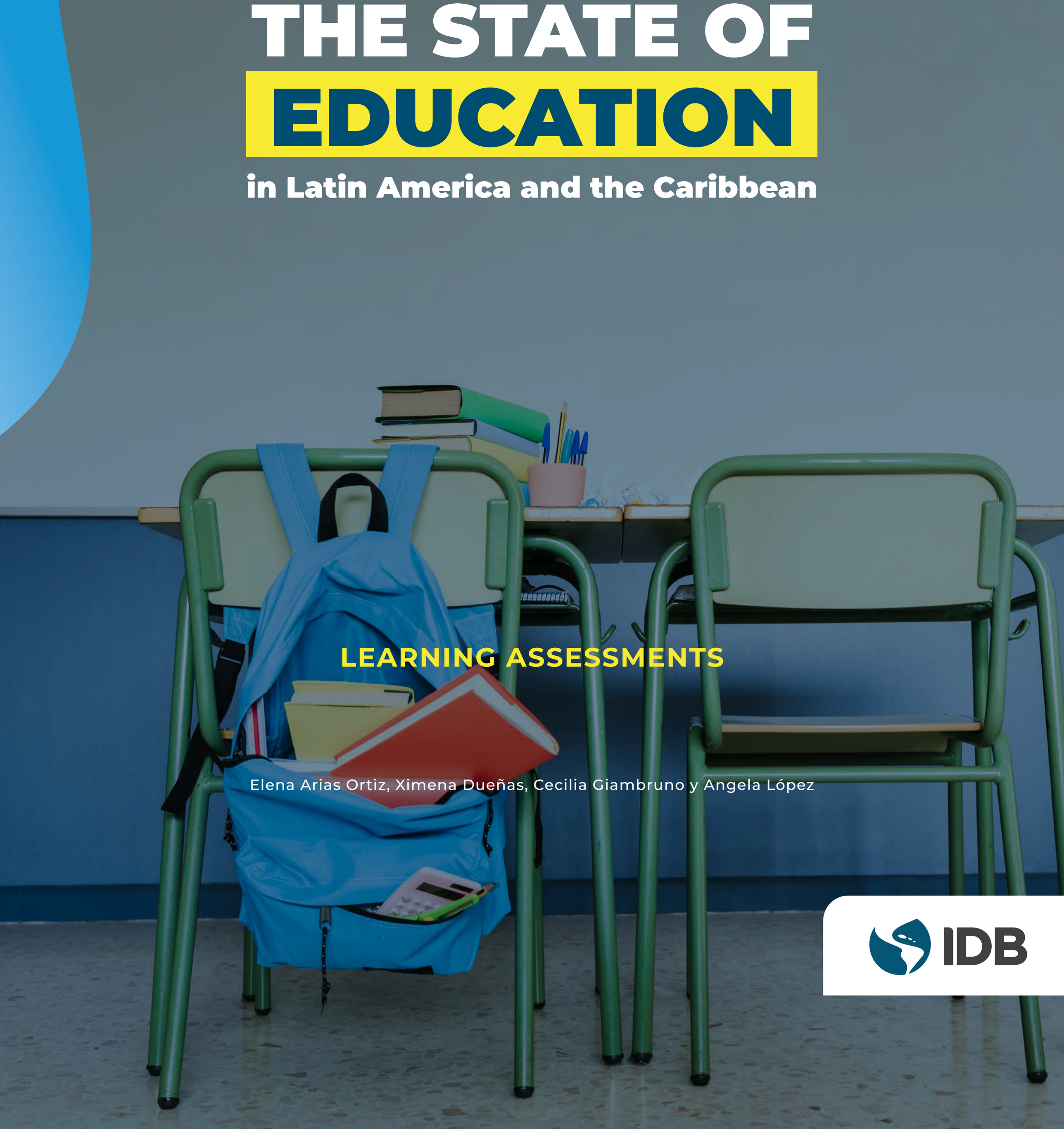
Wagemaker, H. (2014). International large-scale assessments: From research to policy. *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, 11-36.

Zea, M. (2020). La evaluación de la calidad educativa en el paradigma de la educación para vivir en Bolivia. *Journal of Supranational Policies of Education*, 106-126.

2024

THE STATE OF EDUCATION

in Latin America and the Caribbean



LEARNING ASSESSMENTS

Elena Arias Ortiz, Ximena Dueñas, Cecilia Giambruno y Angela López

IDB