EDUCAR EM REVISTA
Universidade Federal do Paraná | Setor de Educação
ISSN 1984-0411

https://doi.org/10.1590/1984-0411.94031-T

**DOSSIER**

*Quality, learning and systemic assessment: discourses from international organizations for Latin American countries*

# Large-scale assessments in Latin America: Problematizing the quality of school education

# *Avaliações em larga escala na América Latina: problematizando a qualidade da educação escolar*

**Cristiane Machado**[a]
crimacha@unicamp.br

**Ocimar Munhoz Alavarse**[b]
ocimar@usp.br

**Raíssa de Oliveira Chappaz**[c]
raissa.chappaz@gmail.com

## ABSTRACT

This article aims to characterize the large-scale assessments implemented in Latin America, of which Brazil is a participant, problematizing the quality that such assessment tools inflict on school education. This is a documentary research based on information available on the website of the National Institute of Studies and Educational Research Anísio Teixeira (INEP). The following assessments are described and analyzed: Comparative and Explanatory Regional Study (ERCE); Progress in Reading International Study (PIRLS); Program for International Student Assessment (PISA); and Trends in International Mathematics and Science Study (TIMSS). We conclude that the quality of school education is central to these assessments, generally translated into scores obtained by students' performance on standardized tests, forming a notion of quality almost exclusively associated with the results of school learning, even though contextual questionnaires accompany these assessments. Furthermore, the characterization of large-scale assessment initiatives present in Latin America reveals the increasingly robust interest of international organizations in creating instruments capable of measuring student learning, enabling the comparison of the quality of school education between and within educational systems.

*Keywords:* Large-Scale Assessment. Latin America. Quality of School Education. International Assessments.

## RESUMO

O objetivo deste artigo é caracterizar as avaliações em larga escala presentes na América Latina, das quais o Brasil é participante, problematizando a noção de qualidade que elas imprimem para a educação escolar. Trata-se de um estudo documental que tomou por base as informações disponíveis no site do Instituto

[a] Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil.

[b] Universidade de São Paulo (USP), São Paulo, São Paulo, Brazil.

[c] Instituto Federal de São Paulo (IFSP), São Paulo, São Paulo, Brazil.

Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). São descritas e analisadas as avaliações: Estudo Regional Comparativo e Explicativo (ERCE); Estudo Internacional de Progresso em Leitura (PIRLS); Programa Internacional de Avaliação de Estudantes (Programme for International Student Assessment – PISA) e Estudo Internacional de Tendências em Matemática e Ciências (Trends in International Mathematics and Science Study – TIMSS). Conclui-se que a qualidade da educação escolar tem centralidade nessas avaliações, traduzida, em geral, em escores obtidos pelo desempenho de estudantes em testes padronizados, conformando uma noção de qualidade quase que exclusivamente associada a resultados das aprendizagens escolares, ainda que questionários contextuais acompanhem essas aferições. Ademais, a caracterização das iniciativas de avaliação em larga escala presentes na América Latina revela o interesse cada vez mais robusto de organismos internacionais em criar instrumentos capazes de aferir a aprendizagem de estudantes, possibilitando a comparação da qualidade da educação escolar entre e dentro dos sistemas educativos.

*Palavras-chave:* Avaliação em Larga Escala. América Latina. Qualidade da Educação Escolar. Avaliações Internacionais.

## Introduction

In Latin America, several large-scale international assessment initiatives have gained support from countries on the continent since the 1990s, despite differences between them, reflecting the movement to consolidate this assessment format in the educational policy scenario, as an expression of a phenomenon on a global scale, as noted, among several studies, by Maroy and Voisin (2013), Verger, Fontdevila and Parcerisa (2019) and Verger and Normand (2015).

In this context, large-scale international assessments have become sources of reference in the search for evidence about students learning in Latin America, with information used for comparisons regarding educational parameters, including associations with the development of these countries. This process was triggered, with differences between and within the nations of the continent, based on the argument that these potential results, especially those of standardized tests, would have to guide educational policies, enabling direct specific interventions in the issues shown both in these results and the contextual data, usually obtained in jointly administered questionnaires.

Following and strengthening the trend that began at the Jomtien Conference in 1990, which valued school education as a way of boosting countries' economic development and creating instruments capable of measuring the effectiveness of this education, in general, improving quality is the justification and, at the same time, the promise of large-scale assessment initiatives. This is the context where these initiatives emerge worldwide, with nuances and particularities specific to the places, cultures, and responsible organizations, but also with a common axis that relates standardized student tests as an expression of the quality of education.

In this article, focusing on Basic Education, we sought to characterize and analyze four large-scale assessments implemented in Latin America and, simultaneously, in Brazil, and to problematize the notion of quality that they convey to school education.

The study, supported by documentary research, was based on consultation of documents available on the website of the National Institute of Studies and Educational Research Anísio Teixeira (INEP) examining the explanations, mainly, about the characteristics and notion of assessment quality: Comparative and Explanatory Regional Study (ERCE); Progress in Reading Study (PIRLS); Program

for International Student Assessment (PISA); and Trends in International Mathematics and Science Study (TIMSS). The information obtained was systematized and categorized to allow comparative analysis of the general characteristics of these four large-scale assessments concerning the year in which they began to be applied, frequency and number of participating countries; beginning of Brazil's participation; grades/years assessed; quantitative coverage of students; applied instruments and purposes.

It is worth noting that, in addition to participating in these four international assessments, since 1990 the National Institute of Studies and Educational Research Anísio Teixeira (INEP) has been responsible for the formulating, applying and systematizing the Basic Education Assessment System (SAEB), a large-scale national sample assessment that, in 2005, was divided into a census named Prova Brasil (Machado; Alavarse, 2014).

This article is organized into three parts, in addition to this introduction. Initially, the large-scale assessments that circulate throughout Latin America are described and analyzed; then, the notion of quality that structures these assessments is problematized; and, finally, final reflections are made as a way of enticing new questions about different aspects of the relationship between large-scale assessment and the quality of school education.

## Large-scale evaluation initiatives in Latin America: characteristics and scope

According to INEP's website[1], the Brazilian agency responsible for planning, organizing and operationalizing international initiatives in the national territory, four assessments were identified in Latin America with a focus on Basic Education in Brazil: the Comparative and Explanatory Regional Study (ERCE)[2]; the Progress in Reading Study (PIRLS); the Program for International Student Assessment (PISA); and the Trends in International Mathematics and Science Study (TIMSS).

In publicizing its responsibilities concerning these international assessment initiatives, INEP performs planning and operational actions, all highlighted in a preamble text on INEP's website. In the case of PISA and PIRLS, the actions performed are specified, namely: representing Brazil before the entity responsible for the large-scale international assessment initiative; coordinating the translation of the applied instruments; coordinating the application of the instruments in participating schools and the collection and coding of participants' responses; analyzing or contributing to the analysis of the results; and preparing the national report. Furthermore, in the case of PIRLS, INEP also collaborates , selecting the participant sample in the application.

The further steps and actions involved in the preparation and implementation of a large-scale international assessment in the country are carried out by the international entities responsible for each of the initiatives. In the case of ERCE, the entity is the Latin American Laboratory for the

[1] The information characterizing and describing the international large-scale assessment initiatives cited in quotation marks in this topic of the article comes from INEP, as well as those that make up the tables presented (Brasil, s.d.).

[2] ERCE underwent a change in its acronym, so that until 2019 a nomenclature was incorporated into its name that indicated the edition number – Perce, Serce and Terce –, but from the fourth edition onwards the Study began to be identified by the acronym ERCE followed by the year of application.

Assessment of the Quality of Education (LLECE), linked to the Regional Office of Education for Latin America and the Caribbean (OREALC/UNESCO Santiago); for PISA, it is the Organization for Economic Cooperation and Development (OECD); for PIRLS and TIMSS, it is the International Association for the Assessment of Educational Performance (IEA), identified as an "international cooperative of national research institutions, academics and analysts".

ERCE, PIRLS, PISA, and TIMSS have different definitions that identify them, also delimiting different stated objectives. Considering the chronological order in which they were formulated, TIMSS has existed since 1995 and is defined as an "international study of trends" that aims to "provide data on performance" in the areas of knowledge of Mathematics and Sciences". ERCE, formulated in 1997 and identified as an "international assessment aimed at countries in Latin America and the Caribbean", has the stated objective of "monitoring progress in the learning of students in the region".

PISA and PIRLS were both developed in the 2000s, with PISA being applied first and identified as an "international comparative study" that aims to "reveal whether educational systems are becoming more or less efficient in preparing students to continue their studies or for the job market", thus seeking to "translate the results of the assessment into thousands of data fragments that, when gathered, reveal the representation of an efficient educational system", to "encourage countries to learn from each other's experiences in building more fair and inclusive educational systems". PIRLS, which emerged in 2001, is identified only as "an initiative" carried out by an international cooperative that works to "evaluate, understand and improve education around the world", with the declared objective of "analyzing trends in reading comprehension, in addition to collecting information on learning contexts, to characterize the reading process of the students assessed in the countries participating in the study".

Thus, with different focuses and scopes, and considering the potential and diffusion of these initiatives in Latin America, it is important to characterize them also with aspects such as purpose, periodicity, participating countries, series/years evaluated, scope, instruments, and target measure, to understand differences and points of similarity. To this end, the document base adopted as a reference is both the information available on INEP's website and also documents provided by this agency, which is responsible for the processes involving the implementation of these large-scale international evaluations in Brazil.

Table 1 covers information from ERCE, PIRLS, PISA and TIMSS regarding the start year of application, the scope of participating countries, Brazil's participation, grades/years evaluated, scope in Brazil, and instruments applied.

**Table 1:** General characteristics of international large-scale assessment initiatives: ERCE, PIRLS, PISA, and TIMSS

| | TIMSS | ERCE | PISA | PIRLS |
|---|---|---|---|---|
| **Starting year** | 1995 | 1997 | 2000 | 2001 |
| **Periodicity** | Every 4 years | 1997; 2006; 2013; 2019 | Every 3 years | Every 5 years |
| **Participating countries** | "About 70 countries use the trend data [...] more countries join the study with each assessment cycle." | 1997: 13 countries<br>2006: 16 countries<br>2013: 15 countries<br>2019: 18 countries | "Since its first edition in 2000, the number of participating countries and economies has increased with each cycle." | 2021: 61 countries |
| **Brazil's participation** | Since 2023 | Since 1997 | Since 2000 | Since 2021 |
| **series/grades evaluated** | Grades 4 and 8 of elementary school | 1997: grades 4 and 5 of elementary school<br>2006: schools series equivalent to grades 4 and 7 of elementary school.<br>2013: same as 2006<br>2019: same as 2006 | "Students aged 15, the age at which compulsory basic education is assumed to have ended in most countries."<br>2000: 200,000 students<br>2003: 250,000 students (Brazil: 4,452)<br>2006: 400,000 students (Brazil: 9,345)<br>2009: 470,000 students (Brazil: 20,127)<br>2018: 600,000 students (Brazil: 10,691) | Grade 4 of elementary school. |
| **Range*** | | 1997: 55 thousand students<br>2006: 196 thousand students<br>2013: 195.752 students | "Students aged 15, the age at which compulsory basic education is assumed to have ended in most countries."<br>2000: 200,000 students<br>2003: 250,000 students (Brazil: 4,452)<br>2006: 400,000 students (Brazil: 9,345)<br>2009: 470,000 students (Brazil: 20,127)<br>2018: 600,000 students (Brazil: 10,691) | 2021: approximately 400 thousand students |
| **Applied instruments** | Cognitive tests and contextual questionnaires | Cognitive tests and contextual questionnaires | Cognitive tests and contextual questionnaires | Cognitive tests and contextual questionnaires |

**Source:** Systematization made by the authors based on information in Brasil (s.d.).
* Total values correspond to the total number of students in the participating countries.

Although quantitative data on the scope of participating countries, students and schools in all editions carried out was not identified, Table 1 shows consolidated initiatives, with a history of applications involving a significant number of countries. Brazil has participated in both ERCE and PISA, since the very first applications in all of those editions carried out to date.

In the case of TIMSS, despite being the oldest international assessment initiative focusing on Basic Education, Brazil only began participating in the last application, carried out in 2023. The same occurred with PIRLS, which is the most recent formulated initiative, with Brazil participating

only after 20 years of its existence. It is worth considering that TIMSS and PIRLS have more specific focuses on the assessment object, respectively "content and cognitive domains in mathematics and science" and "reading skills", which may indicate a priority given to other initiatives with broader assessment objects and therefore the later adherence to these assessments.

Regarding the grades/years assessed, it is possible to identify that the initiatives prioritize the final years of Elementary School I and II. PISA, for example, is more focused on those who have completed the second stage of Elementary School, since its target group is 15-year-old students who, due to their age, should already be in High School. Regarding the scope of participating countries, students and schools involved in each application, data from all years was not found. Analyzing the data from ERCE, an initiative that has the most complete information published on INEP›s website, it is observed that, in general, there was in fact a quantitative growth in the number of countries and students participating from 1997 to 2013. Remainder tests also claim that there is an increase in the number of countries, which cannot be validated by the data available on the INEP website.

Furthermore, a characteristic present in all these initiatives is the application of both tests, named in some cases, as in ERCE and PIRLS, as "cognitive tests", and the application of contextual questionnaires, which are aimed at different actors in the school community and which seek to identify contextual data, understanding factors that are associated with and impact student learning.

Thus, all assessment initiatives discussed here apply these questionnaires to students, families, teachers and principals, only PISA highlights the questionnaires distributed to parents/guardians and teachers as optional. Regarding the topics covered in the questionnaires, shown in Table 2, PISA and PIRLS highlight the existence of specific reference matrices. In general, they exemplify as addressed factors the aspects of learning contexts and access to educational opportunities related to three spheres, the student, the school and classroom, and the family environment. PIRLS covers, in addition to these aforementioned factors, the national scope.

**Table 2:** Aspects covered in the contextual questionnaires

| | |
|---|---|
| **TIMSS** | School factors associated with learning<br>Classroom factors associated with learning<br>Home factors associated with learning |
| **ERCE** | Students' sense of belonging to school<br>Organization of teaching<br>Socioeconomic status of the family<br>Violence in the neighborhood<br>Parent involvement in learning activities at home<br>Parent participation in school |
| **PISA** | Learning approaches<br>Learning environments<br>Students' family background |
| **PIRLS** | Student characteristics<br>School context<br>Classroom context<br>Home context<br>National context |

**Source:** Systematization made by the authors based on information in Brasil (s.d).

Regarding the purposes established for each of the large-scale international assessment initiatives and what they measure, Table 3 summarizes this information. It is worth noting that two of the initiatives, TIMSS and ERCE, include monitoring as a purpose; the first defines "monitoring the effectiveness of their educational systems", while the second highlights quality, indicating monitoring the quality of education as a means of obtaining information and thus guiding decision-making. PISA and PIRLS are more comprehensive and include more than one action in terms of their purpose. PISA defines the purpose of "evaluating" students' knowledge and skills and "learning" from policies and practices in other countries in order to "formulate" policies and programs that are focused on improving the quality of education and equity of results. PIRLS, on the other hand, includes the purpose of "analyzing trends" related to what is its object of assessment and "collecting information" related to the learning context. It is worth noting that in none of the cases in which quality appears mentioned in what constitutes the purpose of the initiatives is it possible to identify what constitutes such quality. For example, factors associated with learning, which are collected in the contextual questionnaires applied in all initiatives, are not mentioned.

**Table 3:** Purposes and evaluation objects of international large-scale evaluation initiatives

| | **purpose** | **object of evaluation** |
|---|---|---|
| TIMSS | "[...] countries use TIMSS trend data to monitor the effectiveness of their education systems in a global context [...]". | "TIMSS assesses core content and cognitive domains in mathematics and science to be tested in grades 4 and 8. The TIMSS frameworks are organized around two dimensions: (i) the content dimension, specifying the subject domains to be assessed; and (ii) the cognitive dimension, specifying the thinking processes to be assessed. Content domains include algebra and geometry in mathematics; and biology and chemistry in science, while the cognitive domains of knowing, applying, and reasoning describe the thinking skills that students are expected to use as they engage with mathematics and science content." |
| ERCE | "ERCE serves as a tool to monitor the quality of education in the Latin American and Caribbean region, providing information for debate and guiding decision-making in this area." | "Students are assessed in reading, writing, mathematics and science."<br>"[...] considering the learning objectives that are common in the Latin American and Caribbean region, according to the curricular bases of each country." |
| PISA | "[...] each country assesses the knowledge and skills of its students in comparison with those of other countries, learns from policies and practices applied elsewhere and formulates its educational policies and programs with a view to improving the quality and equity of learning outcomes". | "PISA assesses three domains – reading, mathematics and science – in all editions or cycles". |
| PIRLS | "[...] analyze reading comprehension trends, in addition to collecting information about learning contexts, to characterize the reading process of students evaluated in the countries participating in the study". | "PIRLS assesses students' reading skills [...]". |

**Source:** Systematization made by the authors based on information in Brasil (s.d).

Regarding what is presented as the object of assessment, the initiatives generally identify skills, content and cognitive domains that are common to the participating countries, with reading being present in all of them. In addition to PIRLS, which has a unique focus, the others generally

also cover Mathematics and Sciences. In 2019, ERCE was administered to students in the 4th and 7th grades of Elementary School, covering constructed response items in both Mathematics and Natural Sciences. Among the initiatives analyzed, it is worth noting that TIMSS is the one that specifies in more detail the contents that are subject to assessment in its initial presentation on INEP›s website.

The characterization of large-scale assessment initiatives in Latin America, in relation to the terms analyzed here, demonstrates several similarities and few differences between them. But, above all, it reveals the increasingly robust interest of international organizations in creating measuring instruments to assess student learning and the possible factors associated with it, attempting to draw a comparison between and within educational systems.

## The quality of school education and large-scale assessments

The debate surrounding the quality of school education, which has gained greater visibility in recent decades, has become central to the proposals for educational policies in different territories, conceptions and nuances, especially due to the association of its conceptualization with the results of standardized tests, the most characteristic instruments of large-scale assessments. This is a recurring scenario on a global scale, as analyzed by Adams, Acedo and Popa (2012), observed after the Jomtien Conference in 1990, which established the consensus that school education is an indispensable condition for economic development and that quality should be a goal to be achieved. Furthermore, this scenario brought to the fore the notion of quality as something measurable.

Recognized as polysemic and ambiguous (Risopatron, 1991, p. 15), permeated by historicities, with the property of changing “in time and space” (Dourado; Oliveira, 2009, p. 203), expressing meanings according to “specific interests and values” of its proponents (Azevedo, 1994, p. 456), the concept of quality requires definition to operationalize its scope through educational policies.

It is inescapable, as several authors do, to highlight the difficulty in defining what constitutes the quality of school education, even more so in the face of an overwhelming international trend supported by international organizations, such as, for example, the World Bank, which make a biunivocal association between quality and standardized test results, or, as Broadfoot (2007) highlights, by imposing a quality agenda that seeks to extract from these results a broader set of information than they can provide.

Harvey and Green (1993), addressing this issue from the perspective of Higher Education, indicated that the quality of education has a dimension that is privileged based on the interests that are held in such education, and that there may be an emphasis on results or processes, with additional differentiations on these elements, connoting a greater or lesser scope. Thus, the discussion on quality is inexorably fraught with political tension, understood as a power struggle with varied arguments, as well as their consequences.

Although not exhaustive, the large-scale assessment initiatives that circulate in Brazil with official support and portrayed in this article configure this international framework that shape a vision of quality, even without a more finished presentation, which must be problematized.

In any case, as already highlighted, these results are relevant to the school process and therefore a position of refuting them *in totum* would not be a strategy that contributes to the debate and the formulation of alternatives. Therefore, the terms of the dialogue between internal assessment and external assessment formulated by Nevo (1998) seem to point to a denser horizon.

The indiscriminate use of quality of education and quality of teaching as analogous terms contributes to deepen the controversies over the concept of quality. Silva (2008, p. 81) asserts that there are countless cases of these expressions being used as equivalents, however, he points out that "the distinction between education and teaching is not irrelevant".

Dourado (2007, p. 11) defines quality of education as "the result of a construction of subjects engaged pedagogically, technically and politically in the educational process" in which it is important to weigh "the objective teaching conditions, the socioeconomic and cultural inequalities of students, the professional devaluation and the limited possibility of permanent updating of education professionals". For Chaves (2009, p. 9), *"cuando hablamos de la educación como derecho estamos entendiendo que la calidad es uno de los atributos del derecho, una de sus condiciones esenciales*" and, therefore, quality of education is a characteristic of the right to education. The quality of teaching is limited to the effects of the pedagogical process, configuring itself as an aspect of the quality of education, pivotal, but not the only one. In this sense, quality of education is an attribute that qualifies broader contexts in which learning takes place, while the quality of teaching refers directly to this learning.

The meaning given to the concept of quality has commonly been associated with measurement, obtained through the implementation of large-scale assessments. Casassus (2007, p. 43) explains that assessing has become necessary to give visibility to the qualitative aspects of education, translated as "academic success" based on standardized test scores and exams. Gil (2021, p. 205) highlights, as an extension of Casassus' (2007) position, that "often, what is considered as measuring the quality of education is the examination of the students' abilities". However, it is known that measurement is an important aspect, but an aspect, of assessment. In this sense, it is worth asking to what extent the methodology primarily used to measure the quality of education, standardized tests, would not be inducing schools to a reductionist perception of the quality of education, without associating other relevant factors with school results, such as social and economic context, for example (Nevo, 1998).

However, in a document from INEP (Brazil, 2019, p. 22), we find an attempt to reconsider this emphasis on student performance by delimiting other conceptual elements, indicating that "[...] the discussion on the quality of education [...] must encompass contextual factors [...] and the results of student performance in the cognitive tests usually disseminated".

In this same document (Brazil, 2019, p. 24) there is a proposition that, despite reaching beyond the scope of this text with regard to the prospects of establishing a confrontation with the trend previously highlighted, points to the need to incorporate other elements in the conceptual treatment of quality, in the sense of:

> Based on the discussions held and the studies produced within the scope of INEP, in summary, the proposed evaluation system aims to measure the quality of education based on seven axes: Equity, Human Rights and Citizenship, Teaching-Learning, Investment, School Services, Management and Teaching Professionals.

In view of this initiative, still in its embryonic stage, even considering the title of the document which highlights that it is a "preliminary version", we must point out that in Brazil, although the Federal Constitution of 1988, in its art. 206, item VII and the Law of Guidelines and Bases of National Education (LDB – Law No. 9,394/96), in art. 3, item IX, establishes that education will be provided based on the principle of guaranteeing quality standards, the first official definition for the concept was the creation of the Basic Education Development Index (IDEB). Disclosed together with the Commitment Everyone for Education Goals Plan, through Decree No. 6,074, of April 24, 2007, with the objective of being an "indicator of educational quality", in addition to enabling "permanent monitoring and measurement of the progress of programs in relation to the goals and results established" in Brazilian education, according to its creator Fernandes (2007, p. 6).

Every two years, IDEB publishes scores on a scale of 0 to 10 for states, municipalities and schools based on the approval rates and proficiency of students in Prova Brasil. In this sense, its principle is that "the quality of education presupposes that the student learns and achieves a passing grade" (Franco, Alves and Bonamino, 2007, p. 991). Examining the repercussions of the theme, Machado and Alavarse (2014, p. 422) argue that:

> Although the concept of quality associated with IDEB is somewhat reductionist, as it does not consider relevant aspects of the pedagogical process, it is possible to consider some potentialities in IDEB due to two characteristics: by facilitating an understanding, even if partial, of the Brazilian educational reality, highlighted by its schools, and, above all, by articulating two elements that have long seemed to be antagonistic: the increase in approval and the increase in performance.

The definition of quality is strongly supported by large-scale assessment, in the Brazilian case, by Prova Brasil, evidencing a movement to shape the meaning of quality through standardized tests and exams. In the proposals for external assessments by multilateral agencies, the definition of quality is advocated by these agencies that are responsible for important international assessments.

If IDEB advances by incorporating school flow, an element that was one of the concerns and meaning of the idea of quality in the 1980s, as Oliveira and Araújo (2005, p. 8) consider, it is necessary to emphasize that in the international scenario, even when using contextual questionnaires to capture associated factors, the concern with the results in standardized tests as a benchmark for quality continues to be the focus, as noted in an official document (Brazil, 2020, p. 15) on PISA, the main large-scale assessment in the world,

> [...] provides information on student performance linked to data on their backgrounds and attitudes towards learning, as well as on the key factors that shape their learning, both in and out of school. The results enable each country to assess the knowledge and skills of its own students in comparison with those of other countries; to learn from policies and practices applied elsewhere; and to shape its own education policies and programmes with a view to improving the quality and equity of learning outcomes.

This is reinforced in an excerpt from a document published on the INEP portal (UNESCO, 2020, p. 64), when reporting on the results of ERCE 2019 for Brazil, as they would be "guidelines to improve the quality of learning of Brazilian students". Mullis and Martin (2021, p. 41) also states that test results, in the case of PIRLS, signal "the high quality of reading performance".

It can be inferred, therefore, that quality is centrally based on student performance in standardized tests as the main objective of large-scale assessment, which shapes a notion of quality associated almost exclusively with the results of school learning, even though contextual questionnaires accompany these assessments, the documents express this tone.

## Final reflections

In this article, we seek to describe, characterize and analyze four large-scale assessments present in Latin America, in which Brazil is a participant ERCE, PIRLS, PISA and TIMSS - and problematize the notion of quality that they impose on school education.

The study enabled us to observe symmetries between such initiatives, mainly in relation to the objectives, commonly the diagnosis obtained through standardized tests for possible actions to improve the quality of school education, but also in relation to the target audience, content evaluated and instruments used.

Although the documents do not provide a clear and objective definition of quality, which would be absolutely relevant given the polysemy of the concept as already indicated, it is possible to infer that the quality advocated by the bodies responsible for these assessments is based almost strictly on student performance. Although all these assessments also apply contextual questionnaires that capture diverse information that can be analyzed as associated and explanatory factors of student performance, there is a strong emphasis on the results of standardized tests, which reflect the quality of teaching, an important but not the only aspect of the quality of education.

In some ways, the definition of quality incorporated by IDEB brings interesting aspects, even if several restrictions are pointed out regarding the procedures for its calculation. First, because quality in IDEB is not exactly the same as results for Saeb tests, nor is it completely associated with the idea that underpinned the cycle and automatic promotion policies of the 1990s; it is a combination that seeks to evoke discussions that took place within schools: how to make students learn and "pass"? IDEB also has, in its formulation, a notion of ipsative evaluation, that is, each school would be compared and evaluated in relation to itself based on the established goals. Additionally, it could be contested that the only results considered in IDEB›s calculations are those related to reading and mathematics, which could be interpreted as a "curricular reduction"; however, on the one hand, these would be contents that are transversal to all school subjects, but on the other, these are contents that long before the external evaluations were already deserving of predominance over others.

In this sense, it is argued that, even though student performance on standardized tests is not the quality of school education, the latter cannot be achieved without it. Furthermore, in an attempt to understand the complexity of the concept of quality of education, and to avoid the reductionist mistake that identifies student performance as its synonym, quality is defended not as excellence, but as characteristics that constitute and structure educational systems.

The literature is eloquent in its criticism, in some cases severe, of the majority of large-scale assessments that focus on measuring student performance, as already systematized by Bauer,

Alavarse and Oliveira (2015). However, the authors warn that "the role that these assessments will effectively assume in the educational policy as a whole is determined by the uses of their results", and it is timely to "analyze how the literature deals with the uses of the results of large-scale assessments" (Bauer, Alavarse and Oliveira, 2015, p. 1376).

Among the criticisms present in the literature involving the prominent role of large-scale assessments, the measurement of student performance and the development of intervention actions in specific aspects of the teaching-learning processes, the curriculum stands out, with the potential to generate direct consequences in improving the quality of education provision.

As problematized in Chappaz (2015), several times, even if the fundamental difference between the reference matrix of a large-scale assessment initiative and the curricular matrix of a school system is not considered (which has direct implications for the idea that large-scale assessments have an impact on the curriculum of schools in the sense of reducing them), it is necessary to reflect that, considering a large-scale assessment initiative that is well-designed - from the point of view of technical, methodological and pedagogical criteria -, better results of students in the tests would be reflecting a greater mastery of the skills and competencies present in the assessment matrices. Since this in turn is a section of a curricular matrix, better results therefore represent an advance in student learning, which may be a positive effect of the assessments on the quality of education. This aspect thus highlights and corroborates the perspective of quality considered in this text, so that the results of large-scale assessments are not the whole of quality, but are an important part of what makes it up.

The Brazilian experience of formulating and using IDEB, the main indicator of educational quality that is circulated and adopted in different spheres: schools, education departments, governments and the media, is an important reference for the debate on quality, as it is based on an international movement that identifies large-scale assessments and their results as a significant instrument in the design of strategies and the development of policies that aim to achieve objectives of improving the efficiency and quality of education. Furthermore, the IDEB, despite nor covering all the aspects that make up and impact the pedagogical process, advances the concept of quality by associating approval rates with the results of students in the large-scale assessment, indicating a concept of quality that values, in addition to performance, attendance at classes and permanence in school.

Finally, despite the general emphasis of large-scale assessments, almost exclusively on students› results in standardized tests, the recognition of the complexity and breadth of the concept of quality of education and the right of all to a quality education are an urgent challenge for educational policies.

## References

ADAMS, Don; ACEDO, Clementina; POPA, Simona. Quality education: challenges and policy choices. *In:* ADAMS, Don; ACEDO, Clementina; POPA, Simona. (Ed.). *Quality and Qualities:* Tensions in Education Reforms. Rotterdam: Sense; Paris: UNESCO: International Bureau of Education, 2012. (Comparative and International Education: a Diversity of Voices, v. 16). p. 197-208.

AZEVEDO, Janete. A temática da qualidade e a política educacional no Brasil. *Educação & Sociedade,* Campinas, n. 49, especial, p. 449-467, 1994.

BAUER, Adriana; ALAVARSE, Ocimar Munhoz; OLIVEIRA, Romualdo Portela. Avaliação em larga escala: uma sistematização do debate. *Educação & Pesquisa*, v. 41, n. especial, p. 1.367-1.382, 2015. https://doi.org/10.1590/S1517-9702201508144607

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Brasil no PISA 2018*. Brasília: INEP, 2020.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. Diretoria de Avaliação da Educação Básica. *Sistema de Avaliação da Educação Básica:* documentos de referência: versão preliminar. Brasília: INEP, 2019.

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira – INEP. *Avaliações e Exames Educacionais*, [s.d.]. https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais

BROADFOOT, Patricia. *An Introduction to Assessment.* New York: Continuum, 2007.

CASASSUS, Juan. *A escola e a desigualdade*. Brasília: Líber Livro, 2007.

CHAPPAZ, Raíssa de Oliveira. *A Prova São Paulo e as tensões das avaliações externas*: diálogos com o currículo oficial da rede municipal de ensino de São Paulo. Dissertação (Mestrado em Educação) - Faculdade de Educação, Universidade de São Paulo, São Paulo, 2015.

CHAVES, Orlando. La cuestión de la calidad de la educación. *Boletín del Foro Latinoamericano de Políticas Educativas*, n. 26, p. 1-11, 2009.

DOURADO, Luiz Fernando (Coord.). *A qualidade da educação*: conceitos e definições. Brasília: INEP, 2007.

DOURADO, Luiz Fernandes; OLIVEIRA, João Ferreira de. A qualidade da educação: perspectivas e desafios. *Cad. Cedes*, v. 29, n. 78, p. 201-215, 2009.

FERNANDES, Reynaldo. Índice de Desenvolvimento da Educação Básica (IDEB). Brasília: INEP, 2007. (Série Documental. Textos para Discussão, n. 26).

FRANCO, Creso; ALVES, Fátima; BONAMINO, Alicia. Qualidade do ensino fundamental: políticas, suas possibilidades, seus limites. *Educação & Sociedade*, v. 28, n. 100, especial, p. 989-1014, 2007. https://doi.org/10.1590/S0101-73302007000300017

GIL, Natália de Lacerda. A quantificação da qualidade: algumas considerações sobre os índices de reprovação escolar no Brasil. *Sociologias*, ano 23, n. 56, p. 184-209, 2021. https://doi.org/10.1590/15174522-109753

HARVEY, Lee; GREEN, Diana. Defining quality. *Assessment & Evaluation in Higher Education,* v. 18, n. 1, p. 9-34, 1993.

MACHADO, Cristiane; ALAVARSE, Ocimar Munhoz. Qualidade das escolas: tensões e potencialidades das avaliações externas. *Educação & Realidade*, v. 39, n. 2, p. 413-436, 2014. https://www.scielo.br/j/edreal/a/fyKb65xtFvXhMw3Hhpv9vNk/abstract/?lang=pt

MAROY, Christian; VOISIN, Annelise. As transformações recentes das políticas de accountability na educação: desafios e incidências das ferramentas de ação política. Trad. Marisa Rossetto. *Educação & Sociedade*, v. 34, n. 124, p. 881-901, 2013. https://doi.org/10.1590/S0101-73302013000300012

MULLIS, Ina; MARTIN, Michael (Org.). *PIRLS 2021*: fundamentos teóricos. Trad. Patrícia Helena Freitag. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais, 2021.

NEVO, David. Avaliação por diálogos: uma contribuição possível para o aprimoramento escolar. *In:* TIANA, Alejandro (Coord.). *Anais do Seminário Internacional de Avaliação Educacional*, 1 a 3 de dezembro de 1997. Trad. John Stephen Morris. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais, 1998. p. 89-98.

OLIVEIRA, Romualdo Portela; ARAUJO, Gilda Cardoso. Qualidade do Ensino: uma nova dimensão da luta pelo direito à educação. *Revista Brasileira de Educação*, n. 28, p. 5-23, 2005. https://www.scielo.br/j/rbedu/a/t64xS8jD8pz6yNFQNCk4n7L/?format=pdf&lang=pt

RISOPATRON, Veronica Edwards. *El concepto de calidad de la educación*. Chile: UNESCO/Orealc, 1991.

SILVA, Vandré Gomes da. *Por um sentido público da qualidade da educação.* Tese (Doutorado em Educação) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2008.

UNESCO. ¿Qué se espera que aprendan los estudiantes de América Latina y el Caribe?: análisis curricular del Estudio Regional Comparativo y Explicativo (ERCE 2019). Santiago: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO): Oficina Regional de Educación para América Latina y el Caribe (Orealc/UNESCO Santiago), 2020.

VERGER, Antoni; FONTDEVILA, Clara; PARCERISA, Luís. Reformando a governança através de instrumentos de políticas públicas: como e em que medida padrões, testes e accountability na educação se espalham pelo mundo. *In:* OLIVEIRA, Dalila Andrade; DUARTE, Adriana Maria Cancella; RODRIGUES, Cibele Maria Lima (Org.). *A política educacional em contexto de desigualdade*: uma análise das redes públicas de ensino da região Nordeste. Campinas: Mercado de Letras, 2019.

VERGER, Antoni; NORMAND, Romuald. Nueva gestión pública y educación: elementos teóricos y conceptuales para el estudio de un modelo de reforma educativa global. *Educação & Sociedade*, v. 36, n. 132, p. 599-622, 2015. https://doi.org/10.1590/ES0101-73302015152799

_____

**CRISTIANE MACHADO**

Phd in Education, Universidade de São Paulo (USP), São Paulo, São Paulo, Brazil; Associate Professor (Free Professor), Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil.

**OCIMAR MUNHOZ ALAVARSE**

Phd in Education, Universidade de São Paulo (USP), São Paulo, São Paulo, Brazil; Professor, Universidade de São Paulo (USP), São Paulo, São Paulo, Brazil.

**RAÍSSA DE OLIVEIRA CHAPPAZ**

Phd in Education, Universidade de São Paulo (USP), São Paulo, São Paulo, Brazil; Pedagogue, Instituto Federal de São Paulo (IFSP), São Paulo, São Paulo, Brazil.

## AUTHORS CONTRIBUTION

Author 1 – Conception and design of the research; analysis and interpretation of data; writing text topics; final review of the article.

Author 2 – Conception and design of the research; collection, processing, analysis and interpretation of data; final review of the article.
Author 3 – Conception and design of the research; construction and processing of data; analysis and interpretation of data; preparation of the final text.

## SUPPORT/FINANCING

There was no funding.

## RESEARCH DATA AVAILABILITY

Does not apply.

## HOW TO CITE THIS ARTICLE

MACHADO, Cristiane; ALAVARSE, Ocimar Munhoz; CHAPPAZ, Raíssa de Oliveira. Large-scale assessments in Latin America: problematizing the quality of school education. Educar em Revista, Curitiba, v. 40, e94031, 2024. https://doi.org/10.1590/1984-0411.94031

*This article was translated by Olivia Lima – E-mail: oliviasable@yahoo.com. After being designed, it was submitted for validation by the author(s) before publication.*

---

**Received:** 01/10/2024
**Approved:** 08/19/2024