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

https://doi.org/10.1590/1984-0411.93646-T

DOSSIER

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

Weaknesses in PISA, a global protagonist in educational policies

As fragilidades do PISA, um protagonista global em políticas educacionais

João Luiz Horta Neto[a]
jlhorta@gmail.com

ABSTRACT

Based on documentary research and academic production, this essay discusses PISA and the cognitive tests results published by OECD. For this purpose, we will present and discuss the main structuring concepts of the three PISA tests until 2021, how they have varied over the years, and some of the weaknesses in results disclosure using proficiency scales as instruments used to explain the students' tests performance tasks. Its influence on educational policies worldwide is briefly discussed, often focusing on achieving the best performance among participating countries. The analysis points to some inconsistencies in the PISA data, hindering its use to extend the educational debate and improve educational policies.

Keywords: PISA. Educational Assessment. Curriculum. Educational Policy.

RESUMO

O objetivo deste texto, baseado em pesquisa documental e na produção acadêmica, é discutir o Pisa e os resultados dos testes cognitivos divulgados pela OECD. Para tanto, serão apresentadas e discutidas as principais concepções que estruturam os três testes do Pisa até 2021, como elas vêm variando ao longo dos anos e algumas das fragilidades na forma de divulgar seus resultados, utilizando, para isso, as escalas de proficiência produzidas, a forma encontrada para explicar as tarefas que os estudantes são capazes de realizar nos testes. Discute-se a enorme influência que este instrumento tem sobre as políticas educacionais ao redor do mundo, na maior parte das vezes, discutindo como alcançar o melhor desempenho entre os países participantes. As análises apontam para algumas inconsistências nas informações geradas pelo Pisa, dificultando seu uso para aprofundar o debate educacional e aprimorar as políticas da área.

Palavras-chave: PISA. Avaliação Educacional. Currículo. Política Educacional.

[a] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Brasília, Distrito Federal, Brazil.

Situating the evaluations in Brazil: SAEB and International Studies[1]

In the last 30 years, the federal government has developed a robust and consistent set of measurement instruments to evaluate Brazilian education. These efforts began during the Brazilian redemocratization in the late 1980s and involved administrations of different ideological profiles. Created in 1988 as a pilot project, the Brazilian Basic Education Evaluation System (SAEB) had its first application cycle in 1990. Since 1993, it has been applied every two years.

In the first two cycles, its initial project was to use a series of instruments, including cognitive tests, to generate three groups of indicators to evaluate Brazilian education's quality. Groups linked to the federal government developed most of these instruments for at least ten years and added them to a logical and harmonious set. At its core, the proposal considered school education as a complex public action, and to evaluate it, measures that captured some of the educational process dimensions would be necessary.

One such instrument was the "*Estudo do Aluno*" (Student Evaluation), which involved applying multiple-choice and constructed response cognitive tests to students from a sample of schools enrolled in the 2nd, 4th, 6th, and 8th grades of basic education, considered fundamental in the student learning process. There was no national curriculum, so educational experts, school principals, and teachers participated in surveys to obtain information about the curriculum taught to build a reference for item elaboration. Once the tests were applied, their results were analyzed using Classical Test Theory to facilitate the teachers' analysis.

From 1995 onwards, the federal government discarded the original model, and SAEB focused only on cognitive tests and questionnaires applied to students, teachers, and school principals, also belonging to a sample. Also, the school years tested changed to the 5th, 9th, and 12th grades, corresponding to the end of the basic education cycles. Furthermore, from that year on, the analysis of the results began to be performed using Item Response Theory (IRT) to allow comparisons between students' performances across the different SAEB cycles. As of 2005, SAEB started to be applied in a census form to all students in those school years enrolled in the public school systems (Horta Neto, 2018).

Based on the data on student performance in the SAEB tests and the school flow collected by the School Census, the Basic Education Development Index (IDEB) was created. IDEB was composed by parameterizing the performance data in the SAEB tests in relation to the average of the Organization for Economic Cooperation and Development (OECD) countries participating in the Program for International Student Assessment (PISA) in 2003. Thus, at its core, the IDEB measure brings an international comparison to put into perspective the Brazilian advances before the international reality (Fernandes, 2017). Based on the IDEB, the Brazilian National Institute for Educational Studies and Research Anísio Teixeira (INEP) set goals for all schools so that in 2022, the year of the bicentennial of independence, all schools would reach the IDEB index of six for the 5th grade of Primary Education, a value equivalent to the average performance of OECD countries in PISA 2003.

[1] Parts of this text were included in the chapter "PISA and curricular reforms in Brazil: the influence of a powerful regulatory instrument" (Horta Neto, 2022).

Besides this national assessment, Brazil has participated in international studies such as those produced by the United Nations Educational, Scientific and Cultural Organization (UNESCO) within the Latin American Laboratory for the Assessment of Quality Education (LLECE) since 1997 (UNESCO, 2001; 2008; 2016) and, as of 2019, has signed cooperation agreements to participate in studies promoted by the International Association for the Evaluation of Educational Achievement (IEA): Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS) and International Civic and Citizenship Education Study (ICCS).

The leading international study in which Brazil participates—and the one that provokes the most discussions – is PISA, organized by the OECD. According to the federal government, the two main objectives of this participation are to have an international reference on student performance and to appropriate methodologies and technology in educational evaluation that help develop national assessments (INEP, 2012).

Table 1 presents the average proficiency of Brazilian students compared with those of OECD countries.

Table 1: Comparison of Brazil’s performance in the PISA tests and the average proficiencies of OECD member countries: 2000 to 2015

Year	Reading		Mathematics		Science	
	Brazil	OECD	Brazil	OECD	Brazil	OECD
2000	396	496	-	-	-	-
2003	403	497	356	499	-	-
2006	393	490	370	497	390	498
2009	412	496	386	495	405	501
2012	407	498	389	496	402	501
2015	407	493	377	490	401	493
2018	413	487	384	489	404	489

Source: INEP, 2019.

In 2012, after several results considered harmful, the federal government highlighted that Brazil had advanced the most in Mathematics proficiency between 2003 and 2012. It also stated that the country is a “model to be followed by countries with greater educational lag and that still face the challenge of including students in the school system” (INEP, 2012, p. 14). In 2015, the following cycle, Brazil’s performance was lower than in 2009, six years ago, without any comment from the federal government[2]. Such a position indicates the political use of PISA results.

The following sections present the constructs that structure the PISA tests and how their results are disclosed. Our goal is to analyze the weaknesses of the data generated and its effects on the public debate from the texts produced by OECD and INEP.

2 The OECD report discusses this fact (OECD, 2016).

Performance measures in PISA

PISA has been standing out as a knowledge-based regulatory instrument (Carvalho, 2020), a developer of educational policies, an agent in the transnationalization of these policies, and a mediator of knowledge (Teodoro, 2020). As such, it has influenced the development of educational policies worldwide. One consequence of this process is the standardization of educational policies, resulting in a discourse homogenization on the most appropriate ways to improve education and prepare better 21st-century citizens, a future designed by OECD. It also influences national instruments that, through cognitive performance tests, seek to regulate education provision and the main results of the educational process. This influence is strongly felt in Brazil, a peripheral country in the world economy and the first non-OECD member to join PISA in 1998, still in its formulation phase.

In each three-year cycle, one area is the central theme among the tests applied, and the other two are tested more limitedly. If 'Reading' was the central theme of the first cycle in 2000, it is once again central in the 2009 cycle, since in 2003, the central theme was Mathematics and, in 2006, Science.

One of the essential steps for developing any measurement instrument is clearly defining what one wants to measure (Horta Neto, 2010). In the case of PISA, reading the reports produced between the 2000 and 2018 cycles revealed four intertwined concepts: knowledge, abilities, competencies, and literacy. Based on these concepts, each area of knowledge develops its constructs, defining what each test is measuring. These are comprehensive formulations that, besides being a reference for PISA, influence several curricular proposals worldwide.

Interpretations of the proficiency scales are to disclose the test results and provide pedagogical information on them, which are made for intervals within the PISA scale called Levels. Each Level describes the performed test tasks. As proficiencies are comparable throughout the different PISA cycles, new skills are described with each new cycle, expanding the knowledge about the characteristics of the tasks students are able to develop at each proficiency level. Hence, the descriptions disclosed in the 2018 cycle are more in-depth than those of the first cycle. Those following cycles produce new information for different scale points, expanding our knowledge about the skills developed. To describe the scale, a group of experts studies the test items[3] and analyzes the tasks demanded and the proficiency needed to solve them. From this analysis, they describe what the students were able to answer.

As for the scale description, Level 2 is considered the basic, indicated by PISA as the minimum expected test performance and reflects the abilities that 15-year-old students should have developed. Conversely, those in Level 1 fall short of the expected abilities, while those in Level 6 have achieved excellence. However, only a tiny percentage of students are classified at this Level.

A probabilistic function known as Rasch's IRT model is used to measure proficiency as it allows comparisons between the different cycles to monitor measurement changes. However, it

[3] Examples of items disclosed by the OECD can be obtained at OECD (2023).

hinders disclosing all items used in one cycle as some of them must be reapplied in the next to ensure comparability between cycles.

The following section will describe the definitions of each area of knowledge's construct, the changes it underwent throughout the various PISA cycles, and the results of the 2018 cycle based on the report on Brazil's participation in the study.

Reading

For the Reading area, the PISA reports show slight variation in the definition of its construct over the cycles. In PISA 2000, when Reading was the main theme of the study, it was defined that "Reading Literacy is understanding, using, and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society" (OECD, 1999, p. 20).

This construct undergoes a single change in the 2009 cycle when Reading is once again the main theme. In that cycle, a small excerpt was included in the Reading Literacy definition referring to students' engagement with Reading. The new definition, with the inclusion highlighted in bold, was as follows: "Reading literacy is understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society" (OECD, 2009, p. 23).

This change reflects the inclusion of the third pillar in the concept of competence, defined as knowledge, ability and attitude. For PISA, attitude is transformed into engagement with Reading, measured through some items of the student, teacher, and school questionnaires, which are then analyzed with those of the cognitive test. The inclusion of engagement with Reading aimed to highlight contextual factors and their influence on the measured proficiencies, seeking to complement the information collected. Hence, engagement with Reading refers to motivational attributes and behavioral characteristics of student reading (OECD, 2009, p.70). Students' attitudinal characteristics refer to the quantity and intensity of their reading activities. In contrast, motivational attributes refer to the extent to which reading is encouraged by the teacher, classroom activities, and the school.

In the 2018 cycle, when Reading was once again the main theme, the report (OECD, 2019a) highlighted its dependence on readers' characteristics, textual factors (formats, complexity of the language used), and the task to be performed (reading for pleasure, for a deeper or lighter understanding). It also states that the selection of what will be used in the test must ensure several types of texts with different purposes, in school and outside, to naturally represent a range of difficulties involving both the texts and the tasks to be performed.

The first doubt that such a definition brings is concerning the diverse cultural realities of each country and the texts produced there, each with its singularities. Choosing texts that cross different cultures is not an easy task. It raises questions about the real possibilities of comparing the results between students who live reading practices and cultural patterns so different. In other words, if it is not possible to ensure comparability between texts, it can induce a homogeneity of curricula at a global level.

Based on the results of the test applied, Table 2 presents what students are able to do in Reading Literacy. It indicates nine Levels, from 6 to 2 plus levels "1a", "1b", "1c", and "below 1c" (OECD, 2019b, p. 87-88). For each of them, sentences describe the OECD's interpretation of what students were able to do based on the test items. Consequently, what was not presented in the test cannot be measured and described.

Due to the limited article space, we cannot analyze the entire content of that table. For this reason, some excerpts from the descriptions were selected that present everyday tasks described at different levels. Just as we selected those shown in Table 2, others could have been chosen without jeopardizing the quality and validity of the analysis performed. For this text, we chose excerpts from the descriptions that referred to the ability to "make inferences," one of the tasks demanded by the Language test. Since it is impossible to encompass all the analysis possibilities here, as highlighted above, we selected some of the Levels to exemplify the extent of this ability, the minimum proficiency to perform each task, and the percentage of OECD and Brazilian students classified in each Level.

Table 2: Description of the Reading Literacy proficiency scale in PISA 2018

Level	Minimum proficiency	Percentage of students at each Level	Characteristics of tasks
6	698	OECD: 1.3% Brazil: 0.2%	(...) They can compare, contrast, and integrate information representing multiple and potentially conflicting perspectives, using multiple criteria and generating **inferences** across distant pieces of information to determine how the information may be used. (...)
4	553	OECD: 18.9% Brazil: 7.4%	(...) They interpret the meaning of nuances of language in a section of text by taking into account the text as a whole. In other interpretative tasks, students demonstrate understanding and application of *ad hoc* categories. They can compare perspectives and draw **inferences** based on multiple sources. (...)
2	407	OECD: 23.7% Brazil: 24.5%	(...) They can understand relationships or construe meaning within a limited part of the text when the information is not prominent by producing basic **inferences** and/or when the text(s) include some distracting information. (...)
1b	262	OECD: 6.2% Brazil: 17.7%	(...) They can also interpret the literal meaning of texts by making simple connections between adjacent pieces of information in the question and/or the text. (...)

Source: author's elaboration based on OECD (2019b).

First, we must comment on some inaccuracies in how PISA presents its results. As previously mentioned, proficiency is calculated using the IRT model, which expresses a probability that the described task has been performed. Hence, we have no certainty that all those classified at a certain level have done so. From the language used in the scale description, one gets the impression that the students classified at a certain Level were able to perform all the tasks described in it, which may not be accurate. Moreover, like any measure, it is associated with an error, i.e., the proficiency value indicated varies within a margin of error. These limitations are not informed in the scale description. One gets the impression that the characteristics expressed in these tables describing the tasks performed are exact information when, in reality, they are approximate interpretations of the measured results.

When comparing the different scale levels regarding the tasks involving the ability to make inferences, only one (Level 1b) mentions no task related to this ability. Hence, the report implies that students classified at that Level were unable to make inferences. Nevertheless, which inferences does PISA refer to? Without having access to all the items in which this task was demanded, it is impossible to understand the tasks to which students were submitted. Just reading OECD's interpretation of what students accomplished does not allow the reader to understand in what context and based on which texts students were asked to make inferences.

Additionally, as the tasks that students classified in Level 1b do not include making inferences, we can suppose that 17.7% of Brazilian students demonstrated not having developed this ability. This supposition can be valid for the texts and tasks demanded by the test items. It is impossible to suppose they cannot make inferences in other situations, even simpler ones. However, the way information is presented to society, including policymakers, the media, and education professionals, illuminates the failure of a school system that allows 17.7% of its students to be unable to make inferences. If, on the one hand, there is a significant difference in performance between OECD and Brazilian students due to the different realities in which they live, on the other, it is pedagogically not credible that such a large percentage of students are unable to make some kind of inference. What we question are possible uses made of the results obtained. If the limitations of the measures are not clarified, countless conclusions can be drawn. Therefore, discourses defending alleged ideal models proliferate, indicating them as silver bullets capable of overcoming social inequalities and ensuring improvements in learning, even without clearly defining what they would be.

Another piece of data highlighted in Table 2, referring to Level 4, points out that students "can compare perspectives and draw inferences based on multiple sources." Such information is minimal, as it does not indicate the actual dimension of the task demanded. This is the problem with analyzing results expressed in scales, whatever they may be. As the number alone means little, the attempt to describe what students can do based on the tasks presented by the items is quite limited and may give rise to multiple interpretations.

Thus, the information in Table 2 indicates that the ability to make inferences can be measured in different tasks with different degrees of difficulty. This can be verified by the gradation of tasks presented to the students. Nothing more than that. Therefore, it is challenging to formulate educational policies to improve students' reading comprehension based only on information like this. What the tables with the descriptions of the proficiency levels inform us are the differences between nations and economies, both in terms of the overall placement in the ranking and the percentage of students classified in each Level. This information is already enough to stimulate competition between countries.

Mathematics

PISA defined this area of knowledge in 2000 as follows:

> Mathematical Literacy is an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgments and to engage in mathematics, in ways that meet the needs of that individual's current and future life as a constructive, concerned and reflective citizen (OECD, 1999 p. 41).

The document states that traditional knowledge and skills, defined in school, are not the focus of PISA, which emphasizes mathematical knowledge used in different contexts requiring reflection and discernment. This definition does not explicitly state that PISA tests 15-year-old students, regardless of the grade they attend, which is a problem for countries with high failure rates, such as Brazil. Additionally, it defines the competencies that the test will measure: mathematical thinking, argumentation and modeling, problem-solving and communication. In preparing the tests, these competencies are used in three dimensions: processes, focusing on the student's ability to analyze, reason, and communicate ideas effectively, presenting, formulating and solving mathematical problems; content, using knowledge related to change and growth, space and form, chance, quantitative reasoning and relations of uncertainty and dependence; context, emphasizing the making and use of mathematics in situations such as personal and school life, work and sports, local community and society (OECD, 1999).

In 2012, the second time that Mathematics was the main PISA theme (the first was in 2003), the document describing the test matrices for this cycle points out that "an understanding of mathematics is central to a young person's preparedness for life in modern society" (OECD, 2013 p.24), bringing to the area a utilitarian vision of education, without clearly defining which life he refers to and what modern society this would be. It also presents changes in the concept of literacy, seeking to clarify better the relation between competencies and what would be demanded by the items, which was unclear in the original formulation. Thus:

> Mathematical Literacy is an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognize the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens (OECD, 2013 p. 25).

As for student engagement, unlike Reading, in which questionnaires are used to measure attitudes and motivations, engagement in Mathematics is measured using mathematical processes within the item itself. As highlighted in the definition of Mathematical Literacy, these processes verify students' ability to formulate situations mathematically, employ mathematical procedures and concepts, and interpret and evaluate mathematical results. The document highlights the importance given to this student engagement by stating that "it is important for both policymakers and those engaged more closely in the day-to-day education of students to know how effectively students are able to engage in each of these processes" (OECD, 2013 p.28).

In the 2021 cycle, when Mathematics was once again the main theme, the concept of literacy remained the same. Still, the way of expressing its model has become more sophisticated, as it makes explicit the mathematical content explored by the test, something that, despite being discussed in previous texts, did not gain prominence (OECD, 2018). Moreover, the proposed model highlights 21st-century skills, called "challenges in a real-world context." Experts define these abilities as "mathematical literacy supports and develops" (OECD, 2018, p. 2). Such conceptualization formulated within PISA, used to guide item elaboration and measure students' performance in Mathematics, given its scope, end up becoming indications of how school systems should guide

Mathematics teaching. By disclosing several years in advance the formulations used to perform measurements in the PISA test, guidelines are produced to induce curricular changes around the world, impacting not only the area of Mathematics but all other curriculum areas. These guidelines become a paradigm to be followed to shape the global citizen. What is posed here is not a direct criticism of any of the so-called "21st-century Skills" that could, if democratically debated, be one of the bases for discussing curricular improvements but a subtle way of inducing global changes in curricula and educational conceptions based on a model to measure student performance in PISA. Not following these guidelines can have a high price and result in worse results in Mathematics in the following cycles. These experts (Carvalho, 2009) define how education should be developed worldwide without the need for significant and lengthy debates. Those debates are vital in any democratic process to determine the directions to be followed in each society.

Mathematics results are presented in seven Levels, from 6 to 1 and "below 1." As in Reading, tables are also produced with interpretations of the proficiency scale in the Mathematics test. We chose the same Levels as Table 2 and some excerpts from the one created in 2018 to simplify the analysis and compose Table 3.

Table 3: Description of the Mathematics Literacy proficiency scale in PISA 2018

Level	Minimum proficiency	Percentage of students at each Level	Characteristics of tasks
6	669	OECD: 2.4% Brazil: 0.1%	(...) Students at this Level can reflect on their actions, and can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments **and the appropriateness of these to the original situation**.
4	545	OECD: 18.5% Brazil: 3.4%	(...) Students at this Level can utilize their limited range of skills and can reason with some insight in straightforward contexts. **They can construct and communicate explanations** and arguments based on their interpretations, arguments and actions.
2	420	OECD: 22.2% Brazil: 18.2%	(...) **They can extract relevant information from a single source and make use of a single representational mode**. (...) They are capable of making literal interpretations of results.
1	358	OECD: 14.8% Brazil: 27.1%	(...) **They are able to identify information and carry out routine procedures according to direct instructions in explicit situations**. They can perform actions that are almost always obvious and follow immediately from the given stimuli.
Below 1		OECD: 9.1% Brazil: 41.0%	No abilities identified.

Source: author's elaboration based on OECD (2019b).

Table 3 shows that 41% of the Brazilian students had a below Level 1 performance. For this same Level, no skills were described. For an expert in psychometrics, this data highlights that the test items were not adequate to measure skills in that proficiency range and that other items should be included in the test if we want to measure what the students are able to do. Thus, a technical problem to be solved. However, for the general public, particularly the media, the information that stands out is that 41% of Brazilian students lack skills in Mathematics or, in other words, have not learned Mathematics and are unprepared to become 21st-century citizens. As we add the percentages of Level 1 with Level below 1, a total of 68.1% of Brazilian students are classified below

Level 2, considered as basic. This information appears in the news published by a major news site in Brazil: "PISA 2018: two-thirds of 15-year-old Brazilians know less than the basics of Mathematics" (Moreno, 2023).

As for the characteristics of tasks described in Levels 4 and 6, the highlighted excerpts refer to the abilities of interpreting and arguing. Comparing the two excerpts from Table 3, "[students at this level] can formulate and precisely communicate their actions and reflections regarding... **the appropriateness of these to the original situation**" and "[students] **can construct and communicate** explanations and arguments based on their interpretations, arguments and actions," what would be the difference between the two levels? Both excerpts refer to the task of explaining something and, at first glance, refer to similar abilities. However, these two levels are 124 points apart within the scale. Similarly, for Levels 1 and 2, if we take the expressions "are able to identify information and carry out routine procedures according to direct instructions in explicit situations" and "can extract relevant information from a single source and make use of a single representational mode," the gradation between them is unclear. Despite being 62 points apart on the scale, they seem to identify very close tasks.

Hence, PISA's goal to be "a politically oriented instrument in the sense of guiding decisions for educational policies"[4] is partially unfulfilled. Moreover, the information is pedagogically ineffective since it is unable to point out ways to improve pedagogical practices. Conversely, they surreptitiously inform that student performance can be enhanced if the curricula are developed based on the proposed model. That is the core of the information PISA wants to give. It is not direct information but uses subterfuges to impose models. It also imposes classroom practices using PISA items, which are very well-prepared and creative. However, these items are produced by experts with several years of training who have spent hours elaborating each of them and have received lots of feedback from other experts who helped them to improve the items. Thus, far from the reality of the vast majority of teachers anywhere in the world.

Science

The Pisa 2000 cycle defined Scientific Literacy as "the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions to understand and help make decisions about the natural world and the changes made to it through human activity" (OECD, 1999, p.60).

The document highlights that the test to measure Scientific Literacy must consider the bases of scientific processes, adapting them to the tasks demanded, the knowledge that must be mobilized, and the context in which the tasks are presented. Each of these three aspects is detailed in the OECD text indicated above.

In the 2006 cycle, Science was the central area, and Pisa presented a new definition of Scientific Literacy. It included attitudinal aspects measured from the students' responses concerning four elements of the individual:

---

[4] This is one of the long-term strategies that the PISA Governing Board, the body in which OECD member countries participate and which decides the test strategies, defined in 2015 (OECD, n.d.).

> • Scientific knowledge and use of that knowledge to identify questions, acquire new knowledge, explain scientific phenomena and draw evidence-based conclusions about science-related issues
>
> • Understanding of the characteristic features of science as a form of human knowledge and inquiry
>
> • Awareness of how science and technology shape our material, intellectual, and cultural environments
>
> • Willingness to engage in science-related issues and with the ideas of science, as a reflective citizen (OECD, 2006 p. 23).

With this new theoretical framework, in 2006, the test was formed by several units of items, each related to a particular context. In addition to the cognitive items, each unit had an item at the end to measure the student's engagement with Science. The latter considered the four aspects mentioned above. In Annex A of the OECD (2006), one can find examples of these units of items. Besides the questions in the cognitive test, others were included in the questionnaires to measure the student's engagement with Science. Thus, the area chose a different model from Reading and Mathematics to measure Science Literacy. However, there is no discussion to indicate the rationale for doing so.

In 2015, when Science was once again the main theme of PISA, the Scientific Literacy proposal of the previous cycle improved:

> Scientific Literacy is the ability to engage with science-related issues, and with the ideas of science, as a reflective citizen.
>
> A scientifically literate person is willing to engage in reasoned discourse about science and technology, which requires the competencies to:
>
> • Explain phenomena scientifically – recognize, offer and evaluate explanations for a range of natural and technological phenomena.
>
> • Evaluate and design scientific inquiry – describe and appraise scientific investigations and propose ways of addressing questions scientifically.
>
> • Interpret data and evidence scientifically – analyze and evaluate data, claims and arguments in a variety of representations and draw appropriate scientific conclusions (OECD, 2017, p. 22).

Without room for a more in-depth analysis in this text, the new formulation is close to that used for Mathematics, especially regarding the explanation of reasoning. Another change was the exclusion of questions to measure engagement with Science, presented at the end of the units of items, which are now part of the questionnaires. According to the OECD text, there were two problems with the previous formulation: the space allocated to cognitive items was reduced in the questions, and there was a mismatch between the answers given in the test and those in the questionnaires. Another critical point in the Science framework's definition is associating Scientific Literacy with a "Key Competency" (OECD, 2017, p. 20).

A proposal for the 2024 cycle is already being discussed within PISA, including new changes in the formulation of Scientific Literacy. Three new areas of knowledge are suggested: informatics, competencies of using scientific knowledge for action and decision-making, and probabilistic thinking. Pisa added another dimension called scientific identity (OECD, 2020). Including this new proposed change, the area of Science underwent the most improvements over the different cycles.

Eight different Levels, from 6 to 2 plus levels “1a”, “1b”, and “below 1b”, presents the Science results. To discuss the 2018 results, Table 4 brings excerpts from the document to describe the characteristics of the tasks that students were able to develop according to their proficiency levels. The selected excerpts refer to the ability to interpret data.

Table 4 – Description of the Science Literacy proficiency scale in PISA 2018

Level	Minimum proficiency	Percentage of students at each Level	Characteristics of tasks
6	708	OECD: 0.8% Brazil: 0.0%	(...) In interpreting data and evidence, they are able to discriminate between relevant and irrelevant information and can draw on knowledge external to the normal school curriculum (...).
4	559	OECD: 18.1% Brazil: 4.6%	(...) Level 4 students can interpret **data drawn from a moderately complex data set** or less familiar context, draw appropriate conclusions that go beyond the data and provide justifications for their choices
2	410	OECD: 25.8% Brazil: 25.3%	(...) students are able to draw on everyday content knowledge and **basic procedural knowledge** to identify an appropriate scientific explanation, interpret data and identify the question being addressed in a simple experimental design. (...).
1a	335	OECD: 16.0% Brazil: 31.4%	(...) With support, they can undertake structured scientific enquiries with no more than two variables. They are able to identify simple causal or correlational relationships and interpret graphical and visual data that require a **low level of cognitive demand**. (...).
Below 1b		OECD: 0.7% Brazil: 4.0%	OECD does not specify the abilities developed.

Source: author’s elaboration based on OECD (2019b).

As with the other two areas, our analysis seeks to point out the weaknesses of the data provided, which, in principle, should serve to guide policymakers. As mentioned before, the information is limited and hinders the ability to clearly understand the characteristics of tasks that students are able to solve.

In Level 6, the document states that students can draw on knowledge external to the school curriculum. However, as PISA is not based on the school curricula of the participating countries and economies, and as no study comparing the different curricula has been conducted, it is impossible to identify what this knowledge would be. Although the task was performed by only 0.8% of OECD students, and none in Brazil, one wonders the practical sense of this information.

As for the following three Levels, despite the gradation seen in the highlighted passages regarding the tasks students are probably capable of doing (data drawn from a moderately complex data set, at Level 4; basic procedural knowledge, at Level 2; low level of cognitive demand, at Level 1a), we have no information on what type of task the descriptions refer to. There are countless possibilities, even more so when they refer to scientific knowledge used to explain phenomena. Another example is the description in Level “1a,” indicating that “with support, they [students] can undertake structured scientific enquiries with no more than two variables.” Despite providing unclear information, it classifies more than 31% of Brazilian students among those who perform activities that require a low level of cognitive demand. This inconsistent and ambiguous information is available for public debate on Science teaching.

Final considerations

Cognitive tests and the PISA questionnaires are well-designed instruments that have counted on contributions from influential world experts for their constant improvement and have served as a model for several national and international studies. Modifications are announced several years in advance, perhaps to prepare countries and economies for the new formats used. This fact indicates a possible concern with the participating countries' ability to prepare themselves for instrument changes, especially cognitive tests. Thus, PISA constitutes a regulatory and policy-inducing instrument, influencing several countries and economies by stimulating competition and driving the educational quasi-market, making it a powerful instrument.

Descriptions of the constructs for each area, rather than indicating what will be measured, present discussions about what should be taught, how each area of knowledge should be structured, and how curricula should be constructed. Thus, texts referencing the cognitive tests go beyond information on what PISA measures. In a more complex way, there are paths to be followed so that improved results in the subsequent cycles translate to the success of the policies adopted. Pisa's essence is mobilizing politicians, the media and influencing societies.

In this debate, surpassing the other means being on the right path to shaping the so-called 21st-century citizen, as if there were a consensus on who that citizen should be and if there was only one educational model to form them. Such logic gains an air of solid scientific basis by presenting several statistical studies from cross-referencing different sources, primarily questionnaires.

However, when examining the cognitive test results in isolation, the data produced, which could, for example, guide the initial and continuing teacher training and education, is fragile. Descriptions of the scales tell us very little and often mistakenly about the student's test performance tasks. This is surprising since the use of cognitive tests as an instrument to monitor the results of educational systems emerged as a significant advance of comparative educational studies, as it allows to inform what and how much students would be learning in school. This is PISA's main weakness.

Criticism is raised not concerning possible methodological problems of the instruments but rather against the logic used to disclose their results. If using IRT allows us to make reliable comparisons over time, it does so from a concept that is incomprehensible to the educational world: that of proficiency, reflected in numbers expressed on a scale completely different from that used by schools. However, even if one has access to the set of items applied in the tests, they are informed about those tasks presented to students every three years, not all those that teachers and the school system must focus on to ensure meaningful learning. IRT ends up prompting test preparation by using a logic that is incomprehensible to education professionals and given the power it exerts over policies.

OECD has been increasingly expanding its power of influence. For example, PISA can be an essential instrument for monitoring Sustainable Development Goal 4 (SDG 4): Quality education, which aims to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (ONU, 2024). Meanwhile, it makes available the "PISA for Development" to the poorest countries, the "PISA for Schools" to schools, and develops the *International Early Learning* *and Child Well-being Study*, known as the PISA-baby. OECD thus contributes to homogenizing educational practices, teacher training, and education and textbook contents.

It also influences policies in Brazil. One such example is the debate and implementation process of Brazil's National Curriculum (BNCC), which was outlined using some of the PISA formulations as a model (Horta Neto, 2022). This resulted in a document that ignores specifying the concepts that should guide Brazilian education in favor of defining learning objectives that seem more like recipes for how and what to teach. Consequently, Brazil will face increasing pressure for more tests and higher performance.

References

CARVALHO, Luís Miguel. Governando a educação pelo espelho do perito: uma análise do PISA como instrumento de regulação. *Educação & Sociedade*, v. 30, n. 109, p. 1009-1036, 2009. https://doi.org/10.1590/S0101-73302009000400005

CARVALHO, Luís Miguel. Revisiting the Fabrications of PISA. *In*: FAN Guori, POPKEWITZ, Thomas. *Handbook of Education Policy Studies*. Berlin: Springer, 2020. p. 259-273.

FERNANDES, Reynaldo. A universalização da avaliação e a criação do IDEB: pressupostos e perspectivas. *Em Aberto,* v. 29, n. 96, p 177-193, 2017. Sistema de Avaliação da Educação Básica (SAEB): 25 anos. https://rbep.inep.gov.br/ojs3/index.php/emaberto/article/view/3154

HORTA NETO, João Luiz. Avaliação externa de escolas e sistemas: questões presentes no debate sobre o tema. *Revista Brasileira de Estudos Pedagógicos*, v. 91, n. 227, p. 84-104, 2010. https://doi.org/10.24109/2176-6681.rbep.91i227.604

HORTA NETO, João Luiz. Avaliação educacional no Brasil para além dos testes cognitivos. *Revista de Educação PUC-Campinas*, v. 23 n. 1, p. 37-53, 2018. https://doi.org/10.24220/2318-0870v23n1a3990

HORTA NETO, João Luiz. PISA and curricular reforms in Brazil: the influence of a powerful regulatory instrument. *In*: TEODORO, António. *Critical perspectives on PISA as a means of global governance:* risks, limitations, and humanistics alternatives. London: Routledge, 2022. p. 70-103.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Relatório Nacional PISA 2012*: resultados brasileiros. INEP, 2012.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Relatório Brasil no PISA 2018*: versão preliminar. Brasília: INEP, 2019. https://download.inep.gov.br/acoes_internacionais/pisa/documentos/2019/relatorio_PISA_2018_preliminar.pdf

MORENO, Ana Carolina. PISA 2018: dois terços dos brasileiros de 15 anos sabem menos que o básico de Matemática. *G1 - Educação*, 13 dec. 2019. https://g1.globo.com/educacao/noticia/2019/12/03/pisa-2018-dois-tercos-dos-brasileiros-de-15-anos-sabem-menos-que-o-basico-de-matematica.ghtml

OECD - Organization for Economic Co-operation and Development. *Measuring student knowledge and skills:* a new framework for assessment. Paris: OCDE, 1999.

OECD - Organization for Economic Co-operation and Development. *Assessing Scientific, Reading and Mathematical Literacy:* a framework for PISA 2006. Paris: OCDE, 2006.

OECD - Organization for Economic Co-operation and Development. *PISA 2009 Assessment Framework*: Key competencies in reading, mathematics and science. Paris: OCDE, 2009.

OECD - Organization for Economic Co-operation and Development. *PISA 2012 Assessment and Analytical Framework:* Mathematics, Reading, Science, Problem Solving and Financial Literacy. Paris: OCDE, 2013.

OECD - Organization for Economic Co-operation and Development. *PISA 2015 Results* (Volume I): Excellence and Equity in Education, 2016. http://dx.doi.org/10.1787/888933432623

OECD - Organization for Economic Co-operation and Development. *PISA 2015 Assessment and Analytical Framework*: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving. Paris: OCDE, 2017.

OECD - Organization for Economic Co-operation and Development. *PISA 2022 Mathematics framework (draft).* November 2018. Paris: OCDE, 2018. https://pisa2022-maths.oecd.org/files/PISA%202022%20Mathematics%20Framework%20Draft.pdf

OECD - Organization for Economic Co-operation and Development. *PISA 2018 Assessment and Analytical Framework.* Paris: OCDE, 2019a.

OECD - Organization for Economic Co-operation and Development. *PISA 2018 Results (Volume I)*: What Students Know and Can Do. Paris: OCDE, 2019b.

OECD - Organization for Economic Co-operation and Development. *PISA 2024 Strategic Vision and Direction for Science.* Paris: OCDE, 2020. https://www.oecd.org/pisa/publications/PISA-2024-Science-Strategic-Vision-Proposal.pdf

OECD - Organization for Economic Co-operation and Development. *Beyond PISA* 2015*:* A longer-term strategy of PISA. [s.d.] http://www.oecd.org/pisa/pisaproducts/Longer-term-strategy-of-PISA.pdf

OECD - Organization for Economic Co-operation and Development. *Programme for International Student Assessment* (PISA). Paris: OCDE, 2023. https://www.oecd.org/Pisa/38709396.pdf

ONU – Organização das Nações Unidas. Agenda 2030. *Objetivo de Desenvolvimento Sustentável 4, ODS 4*: Educação de qualidade, 2015.

TEODORO, António. A OCDE e o sonho de uma governação mundial da educação: pressupostos e análise crítica. *In:* GONZÁLEZ-DELGADO, Mariano; FERRAZ, Manuel; MACHADO-TRUJILLO, Cristian. *Transferencia, transnacionalización y transformación de las políticas educativas (1945-2018).* Salamanca: FahrenHouse, 2020. p. 283-292.

UNESCO - United Nations Educational, Scientific and Cultural Organization. *Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos del tercer y cuarto grado de la educación básica*: informe tecnico. UNESCO, 2001

UNESCO - United Nations Educational, Scientific and Cultural Organization*. Los Aprendizajes de los estudiantes de América Latina y el Caribe*: primer reporte de los resultados del Segundo Estudio Regional Comparativo y Explicativo (SERCE). UNESCO, 2008.

UNESCO - United Nations Educational, Scientific and Cultural Organization. *Informe de resultados TERCE*: logros de aprendizaje. UNESCO, 2016.

JOÃO LUIZ HORTA NETO
PhD in Social Policy, Universidade de Brasília (UNB), Brasília, Distrito Federal, Brazil; Researcher, Instituto Nacional de Estudos e Pesquisas Educacionais "Anísio Teixeira" (INEP), Brasília, Distrito Federal, Brazil.

SUPPORT/FINANCING

There was no funding.

RESEARCH DATA AVAILABILITY

All data was generated/analyzed in this article.

HOW TO CITE THIS ARTICLE

HORTA NETO, João Luiz. Weaknesses in PISA, a global protagonist in educational policies. *Educar em Revista*, Curitiba, v. 40, e93646, 2024. https://doi.org/10.1590/1984-0411.93646

This article was reviewed by TIKINET Academic – E-mail: comercial@tikinet.com.br. After being designed, it was submitted for validation by the author(s) before publication.

Received: 02/21/24
Approved: 08/08/24