

ARTICLE

BI-MONTHLY LARGE SCALE ASSESSMENT: AN ANALYSIS OF AVALIAÇÃO DA APRENDIZAGEM EM PROCESSO AND PROVA PAULISTA¹

GUILHERME LUIS DESIDERIO¹

ORCID: <https://orcid.org/0009-0007-1040-1130>

<guilherme.desiderio@usp.br>

SANDRA ZAKIA SOUSA¹

ORCID: <https://orcid.org/0000-0001-5171-8301>

<sanzakia@usp.br>

¹ Universidade de São Paulo. São Paulo, São Paulo (SP), Brasil.

ABSTRACT: This article analyses the trajectory and purposes of *Avaliação da Aprendizagem em Processo* and *Prova Paulista* in the public school system in the state of São Paulo between 2011 and 2024. Through documentary research and literature review, it investigates how these bi-monthly large-scale assessments were implemented, their impacts on teaching practices, and their potential contributions to improving education quality. The analysis reveals an ongoing transformation of these assessments over time, marked by the expansion of the evaluated grade levels and subjects, increased frequency, change in purposes, broader objectives, digitalization and standardization of administration, dissemination of results to educational management bodies, and the use of data in accountability policies. The study concludes that *Avaliação da Aprendizagem em Processo* and *Prova Paulista* have contributed to the consolidation of a standardized teaching model, guided by logics of control and accountability, which tends to undermine teachers' professional autonomy and restrict school's role as a space for students' full development.

Keywords: large-scale assessment, managerialism, educational management, teaching practice.

AVALIAÇÃO EXTERNA BIMESTRAL: ANÁLISE DA AVALIAÇÃO DA APRENDIZAGEM EM PROCESSO E DA PROVA PAULISTA

RESUMO: O presente artigo analisa a trajetória e os propósitos da Avaliação da Aprendizagem em Processo e da Prova Paulista na rede estadual de São Paulo, entre os anos de 2011 e 2024. Por meio de pesquisa documental e revisão bibliográfica, investiga-se como essas avaliações externas bimestrais foram implementadas, seus impactos no trabalho docente e suas possíveis contribuições para a melhoria da qualidade da educação. A análise revela um processo de transformações das avaliações ao longo do tempo, marcado pela ampliação dos anos e componentes curriculares avaliados, aumento da frequência de aplicação, mudança de finalidade, alargamento dos objetivos, digitalização e padronização da aplicação,

¹ Article published with funding from the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* - CNPq/Brazil for editing, layout and XML conversion services.

disponibilização dos resultados para a gestão educacional, além do uso dos dados em políticas de responsabilização. Conclui-se que a Avaliação da Aprendizagem em Processo e a Prova Paulista vêm consolidando um modelo de ensino padronizado, orientado por lógicas de controle e responsabilização, que tende a esvaziar o trabalho docente e a restringir o papel da escola como espaço de formação integral dos estudantes.

Palavras-chave: avaliação externa, gerencialismo, gestão educacional, trabalho docente.

EVALUACIÓN EXTERNA BIMESTRAL: ANÁLISIS DE LA AVALIAÇÃO DA APRENDIZAGEM EM PROCESSO Y DE LA PROVA PAULISTA

RESUMEN: El artículo analiza la trayectoria y los propósitos de la *Avaliação da Aprendizagem em Processo* y de la *Prova Paulista* en la red estatal de educación de São Paulo, entre los años 2011 y 2024. A través de investigación documental y revisión bibliográfica, se investiga cómo se implementaron estas evaluaciones externas bimestrales, sus impactos en el trabajo docente y sus posibles contribuciones a la mejora de la calidad educativa. El análisis revela un proceso de transformaciones de las evaluaciones a lo largo del tiempo, caracterizado por la ampliación de los años y componentes curriculares evaluados, el aumento de la frecuencia de aplicación, el cambio de finalidad, la expansión de los objetivos, la digitalización y estandarización de la aplicación, la disponibilidad de los resultados para la gestión educativa, además del uso de los datos en políticas de responsabilización. Se concluye que la *Avaliação da Aprendizagem em Processo* y la *Prova Paulista* vienen consolidando un modelo de enseñanza estandarizado, orientado por lógicas de control y responsabilización, que suele vaciar el trabajo docente y restringir el papel de la escuela como espacio de formación integral de los estudiantes.

Palabras clave: evaluación externa, gerencialismo, gestión educativa, trabajo docente.

INTRODUCTION

Since 2011, the São Paulo State Department of Education (Seduc-SP-*Secretaria de Educação*) has applied a bimonthly external assessment instrument in state schools. Initially called the Learning in Process Assessment (AAP-*Avaliação da Aprendizagem em Processo*), this instrument has been modified over time, being replaced in 2023 by the Paulista Test (*Prova Paulista*).

The creation of the AAP and the *Prova Paulista* is part of a broader movement to implement initiatives for evaluating education systems and networks by federated entities, which have been used as instruments both to induce improvements in the quality of education and for the management of educational policies (Machado, Alavarse, Arcas, 2015). Within the paradigm shift in management that has taken hold of public administration in recent decades, in the context of the managerial reform of the State (Bresser-Pereira, 2000), external and large-scale assessments were the first results-based management instruments to be introduced in education, allowing the development of performance indicators and the adoption of accountability mechanisms (Segatto; Abrucio, 2017).

In Brazil, this process began with the Basic Education Assessment System (Saeb-*Sistema de Avaliação da Educação Básica*), created in 1990 to improve learning through the public report of results (Gatti, 2009). Since then, states and municipalities have implemented their systems, consolidating external assessments as central tools in educational management.

Subnational entities cite several motivations for developing their assessment systems, with the primary goal being the use of results by schools and teachers to improve pedagogical practices (Bauer et al., 2015; Gimenes et al., 2013; Silva & Sousa, 2022). In this context, the establishment of these systems

reflects a desire for greater control over the assessment process. The aim is to ensure that results can be appropriated and utilized to formulate precise and effective strategies and actions (Brooke & Cunha, 2011).

The introduction of new assessments by federated entities aimed at disseminating results by school, as well as the census-based application of the Saeb and the creation of the Basic Education Development Index (Ideb-*Índice de Desenvolvimento da Educação Básica*), indicates a trend of using this data to promote accountability policies (Bonamino, Sousa, 2012). With the public report of the results of systems and schools, and the eventual adoption of incentive programs linked to performance, teachers, managers, and schools – and not just students – have become accountable for the results achieved (Fernandes, Gremaud, 2009).

The multiplication of external assessment initiatives reinforces the supposed belief that more measurements would be the solution to the problem of low educational quality. However, “is applying more tests an answer or, better yet, a solution to improve student performance on tests? Or, even, have large-scale assessments provided an improvement in the quality of education?” (Sousa, 2014, p. 410).

Considering this scenario, this work aims to analyze the trajectory and purposes of the AAP and the *Prova Paulista* from 2011 to 2024. To this end, their characteristics, transformations, and implications for teaching and teacher practice were investigated, as well as their possible contribution to improving the quality of education.

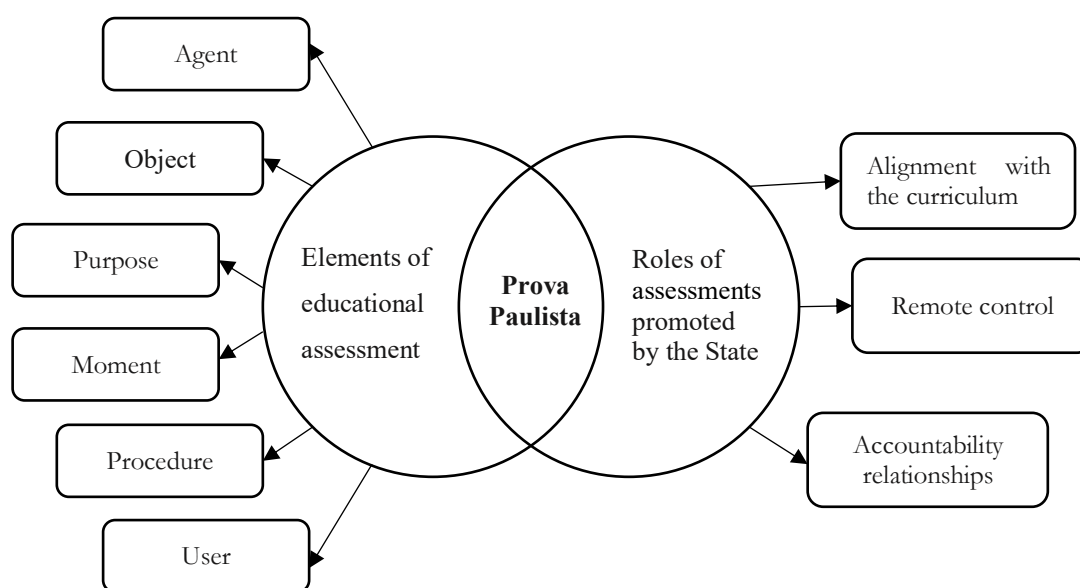
Thus, a documentary-based study was carried out, in which the documents were a source of data and information for understanding the phenomenon studied (Kripka; Scheller; Bonotto, 2015). The objective was to gather official documents about the *Prova Paulista* and its predecessor, the AAP, to characterize these assessment instruments and understand the main changes that occurred in their scope over time.

As Cellard (2008, p. 296) states, documents are considered to be anything that is “a vestige of the past.” In this sense, the documentary survey consisted of identifying communications, decrees, resolutions, news items, and videoconferences published by Seduc-SP and by regional education boards, referring to the AAP and the *Prova Paulista*, between 2011 and 2024. The temporal scope is justified by the fact that 2011 marks the year of creation of the AAP, and 2024 is the year immediately preceding the research. Complementarily, the educational policy of the State of São Paulo was also analyzed from 1995 onwards, the year in which a new management logic began to operate in the state network, aiming to understand the evolution of the reforms and programs implemented by the different administrations.

CONCEPTS AND ROLES OF EDUCATIONAL ASSESSMENT

According to Arredondo and Diago (2013, p. 27), “the variety of functions and purposes that are being incorporated into assessment increases its complexity and obliges authors to mark or point out its conceptual delimitation before entering into the development of any theme related to it”. In this sense, the theoretical framework used to analyze the AAP and the *Prova Paulista* in this work is articulated in two dimensions: the elements that allow characterizing educational assessment and the roles played by assessments promoted by the State, considering the context of State reform under the paradigm of new public management (NPM) (Figure 1).

Figure 1 - Dimensions of the theoretical framework



Source: Authors' elaboration (2025).

Elements of Educational Assessment

Assessment can be understood as a process involving the collection of information, the issuance of a value judgment based on predefined criteria, and the subsequent making of a decision (Lukas Mujika, Santiago Etxeberria, 2009). Within this process, different components warrant the evaluator's attention, and their characterization helps define the concept of educational assessment (Alavarse, 2013). Among these are the agent, the object, the purpose, the moment, the procedures, and the users of the assessment results.

The *agent* of assessment is determined by the origin and location of those who exercise control over it. We can distinguish between *internal assessment* and *external assessment*. Generally, internal assessment refers to assessments initiated and conducted within the school by its own professionals, such as teachers assessing their own classrooms (Madaus, 1988). External assessment is controlled and/or required by entities outside the school, such as a department of education, the Ministry of Education, or a private agency (Madaus, 1988). This type of assessment is developed, organized, and conducted by individuals or groups that do not belong to the school.

The *object* refers to what will be assessed, considering the defined objectives of the assessment (Lukas Mujika, Santiago Etxeberria, 2009). In the context of the teaching-learning process, this object is generally the students' *learning* (Amaral, Alavarse, Silva, 2020). In external assessments, the focus may fall on the *school* or the *teacher* (Alavarse, 2013).

The *purposes* of the assessment are related to the use made of its results. Formative assessment “is an improvement strategy to adjust and regulate ongoing educational processes to achieve the planned goals or objectives” (Arredondo, Diago, 2013, p. 63). In this sense, it occurs in parallel with the activity being assessed (Lukas Mujika; Santiago Etxeberria, 2009), allowing students to identify their errors, successes, and difficulties, and enabling teachers to reformulate their teaching practices (Haydt, 1988).

Summative assessment “is applied at the end of a given period of time as proof of the achievements” (Arredondo, Diago, 2013, p. 63), such as the end of a teaching unit, a term, a semester, or

the end of the school year. The objective of this assessment is to offer a cumulative view of student learning, having a character of control over the results or final products obtained, leaving for a subsequent moment the implementation (or not) of possible interventions that may support the improvement of learning (Lukas Mujika, Santiago Etxeberria, 2009).

In addition to formative and summative assessments, many authors point to the existence of another type of assessment: *diagnostic* assessment (Haydt, 1988; Hadji, 2001; Arredondo, Diago, 2009). This assessment is carried out at the beginning of a course, academic period, or teaching unit, with the objective of “ascertaining whether the students [...] possess the essential knowledge and skills for new learning” (Haydt, 1988, p.16-17), allowing the teacher to adjust the teaching plan according to their needs (Hadji, 2001).

Assessments can occur at different *moments* in the teaching-learning process, which leads to the definitions of *initial*, *process*, and *final* assessments (Lukas Mujika; Santiago Etxeberria, 2009). The initial assessment is conducted before the teaching-learning process begins, typically at the start of a teaching unit, term, semester, or school year. Process assessment aims to regulate and improve the ongoing educational process. It is carried out continuously throughout the teaching-learning process, thus relating to formative assessment. The final assessment is carried out at the end of the teaching-learning process to verify the achievement of the planned objectives. Therefore, it collects and analyzes information after the period foreseen for the achievement of learning, often being identified with summative assessment.

An important aspect of the assessment process is the *procedures* used to obtain information about the object being assessed. According to Arredondo and Diago (2013, p. 265), “each moment in the assessment process requires a specific *technique* that aligns with the functions it performs and the purposes it aims to fulfill, and the necessary *tools*, such as specific resources or means used to obtain it”. Thus, the instruments are chosen considering the objectives of the assessment and the intended uses of its results, so that the same instrument can be used for different assessment purposes.

The procedures used to assess student learning can be divided into two groups: *standardized* and *non-standardized*. Russel and Airasian (2014, p. 22-23) consider that an assessment is standardized when

[...] is administered, graded, and interpreted in the same way for all students, regardless of where or when they were assessed. Standardized assessments should be given to students from different classrooms, but always under the same conditions, with the same instructions, and in the same amount of time as all other students taking the test at that moment. Furthermore, test results will be graded and interpreted in the same way for all students.

In this sense, the use of standardized procedures aims to compare the results obtained by students, in addition to allowing the assessment of the performance of students from different educational institutions using the same instrument (Russel; Airasian, 2014).

Finally, an additional important element is the *user* of the assessment results (Alavarse, 2013), which relates to the purposes and actors involved in the assessment process. For example, in the case of internal assessments, users vary according to the goals established for the assessment: if it is a formative assessment, the main users are the students and the teacher themselves; in a summative assessment, they may include families and the school's leadership team. In external assessments, users of the results may include public policy makers, the media, and society in general, as well as professionals from the schools.

Roles of State-Promoted Assessments

State-promoted educational assessments have gained centrality in recent decades, becoming “one of the structuring axes of contemporary educational policies” (Afonso, 2000, p. 9). They have become instruments for controlling and regulating schools, monitoring their efficiency based on students' academic performance (Oliveira, 2015).

Historically used to certify students or diagnose problems in educational systems, external and large-scale assessments have been increasingly employed to hold schools, principals, and teachers accountable for student learning outcomes, as well as to monitor the distribution of standardized curricula by educational institutions. With the adoption of NPM principles for educational management, these assessments have assumed the role of political instruments that support the implementation of principles that shaped global education reforms, especially standardization, decentralization, and accountability (Verger, Parcerisa, Fontdevila, 2018).

In the context of standardization, which refers to establishing common and measurable learning standards, assessments are used by governments to monitor whether the official curriculum is being implemented by schools. On the other hand, decentralization can occur when educational management responsibilities are transferred from the central government to regional, state, or local authorities. It can also involve assigning organizational and pedagogical duties to schools and granting managerial authority to school administrators. In these cases, assessments serve as tools for remote state oversight of the various authorities and actors within the educational system (Verger, Parcerisa, Fontdevila, 2018).

In the case of accountability, which implies holding teachers and schools responsible for their actions and results, it is assumed that the adoption of incentives and/or sanctions will lead to improvements in teaching and learning. From this perspective, assessments are used to generate accountability relationships of an administrative nature – linked to the consequences and incentives distributed according to the results obtained – or of a market nature – related to the awareness and school competition fostered by the public report of results (Verger, Parcerisa, Fontdevila, 2018).

Thus, external and large-scale assessments contribute to shifting the exercise of control over educational institutions, which ceases to be an internal bureaucratic control, based on ensuring compliance with rules, norms, and processes, and becomes an external managerial control, carried out remotely by the State, with a focus on achieving results (Duarte, 2020). This control is enhanced by the adoption of technologies, platforms, and digital communication and information processes, which expand the educational administration's capacity to control pedagogical work and schools (Lima, 2012).

It became possible to conduct systematic surveillance, based on pre-defined and unilaterally imposed categories, which shape school realities, control time and space, generalize methodologies and work processes, and promote the standardization and measurement of professionals (Lima, 2012, p. 148).

Therefore, the control exerted by assessment ends up influencing the curriculum and the organizational forms of schools (Lima, 2011), as well as the work and personal and professional identity of teachers (Anderson, 2017). In this regard, Freitas (2014, p. 1092) highlights that external assessment has been used by the State as a means of controlling and strengthening internal classroom assessment processes, with the aim of “preserving and increasing control over the objectives, content and even the methods of the school”, guiding the entire pedagogical process, from beginning to end. As the author

observes, “assessment, regardless of content and method, imposes a ‘model of reasoning’, a ‘way of thinking’, a way for the teacher to relate to the student, embedded in their specific practices” (Freitas et al., 2009, p. 25).

Thus, external and large-scale assessment strains the relationships of trust between the actors involved in the pedagogical process by imposing, on the one hand, pressures and accountability for results and, on the other, a workflow determined from outside to inside the school. As a result, teachers are pressured to follow an imposed pace, which limits the possibility of addressing students' learning difficulties, contributing to accentuating performance inequalities and processes of exclusion (Freitas, 2014, p. 1100).

BIMONTHLY EXTERNAL ASSESSMENTS OF THE STATE OF SÃO PAULO

The creation of the AAP and the *Prova Paulista* is situated in a context in which the educational policy of the State of São Paulo began to adopt “a conception of quality of school education linked to the measurement of learning through external and standardized assessments” (Jacomini et al., 2022, p. 27), an axis that guided the development of programs, projects, and initiatives of different administrations since 1995.

The restructuring of the Department of Education initiated in 1995 marks a significant milestone for the state education system. This restructuring introduced a set of guidelines designed to enhance the rationality and efficiency of the system, based on the principles and concepts of the New Public Management (NGP) (Jacomini, Nascimento, & Stoco, 2023). These guidelines emphasized the importance of assessment in managing the state network, positioning it as a central component of the system's operations.

[...] a “sine qua non” condition for the State to fulfill its equalizing role, insofar as it provides data to act in overcoming existing inequalities among São Paulo schools. Furthermore, the results of school performance should be widely disseminated so that both the school staff and the user community are able to identify the position of their school within the group of schools in their district, neighborhood, and municipality. This will enable the school to seek diversified forms of action, to improve school results. On the other hand, it will also allow the population access to information, so that they can monitor, participate in, and demand quality in the services that should be provided to them (São Paulo, 1995, p. 10).

Not surprisingly, in 1996, the state government created the São Paulo State School Performance Assessment System (Saresp) with the objective of verifying the performance of elementary (EF) and high school (EM) students through external assessments (São Paulo, 1996). The introduction of this system aimed to support schools, education boards, and the Department of Education in making decisions about school planning and educational policy for the state network. Until the 2000s, Saresp had a diagnostic character, since it was applied at the beginning of the school year. From then on, the exam began to be carried out at the end of the year, becoming an instrument for measuring student results, with a great influence on the development of the pedagogical work of schools (Jacomini et al., 2022).

Starting in 2008, the results of Saresp began to be part of the São Paulo State Education Development Index (Idesp), an indicator taken as an expression of the quality of state schools that allows the establishment of individual performance goals for each school (São Paulo, 2008b). In 2009, Idesp also began to be used for the payment of bonuses to state education professionals (São Paulo, 2009), linking their compensation to the achievement of results and the fulfillment of goals.

In 2007, alongside the development of Idesp and the introduction of bonuses, Seduc-SP implemented the São Paulo State Curriculum Proposal for Elementary and High School as part of the *São Paulo Faz Escola* Program. This proposal established a common curriculum for all schools in the state network (São Paulo, 2008a). The Department of Education justified the adoption of a centralized curriculum due to the low performance of students in external assessments, particularly in Saresp (Ramos, Goulart, & Jacomini, 2023). The expectation was that by providing “precise guidelines on what and how to teach and controlling the teacher's work, students would be able to demonstrate improved performance in these external assessments” (Jacomini et al., 2021, p. 10).

The articulation between curriculum, Saresp, Idesp, and performance-based bonuses reflects the managerial logic of results-based management that has been consolidated in the São Paulo state network since 1995, configuring itself as a mechanism for controlling and managing teaching work (Bernardes, 2018). As a result of the deepening of this management process, the creation of the AAP and the Prova Paulista represented the introduction of yet another instrument aimed at monitoring the pedagogical work carried out within schools (Rodrigues, 2018).

Learning in Progress Assessment

Created in 2011 by Seduc-SP, the AAP was an assessment instrument intended to diagnose student's learning on an individual basis, with the aim of providing teachers and schools with quick results to support planning and replanning of instructional activities, aiming to promote learning for all students. The analysis of documents on the AAP reveals that, since its creation, the assessment has undergone a series of changes in its structure and use. Based on these transformations, as part of this research, the implementation of the different editions of the AAP was organized into five phases, detailed below.

Phase 1 – Pilot Action (2011)

The first phase corresponds to the initial edition of the AAP, carried out in 2011 and considered a pilot action by the Department of Education. Launched through a videoconference, the test was presented as a diagnostic assessment, without any link to external assessments, aimed at "monitoring student learning at the beginning of a cycle and/or level of education" (São Paulo, 2011). In addition to not being used to compose students' grades, the test also differed from Saresp, as its results would not be used to pay bonuses or calculate indicators such as Idesp.

That year, the AAP was intended for the 6th grade of middle school and the 1st year of high school, chosen because they represent transition years between stages of schooling. In the words of the Deputy Secretary of Education of São Paulo, these grade levels were selected because:

In many inland cities, especially, and even in some cities in Greater São Paulo, the 6th grade is when the state school system receives the student, because they attended municipal schools up to the 5th grade. Therefore, it is very important to know exactly how this student is doing when they enter our system. The same applies to the 1st year of high school; that is, they have completed a stage of basic education, the middle stage, which is the 9-year elementary school, and it is also important that upon entering high school, we have clarity about what they have learned and what they have not learned. So, it is a working tool for the teacher (São Paulo, 2011).

Due to delays in making Saresp data available to schools and the fact that many students changed schools each year, Seduc-SP aimed to develop a tool that would allow for the quick identification and addressing of students' needs. Additionally, it was anticipated that, in the future, teachers would be

able to select questions from an item bank to create tests based on the material they were teaching their students during a specific period.

The test assessed Portuguese and Mathematics, and it was expected that its results would complement the monitoring already carried out by teachers. Although it was developed externally by Seduc-SP, it was up to the schools to administer, correct, use the results, and develop the action plan. It was even recommended that the subject teachers administer the test in their classes to minimize the impact on the school routine. The regional and central bodies were responsible for monitoring the execution of the action plan, according to the reality and conditions of each teaching unit.

The assessment instruments generally included multiple-choice questions, open-ended questions, and a written production component. For both subjects, the skills that students were expected to have developed throughout their educational journey, leading up to the year prior to the assessment, were taken into account. Although the focus was on content deemed essential, Seduc-SP acknowledged that the instruments were limited and inadequate for evaluating the full scope of the school's work. As stated, "the school does much more than an externally constructed instrument is capable of observing" (São Paulo, 2011).

Phase 2 – Expansion of Participating Grade Levels (2012 to 2014)

The second phase of AAP implementation was characterized by the progressive expansion of the group of students participating in the assessment. In 2012, the 7th grade and the 2nd year of high school were included; in 2013, students from the 6th and 9th grades and the 3rd year of high school also began to participate; and, in 2014, in the 7th edition, all grade levels of elementary, middle and high school (except the 1st grade) were incorporated into the assessment.

Between 2012 and 2014, the AAP was conducted twice a year, always at the beginning of each semester. According to the guiding documents, its purpose was exclusively to diagnose student's learning, with the aim of supporting the monitoring of pedagogical proposals, school planning, and the development of action plans for learning recovery processes.

The tests, which are exclusively diagnostic in nature, constitute investigative instruments of learning for the subsequent mobilization of procedures, attitudes, and concepts by the school and in the classroom, aiming at the elaboration and execution of plans intended to provide immediate support to overcome students' difficulties (São Paulo, 2012, emphasis added).

At this stage, the AAP continued to assess Portuguese, Mathematics, and writing skills of the classes participating in each edition. In general, Portuguese instruments consisted only of multiple-choice items, while the Mathematics instruments combined multiple-choice items with open-ended questions.

In two editions (2nd and 7th), the guidelines specified that the tests should be administered before the semester planning meetings so that their results could be analyzed and discussed during those meetings. Furthermore, from the 6th edition onwards, the AAP began to be seen as a tool to support the continuous progression system within the learning cycles, fulfilling the function of "identifying the knowledge not acquired by students to support the promotion of pedagogical reinforcement and/or recovery interventions" (São Paulo, 2013).

The official notices stipulated that the assessment instruments should be administered and corrected by the subject teachers themselves, preferably in double classes. They emphasized the

importance of prioritizing qualitative analyses of the results, even though they could also be quantified. Except for the 6th edition, the week of application of the tests was determined by Seduc-SP.

From the 7th edition onwards, teachers were instructed to record the AAP results in the Assessment Results Monitoring System (*Sara-Sistema de Acompanhamento dos Resultados de Avaliações*). In that edition, registration was mandatory only for the 6th grade of elementary school, being optional for the others. According to Seduc-SP, this system made it possible to view results and carry out comparative analyses, contributing to the implementation of plans for pedagogical intervention by teachers. It also allowed the education boards to monitor data entry and plan support actions for schools (São Paulo, [2017]).

Phase 3 – Broadening the Objectives of the Assessment (2015)

The third phase of AAP implementation was marked by a broadening of the objectives assigned to the assessment. In the 8th edition (2015), although the emphasis on the *exclusively diagnostic* nature of the test was maintained, the guidelines began to highlight the importance of the assessment in supporting “the teacher in the development of the curriculum in the classroom” (São Paulo, 2015a). In the following edition, the emphasis on diagnostic *exclusivity* disappeared, and the monitoring of the school's curriculum development was explicitly incorporated among the assessment objectives:

The AAP assessments, of a diagnostic nature, **in addition to monitoring curriculum development**, also constitute investigative instruments of student learning, in terms of their skills and abilities, aiming to support the progression of learning that has not yet been consolidated, through the elaboration and execution of plans developed by the teacher **for the development of the curriculum in the classroom** (São Paulo, 2015b, emphasis added).

A notable change observed in relation to previous editions was the inclusion of the 1st grade of elementary school in the 9th edition of the AAP. Beginning in 2015, the assessment came to encompass all elementary, middle and high school grades. In that edition, it also ceased to be mandatory for the test to be administered by the subject teacher in middle and high school. The assessments continued to focus on Portuguese, Mathematics, and writing, featuring both multiple-choice items and open-ended questions.

Following the trend of the previous phase, from 2015 onwards, teachers of all grades were required to register the AAP results in Sara, in order to allow the monitoring of class performance by teachers, schools and regional education boards.

Phase 4 – Change in Periodicity (2016 to 2019)

The fourth phase was marked by a change in the periodicity of the AAP application. Until 2015, the test was administered at the beginning of each semester. However, starting in 2016, it began to be conducted at the end of the first, second, and third bimesters. This new schedule was accompanied by the expectation that teachers would use the results to plan and adjust their curriculum for the following bimester.

In 2016, the AAP began to be referenced in the Processual Assessment Matrix (*MAP-Matriz de Avaliação Processual*), which organized content and skills by bimester, subject, and year/grade of Elementary, Middle and High School. During this period, the use of multiple-choice items was consolidated for assessing Portuguese and Mathematics from 4th grade of elementary school through the 3rd year of high school. In the 1st, 2nd, and 3rd grades of elementary school, open-ended questions

prevailed, although some editions also included multiple-choice items. Written composition was restricted to 4th and 5th grades of elementary school.

Between the 10th and 15th editions (2016–2017), the guidance indicated that teachers were responsible for administering the AAP in the early years of elementary school; without this requirement for the other grade levels. However, with the 16th edition (2017), the guidance changed to state that the subject teacher must administer the assessment in all classes. Additionally, the official notices specified the period for taking the test and recording the results. Starting with the 21st edition (2018), regional education boards were allowed to choose two days on which the AAP would be administered in the schools under their jurisdiction.

In all editions, teachers were instructed to record students' answers in Sara. From the 11th edition (2016), these data began to be incorporated into the *Plataforma Foco Aprendizagem*, enabling teachers, schools, and regional education boards to access and use of the results. Starting with the 15th edition (2017), the tests began to include an answer sheet on the cover for students to mark their responses, allowing for automated reading via the *Leitor Resposta* app.

Phase 5 – Application in Digital Format (2020 to 2022)

The fifth phase of AAP implementation was impacted by the COVID-19 pandemic, which brought changes in the application of the tests. Despite the health crisis, which led to the suspension of classes and the introduction of emergency remote teaching, Seduc-SP maintained the administration of the AAP. In the 1st and 2nd bimesters of 2020, the tests were sent to students along with other printed materials distributed by schools to families. Students were expected to complete the AAP at home and record their answers in online forms created by the schools.

The test for the 3rd bimester of 2020, in turn, represented a significant change: it was the first time that the assessment was administered through a digital platform specifically designed for this purpose, in partnership with the Center for Public Policy and Education Assessment (CAEd-*Centro de Políticas Públicas e Avaliação da Educação*) of the Federal University of Juiz de Fora. This is the Formative Assessment and Activities Platform, also known as the São Paulo Platform or CAEd Platform.

Until the adoption of the CAEd Platform, only the assessments from the 6th grade of elementary school to the 3rd year of high school consisted exclusively of multiple-choice items. In the early years of elementary school, open-ended questions predominated, and a written composition was required from students in the 4th and 5th grades. With the shift to the digital platform, all tests – for all grade levels – began to consist only of multiple-choice items, and the written composition for students in the 4th and 5th grades was eliminated.

The use of the CAEd Platform was seen by Seduc-SP as an opportunity to improve the monitoring of network results by enabling the visualization of participation and performance data by student, class, school, and regional education board, aiming to support pedagogical planning. These results were automatically integrated into the Digital School Secretariat (SED-*Secretaria Escolar Digital*). In parallel with the digital application, Seduc-SP also made a PDF version of the assessment available.

Beginning in the 1st semester of 2021, the AAP underwent another change and expanded the areas of knowledge assessed, also including Human Sciences and Natural Sciences, in addition to Portuguese and Mathematics. The latter two continued to be administered to all grade levels, while the new areas were applied from the 4th grade of elementary school to the 3rd year of high school. In 2021, these tests were called “Intermediate AAPs” and, in 2022, “Digital Sequences”.

Prova Paulista

Starting in 2023, the *Prova Paulista* replaced the AAP as the bimonthly assessment of the state school system, becoming one of the initiatives of the “Classroom of the Future” Program, from Seduc-SP. In this new model, the exam came to cover students from the 5th grade of elementary school to the 3rd year of high school, excluding classes from the 1st to the 4th grade of elementary school. According to Seduc-SP, the *Prova Paulista* represented an innovation by expanding the curricular components assessed and by adopting a new digital application tool: the São Paulo Education Media Center (CMSP- *Centro de Mídias da Educação de São Paulo*) app.

Unlike the AAP, the *Prova Paulista* includes different curricular components. In 5th grade of elementary school, the subjects assessed are Portuguese, Science, Mathematics, History, and Geography. From 6th to 9th grade, English was added beginning in the 2nd bimester of 2023. In high school, the test assesses the components of Portuguese, English, Biology, Physics, Chemistry, Mathematics, Philosophy, Sociology, History, and Geography. In 2024, the *Itinerários Formativos* were also included, in addition to the subjects of basic general education.

The administration of the *Prova Paulista* was organized over two days for each edition, both in 2023 and 2024, following a predefined schedule. The 5th grade tests contained 30 items for each day; from the 6th to the 9th grade, 40 items for each day; and from the 1st to the 3rd year of high school, there were 45 items per day. In 2023, all items were multiple choice, with no open-ended questions or writing assessments. In 2024, the tests began to include single-answer and multiple-answer items – with up to three correct alternatives.

In the 1st and 2nd bimester editions of 2023, the *Prova Paulista* was based on the contents of the São Paulo Curriculum (*Currículo Paulista*). From the 3rd bimester onwards, however, the reference shifted to the Digital Material (*Material Digital*)². This change coincided with an alteration in the objective assigned to the assessment, which once again included the perspective that the test also “monitors the development of the curriculum, in all areas, [...] through the application of standardized instruments” (São Paulo, 2023a).

According to Seduc-SP, conducting the *Prova Paulista* through the CMSP app aimed to facilitate its application and provide faster feedback on results, enabling teachers to make immediate interventions in both the learning process and lesson planning. Performance data were made available on the Escola Total Panel (*Painel Escola Total*), a business intelligence platform that offers access to results for teachers, schools, regional education boards, and the central office, allowing their use for decision-making. Although the *Prova Paulista* is a digital assessment, the instruments were also made available in PDF format along with an answer key document containing pedagogical observations for each item.

In 2024, the results of the *Prova Paulista* came to be used for different purposes, such as: composing the bimonthly evaluation of school principals, carried out by the regional education boards (São Paulo, 2024a); integrating students’ bimonthly grades (São Paulo, 2024d); and serving as a classification criterion for other initiatives of the Department of Education, such as the Mathematics Olympiad of São Paulo state schools (São Paulo, 2024b) and the semester recovery program (São Paulo, 2024c).

² Digital Material is another initiative of the “Classroom of the Future” Program and refers to fully digital classroom material that can be used in PDF format or edited by the teacher before being taken to the classroom (São Paulo, 2023b).

BIMONTHLY EXTERNAL ASSESSMENT IN THE SERVICE OF MANAGING TEACHERS' WORK?

Since 2011, state schools have been required to administer another external assessment instrument: initially the AAP and, currently, the *Prova Paulista*. Table 1 presents a summary of the characteristics that marked each phase of implementation of these two assessments between 2011 and 2024, whose elements were progressively incorporated in subsequent stages.

Table 1 – Summary of the implementation phases of the AAP and the *Prova Paulista*

Test	Year - Phase	Main characteristics and changes observed
AAP	2011 – Phase 1. Pilot Action	Frequency: Semiannual (administered at the beginning of the semester) Objective: To identify student's learning Grade levels involved: 6th grade of Elementary School and 3rd year of High School (transition years) Subjects assessed: Portuguese and Mathematics
	2012 to 2014 – Phase 2. Expansion of Participating Grade Levels	Grade levels involved: gradual expansion from the 2nd grade of Elementary School to the 3rd year of High School Entry of results for 6th grade into Sara
	2015 – Phase 3. Broadening the Objectives of the Assessment	Objective: Incorporation of the expectation of monitoring curriculum development in the classroom Grade levels involved: Inclusion of 1st grade of Elementary School Entry of results for all grade levels into Sara
	2016 to 2019 – Phase 4. Change in Periodicity	Frequency: Change to bimonthly (administered at the end of the bimester) Developed based on MAP Inclusion of an answer sheet on the cover of the test
	2020 to 2022 – Phase 5. Application in Digital Format.	Objective: To identify students' learning. Administration: In digital format via the CAEd Platform. Items become exclusively multiple-choice. Subjects assessed: Inclusion of Humanities and Natural Sciences from 4th grade of Elementary School to the 3rd year of High School in 2021 and 2022.
Prova Paulista	2023 and 2024 – Expansion of the subjects assessed	Objective: Incorporation of the expectation of monitoring curriculum development in the classroom Administration: In digital format via the CMSP app Grade levels involved: 5th grade of Elementary School through 3rd year of High School Subjects assessed: All except Art and Physical Education Standardization of the administration process Development based on Digital Material Visualization of results on the Escola Total Panel Use of results in accountability policies

Source: Authors' elaboration using data from the research. (2025).

An analysis of the history of these two instruments reveals transformations that point to a greater incidence of external assessment on teachers' work, resulting in intensified monitoring of results

and increased control over the teaching-learning process. The main trends and patterns observed are discussed below.

Expansion of the Grade Levels Involved

The AAP was originally designed to assess the skills and knowledge of students transitioning to a new stage of schooling, specifically those entering 6th grade in elementary school and the 1st year of high school. This gave the AAP a very specific diagnostic purpose. However, with the gradual expansion of participating grade levels, the Department of Education started to assess the performance of an increasingly larger portion of the state education system. This shift distorted the AAP's original intentions and led to a broader logic of assessment and control. Consequently, the expansion of the assessed classes indicates that both the AAP and the *Prova Paulista* have taken on additional goals. These now include not only supporting teachers in transitional grades but also monitoring the performance of schools throughout primary and secondary education.

Expansion of the Areas of Knowledge Assessed

The scope of the *Prova Paulista* highlights the growing influence of external assessments in the state education system. While the inclusion of additional subjects beyond Portuguese and Mathematics might suggest a shift away from focusing solely on these two areas — common in external assessments — it has, in reality, led to the test becoming a central aspect of the pedagogical organization in schools. Its connection with the Digital Material, introduced in the context of the platformization of education in São Paulo, further emphasizes the standardization of classroom practices and increased managerial control over teaching activities. The material outlines what should be taught in each class, while the test specifies what should be assessed at the end of the term. Therefore, the expansion of assessed knowledge areas, combined with the use of structured instructional materials, indicates that the Seduc-SP aims to enhance control over teaching practices, which ultimately contributes to the decharacterization, de-intellectualization, and loss of autonomy for teachers.

Increased Frequency of Application

Unlike Saresp, whose punctual administration at the end of the year allows for a certain degree of flexibility in teachers' work, the bimonthly implementation of the AAP and the *Prova Paulista* intensified the impact of external assessment on teachers' pedagogical practice, potentially restricting their autonomy in planning and proposing learning experiences for their classes (Jacomini, Nascimento, Stocco, 2023).

In a study with fifth-grade elementary school teachers, Pereira (2023) found that their assessment practices were influenced by Saresp and the AAP. As a result, most classroom assessments mirrored the format of these external exams, primarily aiming to prepare students for these tests. While teachers indicated that they used AAP results to monitor student learning, this instrument acted as a regulatory assessment, shaping both the content and methods of instruction.

Ravagnani Pinto (2016) identified in his study that sometimes teachers were unable to cover all the syllabus content before the AAP was administered, and the pressure to teach the content that would be tested in the following exam limited the time available to address the identified difficulties. According to the author, the instrument did not take into account the specificities of the school and its surroundings, and its bimonthly frequency disregarded the learning time and differences of the students.

In this context, the bimonthly application of an externally designed instrument tends to influence the assessment conducted in the classroom, having the potential to shape the way teaching is structured and how pedagogical interactions occur in the school. In addition, its bimonthly administration contributes to a schooling model that values the homogenization of processes, paths and results, to the detriment of the diversity of educational trajectories, by disregarding the different times and ways in which children and adolescents learn.

Change of Purpose

Between 2011 and 2015, the AAP was characterized as a diagnostic instrument, administered at the beginning of each semester. From 2016 onwards, however, although it officially maintained this purpose, the AAP began to be carried out at the end of the teaching-learning process, assessing the skills planned for the recently completed bimester.

This change in timing and frequency generated ambiguity regarding the nature of the instrument, reinforced by its original name. The expression “in process” suggests a formative purpose aimed at monitoring learning as it is being developed. Although the AAP was administered throughout the school year – which could justify the notion of being “in process” – and had the stated objective of identifying students' knowledge – which could justify the use of the term “diagnostic” –, its administration at the end of each bimester points to an eminently summative assessment logic, also maintained by the *Prova Paulista*.

In practice, since 2016, both the AAP and the *Prova Paulista* have been tests administered at the end of the bimester, after the conclusion of the teaching-learning cycle, aimed at verifying the development (or lack thereof) of the skills foreseen in the curriculum for the period. As Arredondo and Diago (2013) state, summative assessment can also serve as an initial assessment for the next instructional period, since the information obtained can support decision-making that reorients the teacher's work with their students. Thus, although the AAP and the *Prova Paulista* maintain a diagnostic rhetoric, their function in the state school system is closer to that of a summative assessment.

Expansion of the Assessment Objectives

Over the years, the AAP and the *Prova Paulista* have incorporated among their objectives the monitoring of curriculum implementation in the classroom – a curriculum that has been standardized and centralized by the Department of Education since 2007. This shift highlights an important redefinition of expectations regarding the role of the bimonthly external assessment. For Seduc-SP, it was not enough to simply establish a common curriculum to guide teachers' work: it became equally necessary to systematically and frequently monitor pedagogical practice through these tests.

As a result of this process, there has been a growing alignment between the taught curriculum and the assessed curriculum over time, with the bimonthly external assessment guiding the content taught in classrooms. This movement began with the MAP and has been deepened more recently with the development of the Digital Material. In this sense, the influence of assessment on the curriculum reveals that the AAP and the *Prova Paulista* have come to play a regulatory role in teachers' work, establishing a logic of curricular control that narrows the spaces for teacher autonomy and standardizes the teaching-learning process across state schools.

Dissemination of Results to Central and Intermediate Management Bodies

In its initial conception, the bimonthly external assessment had teachers and schools as the main users of its results. Although external actors were involved in some stages of the assessment process, such as the development of instruments and monitoring the implementation of action plans, the test results were not made public beyond the school environment, remaining under the exclusive control of teachers and schools.

Over time, however, this scenario changed. The requirement to record results in Sara and, later, their automatic availability on computerized platforms such as SED and Escola Total Panel, enabled monitoring of the performance of students, schools, and the state network in real time by the central and intermediate levels of educational management. This process opened space for the use of bimonthly external assessment as a remote-control instrument, in addition to enabling the adoption of accountability mechanisms based on results. In this way, assessment data came to fulfill not only a pedagogical function, but also a managerial one, being incorporated as a tool for governing the state school system.

Digitization of Test Administration

The administration of the bimonthly external assessment through the CAEd Platform and the CMSP app expanded the monitoring capacity of the central and intermediate management bodies, since data began to be collected, systematized, and automatically made available by computerized systems, intensifying supervision and control over schools' pedagogical work. With digitization, teachers were removed from the role of directing test administration and scoring, being relegated to an increasingly executory role in the process and a consumer of its results.

Moreover, digital administration raises questions about the validity of the data produced. Factors such as the possibility of consulting the internet or classmates, as well as technical failures in devices or connectivity, may compromise the reliability of results. These weaknesses may distort the reality represented by the collected data. As a result, teachers, schools, regional education boards, and the Department of Education might make decisions that do not accurately address the actual learning needs of students.

Standardization of Test Administration

The changes that have occurred throughout the editions of the AAP and the *Prova Paulista* demonstrate a gradual process of standardizing the administration of the bimonthly external assessment. This has made it easier for both central and intermediate management bodies to access and utilize assessment data. Firstly, the test is no longer administered by the teacher of the subject being assessed during their class time; instead, it is administered by any available teacher at the scheduled time. Secondly, the assessment tools have transitioned to using only multiple-choice items, which has helped standardize both the administration and grading processes. Thirdly, the tests are now conducted during a single designated testing window, determined by the Department of Education or the regional education board, which allows for more rigorous monitoring of the assessment process. Finally, the development of a computerized system has facilitated continuous tracking of both the application process and the results achieved by schools, thereby enhancing the educational administration's capacity for control.

Use of Results in Accountability Policies

Unlike the AAP, the *Prova Paulista* has come to be used in processes of holding students, teachers, and schools accountable for results, either through sanctions or rewards. This change represents a shift in the pedagogical focus of assessment to a managerial logic, in which the results cease to primarily serve to monitor learning and improve educational practices, and instead function as a basis for mechanisms of control and inducement of behaviors within the school system.

FINAL CONSIDERATIONS

Originally designed as an instrument to support teachers' work by providing complementary information about student learning, the bimonthly external assessment of the São Paulo state school system has undergone several transformations over time. It has become increasingly integrated into daily school life, potentially influencing the pace, methods, and content of teaching. There is an expectation that teachers will quickly use the assessment results to enhance their practices. The digital application now enables real-time visualization of results for teachers, schools, regional education boards, and the Department of Education. This development has expanded the possibilities for remote control of classroom activities and has facilitated the adoption of accountability policies.

In summary, the current configuration of the *Prova Paulista* reflects the consolidation of an educational model designed to operate "teacher-proof," supported by standardized curricula and instructional materials, frequent assessments, and control mechanisms based on incentives and sanctions. Although it provides data on student learning at the end of each bimester, the design and implementation of the *Prova Paulista* indicate that it has played a role in monitoring the work of the school system, revealing a shift from its initial purpose to a predominantly managerial function.

However, the continued persistence of low learning indicators shows that the expected objectives are not being met. Instead, this logic tends to contribute to diminishing the school's role as a space for the holistic development of students, ultimately compromising the quality and the very meaning of the educational process.

REFERENCES

AFONSO, Almerindo J. *Avaliação educacional: regulação e emancipação: para uma sociologia das políticas avaliativas contemporâneas*. Cortez, 2000.

ALAVARSE, Ocimar M. Desafios da avaliação educacional: ensino e aprendizagem como objetos de avaliação para a igualdade de resultados. *Cadernos Cenpec*, v. 3, n. 1, p. 135-153, 2013. Disponível em: <https://www.researchgate.net/publication/304450348_Desafios_da_avaliacao_educacional_ensino_e_aprendizagem_como_objetos_de_avaliacao_para_a_igualdade_de_resultados>. Acesso em: 15/02/2024.

AMARAL, Michelly F. B.; ALAVARSE, Ocimar M.; SILVA, Fernando A. O desafio da construção de um modelo alternativo de avaliação externa em larga escala da aprendizagem: um desafio, simultaneamente, conceitual, técnico e político como ponto de apoio para o sucesso de todos os alunos. *Olhar de Professor*, v. 23, p. 1-18, 2020. <<https://doi.org/10.5212/OlharProfr.v.23.2020.15616.209209225971.0530>>

ANDERSON, Gary. Privatizando subjetividades: como a Nova Gestão Pública (NGP) está criando o “novo” profissional da educação. *Revista Brasileira de Política e Administração da Educação*, v. 33, n. 3, p. 593-626, 2017. <<https://doi.org/10.21573/vol33n32017.79297>>

ARREDONDO, Santiago Castillo; DIAGO, Jesús Cabrerizo. *Avaliação educacional e promoção escolar*. Curitiba: InterSaber, 2013.

BAUER, Adriana *et al.* Avaliação em larga escala em municípios brasileiros: o que dizem os números? *Estudos em Avaliação Educacional*, v. 26, n. 62, p. 326-352, 2015. <<https://doi.org/10.18222/ae266203207>>

BERNARDES, Adilson T. *A Precarização do Trabalho Docente na Rede Pública Estadual Paulista: possibilidades de uma análise territorial*. Dissertação (Mestrado em Geografia Humana). São Paulo: Universidade de São Paulo, 2018. Disponível em: <<https://www.teses.usp.br/teses/disponiveis/8/8136/tde-23072019-154906/pt-br.php>>. Acesso em 18/02/2024.

BONAMINO, Alicia; SOUSA, Sandra Z. Três gerações de avaliação da educação básica no Brasil: interfaces com o currículo da/na escola. *Educação e Pesquisa*, v. 38, p. 373-388, 2012. <<https://doi.org/10.1590/S1517-97022012005000006>>

BRESSER-PEREIRA, Luiz C. A Reforma Gerencial do Estado de 1995. *Revista de Administração Pública*. v. 34, n. 4, p. 55-72, 2000. Disponível em: <<https://periodicos.fgv.br/rap/article/view/6289/4880>>. Acesso em: 08/03/2024.

BROOKE, Nigel; CUNHA, Maria Amália de A. A avaliação externa como instrumento da gestão educacional nos estados. *Estudos & Pesquisas Educacionais*, São Paulo, v. 2, n. 5, p. 17-79, 2011. Disponível em: <https://fvc.org.br/wp-content/uploads/2018/04/estudos_e_pesquisas_educacionais_vol_2.pdf>. Acesso em: 04/04/2024.

CELLARD, André. A análise documental. In: POUPART, Jean *et al.* *A pesquisa qualitativa: enfoques epistemológicos e metodológicos*. Petrópolis: Vozes, 2008, p. 295-316.

DUARTE, Alexandre W. B. Políticas Avaliativas, Nova Gestão Pública e Trabalho Docente no Brasil: Reformas nos Sistemas de Ensino e a Repercussão Sobre o Trabalho do Diretor Escolar no Estado de Minas Gerais. *Sisyphus – Journal of Education*, v. 8, n. 1, p. 28-54, 2020. <<https://doi.org/10.25749/sis.19005>>

FERNANDES, Reynaldo; GREMAUD, Amaury P. Qualidade da educação básica: avaliação, indicadores e metas. In: VELOSO, Fernando *et al.* (Orgs.). *Educação básica no Brasil: construindo o país do futuro*. Rio de Janeiro: Elsevier, 2009, p. 213-238.

FREITAS, Luiz C. Os reformadores empresariais da educação e a disputa pelo controle do processo pedagógico na escola. *Educação & Sociedade*, v. 35, p. 1085-1114, 2014. <<https://doi.org/10.1590/ES0101-73302014143817>>

FREITAS, Luiz C. *et al.* *Avaliação educacional: caminhando pela contramão*. 2ª ed. Petrópolis: Vozes, 2009.

GATTI, Bernadete A. Avaliação de sistemas educacionais no Brasil. *Revista de Ciências da Educação*, n. 9, p. 7-18, 2009. Disponível em: <https://www.researchgate.net/publication/28320450_Avaliacao_de_sistemas_educacionais_no_Brasil>. Acesso em: 06/04/2024.

GIMENES, Nelson *et al.* Além da Prova Brasil: investimento em sistemas próprios de avaliação externa. *Estudos em Avaliação Educacional*, São Paulo, v. 24, n. 55, p. 12-32, 2013. <<https://doi.org/10.18222/ae245520132718>>

HADJI, Charles. *Avaliação desmistificada*. Porto Alegre: Artmed Editora. 2001.

HAYDT, Regina C. *Avaliação do processo ensino-aprendizagem*. São Paulo: Ática, 1988

JACOMINI, Márcia A. *et al.* Padronização, controle e accountability na política curricular paulista (2007-2018). *Revista Brasileira de Educação*, v. 26, e260074, p. 1-24, 2021. <<https://doi.org/10.1590/S1413-24782021260074>>

JACOMINI, Márcia A. *et al.* Política educacional na rede estadual paulista e qualidade do ensino sob a Nova Gestão Pública, 1998 a 2018. *Education Policy Analysis Archives*, v. 30, n. 27, p. 1-34, 2022. <<https://doi.org/10.14507/epaa.30.6465>>

JACOMINI, Márcia A.; NASCIMENTO, Iracema S. do; STOCO, Sergio. Política educacional na rede estadual paulista sob a nova gestão pública (1995-2018). *Educação em Revista*, v. 39, 2023. <<https://doi.org/10.1590/0102-469826145>>

KRIPKA, Rosana M. L.; SCHELLER, Morgana; BONOTTO, Danusa de L. Pesquisa documental na pesquisa qualitativa: conceitos e caracterização. *Revista de Investigaciones de la UNAD*, v. 14, n. 2, p. 55-73, 2015. <<https://doi.org/10.22490/25391887.1455>>

LIMA, Licínio C. Avaliação, competitividade e hiperburocracia. In: ALVES, Maria P.; KETELE, Jean-Marie de. *Do currículo à avaliação, da avaliação ao currículo*. Porto: Porto Editora, 2011, p.71-82.

LIMA, Licínio C. Elementos de hiperburocratização da administração educacional. In: LUCENA, Carlos A.; SILVA JÚNIOR, João dos R. *Trabalho e educação no século XXI: Experiências internacionais*. São Paulo: Xamã, 2012, p.129-158.

LUKAS MUJICA, Jose F.; SANTIAGO ETXEBERRÍA, Karlos. Concepto, componentes y fases de la evaluación. In: LUKAS MUJICA, Jose F.; SANTIAGO ETXEBERRÍA, Karlos. *Evaluación educativa*. 2. ed. Madrid: Alianza, 2009, p. 87-126.

MACHADO, Cristiane; ALAVARSE, Ocimar M.; ARCAS, Paulo H. Sistemas estaduais de avaliação: interfaces com qualidade e gestão da educação. *Revista Brasileira de Política e Administração da Educação – Periódico científico editado pela ANPAE*, v. 31, n. 3, p. 667-680, 2015. <<https://doi.org/10.21573/vol31n32015.63800>>

MADAUS, George F. The influence of testing on the curriculum. In: TANNER, Laurel N. (Ed.). *Critical issues in curriculum*. Eighty-seventh Yearbook of the National Society for the Study of Education. Part I. Chicago, IL: University of Chicago Press, 1988, p. 83-121.

OLIVEIRA, Dalila A. Nova gestão pública e governos democrático-populares: contradições entre a busca da eficiência e a ampliação do direito à educação. *Educação & Sociedade*, v. 36, p. 625-646, 2015. <<https://doi.org/10.1590/ES0101-73302015152440>>

PEREIRA, Kalina E. L. *Avaliações Externas: implicações e contribuições para a prática avaliativa do professor*. Dissertação (Mestrado em Educação). São Paulo: Pontifícia Universidade Católica de São Paulo, 2023. Disponível em: <<https://repositorio.pucsp.br/jspui/handle/handle/32576>>. Acesso em: 15/03/2024.

RAMOS, Rubia de A.; GOULART, Débora C.; JACOMINI, Márcia A. Experiências, percepções e expectativas de estudantes na escola gerencialista da rede estadual paulista. *Pro-Posições*, v. 34, e20210129, p. 1-29, 2023. <<https://doi.org/10.1590/1980-6248-2021-0129>>

RAVAGNANI PINTO, Marcio A. *A Avaliação da Aprendizagem em Processo (AAP): SEE-SP (2011-2016): da proclamação à execução: estudo de caso do programa em uma escola*. Tese (Doutorado em Educação Escolar). Araraquara: Universidade Estadual Paulista “Júlio de Mesquita Filho”, 2016. Disponível em: <<https://repositorio.unesp.br/entities/publication/3ddb86c2-fe35-4bd5-a7d2-929b6bdf8596>>. Acesso em: 12/03/2024.

RODRIGUES, Jean D. Z. *Gerencialismo e Responsabilização: Repercussões para o trabalho docente nas escolas estaduais de Ensino Médio de Campinas/SP*. Tese (Doutorado em Educação). Campinas: Universidade Estadual de Campinas, 2018. Disponível em: <<https://repositorio.unicamp.br/acervo/detalhe/1079929>>. Acesso em: 12/03/2024.

RUSSEL, Michael K.; AIRASIAN, Peter W. *Avaliação em sala de aula: conceitos e aplicações*. Trad. Marcelo de Abreu Almeida. 7ª ed. Porto Alegre: AMGH, 2014

SÃO PAULO. *Comunicado SE, de 22 de março de 1995*. Diário Oficial do Estado de São Paulo: São Paulo, 23 de mar. 1995, Seção I, p. 8-10.

SÃO PAULO. *Resolução SE n.º 27, de 29 de março de 1996*. Dispõe sobre o Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo. São Paulo, 1996.

SÃO PAULO. *Proposta curricular do Estado de São Paulo*. São Paulo, 2008a. Disponível em: <http://www.rededosaber.sp.gov.br/portais/portals/18/arquivos/propostacurriculargeral_internet_md.pdf> Acesso em: 18/04/2024.

SÃO PAULO. *Resolução SE-74, de 6 de novembro de 2008*. São Paulo, 2008b.

SÃO PAULO. *Resolução SE-22, de 27 de março de 2009*. Dispõe sobre a definição dos indicadores específicos da Secretaria da Educação, para fins de pagamento da Bonificação por Resultados – BR, instituída pela Lei Complementar n.º 1078, de 17 de dezembro de 2008, seus critérios de apuração e avaliação. São Paulo, 2009.

SÃO PAULO. *Avaliação da Aprendizagem em Processo*. Videoconferência. Videoconferencistas João Cardoso Palma Filho; Regina Resek; Angélica Fontoura. Rede do Saber. São Paulo, 2011. Disponível em: <<https://www.youtube.com/watch?v=1MJGTjtVhbQ>>. Acesso em: 21/03/2024.

SÃO PAULO. *Avaliação da Aprendizagem em Processo. Terceira Edição – Segundo Semestre de 2012*. Comunicado Conjunto CGEB-CIMA, de 31 jul. 2012. Diário Oficial do Estado de São Paulo: São Paulo, 01 jun. 2012. Seção I, p. 38.

SÃO PAULO. *Resolução SE n.º 74, de 08 de novembro de 2013*. Dispõe sobre a reorganização do Ensino Fundamental em Regime de Progressão Continuada, oferecido pelas escolas públicas estaduais, e dá providências correlatas. São Paulo: SEE, 2013.

SÃO PAULO. *Avaliação da Aprendizagem em Processo. Oitava Edição – Primeiro Semestre de 2015*. Comunicado Conjunto CGEB-CIMA, de 23 jan. 2015. Diário Oficial do Estado de São Paulo: São Paulo, 24 de jan. 2015a. Seção I, p.27.

SÃO PAULO. *Avaliação da Aprendizagem em Processo. Nona Edição – Agosto de 2015*. Comunicado Conjunto CGEB-CIMA, de 25 jul. 2015. Diário Oficial do Estado de São Paulo: São Paulo, 25 de jul. 2015b., Seção I, p.28.

SÃO PAULO. *Sistema de Acompanhamento dos Resultados das Avaliações*. Coordenadoria de Informação, Monitoramento e Avaliação Educacional, São Paulo, [2017]. Disponível em: <https://midiasstoragesec.blob.core.windows.net/001/2017/02/tutorial_de_utilizacao_sara-professor_v4.pdf>. Acesso em: 21/03/2024.

SÃO PAULO. *Prova Paulista – 3.º bimestre*. Comunicado Externo Conjunto Subsecretaria/COPED, 2023, n.º 206, de 12 set. 2023a.

SÃO PAULO. *Educação de SP oferece material digital inédito para professores da rede; assista ao vídeo*, 24 abr. 2023b. Disponível em: <<https://www.educacao.sp.gov.br/sala-futuro-educacao-de-sp-oferece-material-digital-inedito-para-professores-da-rede/>>. Acesso em: 27/03/2024.

SÃO PAULO. *Resolução SEDUC n.º 4, de 19 de janeiro de 2024*. Dispõe sobre a Avaliação de Desempenho de Diretores Escolares/Diretores de Escola e dá providências correlatas. São Paulo, 2024a.

SÃO PAULO. *Avaliação bimestral: Prova Paulista começa na próxima segunda (8) nas escolas estaduais*. 02 abr. 2024b. Disponível em: <<https://www.educacao.sp.gov.br/avaliacao-bimestral-prova-paulista-comeca-na-proxima-segunda-8-nas-escolas-estaduais/>>. Acesso em 01/09/2024.

SÃO PAULO. *Prova Paulista: Exame será aplicado até dia 24 nas escolas de SP*. 19 jun. 2024c. Disponível em: <<https://www.saopaulo.sp.gov.br/spnoticias/ultimas-noticias/prova-paulista-exame-sera-aplicado-ate-dia-24-nas-escolas-de-sp/>>. Acesso em 01/09/2024.

SÃO PAULO. *Resolução SEDUC n.º 104, de 26 de novembro de 2024*. Dispõe sobre o registro do rendimento escolar dos estudantes das escolas da Rede Estadual e dá providências correlatas. São Paulo, 2024d.

SEGATTO, Catarina; ABRUCIO, Fernando. A gestão por resultados na educação em quatro estados brasileiros. *Revista do Serviço Público*, v. 68, n. 1, p. 85-106, 2017. <<https://doi.org/10.21874/rsp.v68i1.762>>

SILVA, Antonia B. da; SOUSA, Sandra. Z. Motivações para a criação de avaliação educacional em larga escala por municípios brasileiros. *Revista Educação em Questão*, v. 60, n. 64, p. 1-24, 2022. <<https://doi.org/10.21680/1981-1802.2022v60n64ID29326>>

SOUSA, Sandra. Záquia. Concepções de qualidade da educação básica forjadas por meio de avaliações em larga escala. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, v. 19, p. 407-420, 2014. <<https://doi.org/10.1590/S1414-40772014000200008>>

VERGER, Antoni; PARCERISA, Luís; FONTDEVILA, Clara. Crescimento e disseminação de avaliações em larga escala e de responsabilizações baseadas em testes: uma sociologia política das reformas educacionais globais. *Revista da FAEEDBA: Educação e Contemporaneidade*, Salvador, v. 27, n. 53, p. 60-82, 2018. <<https://doi.org/10.21879/faceba2358-0194.2018.v27.n53.p60-82>>

Submitted: 07/11/2025

Preprint: 07/06/2025

Approved: 09/24/2025

DECLARATION ON DATA AVAILABILITY

The research data is contained within the manuscript itself.

AUTHOR'S CONTRIBUTIONS

Author 1 – Conception of the study, data collection, data analysis, and writing of the text.

Author 2 – Conception and coordination of the study, supervision, and revision of the final writing.

DECLARATION OF CONFLICT OF INTEREST

The authors declare that there is no conflict of interest with this article.