Articles Section
2025, Volume 7, e2025016
Pages 1–10
10.33532/revemop.e2025016

revemop
ISSN 2596-0245

Development of mathematics assessments: dialogue between the SipaeDF matrix and the SME-SP diagnostic tool

Desenvolvimento de avaliações em matemática: diálogo entre a matriz do SipaeDF e a sondagem da SME-SP

Desarrollo de evaluaciones en matemáticas: diálogo entre la matriz Sipae-DF y el instrumento diagnóstico de la SME-SP

Janaina Segatto[1]
Edvonete Souza de Alencar[2]

Abstract
This article aims to identify key elements in the performance reference matrix (SipaeDF) and the mathematics diagnostic tool (SME-SP) that may contribute to the development of assessment instruments to analyze the mathematical strategies employed by 2nd-grade elementary students. This is a qualitative research study based on document analysis of two instruments: the Federal District's assessment matrix and the diagnostic tool used by São Paulo's municipal education system. While the SipaeDF matrix uses multiple-choice questions to measure performance, the SME-SP tool allows for the documentation of students' problem-solving strategies. Drawing on Buriasco's theoretical contributions, we argue that it is possible to integrate both summative and formative purposes in assessment by focusing on the learning process.
Keywords: SipaeDF. Mathematics Survey. Development. Assessment Tool. Early Years.

Resumo
Este artigo busca identificar elementos presentes na matriz de referência de desempenho do SipaeDF e na sondagem de matemática da SME-SP que possam contribuir para a construção de instrumentos avaliativos voltados à análise das estratégias matemáticas de alunos do 2º ano do ensino fundamental. Trata-se de uma pesquisa qualitativa, baseada na análise documental de dois instrumentos: a matriz do Distrito Federal e a sondagem da rede municipal de São Paulo. A matriz do SipaeDF utiliza itens de múltipla escolha para mensurar desempenho, enquanto a sondagem permite o registro das estratégias de resolução. Com base em Buriasco, conclui-se que é possível articular a avaliação de rendimento e a diagnóstica ao considerar o processo de aprendizagem.
Palavras-chave: SipaeDF. Sondagem Matemática. Desenvolvimento. Instrumento Avaliativo. Anos Iniciais.

Resumen
Este artículo tiene como objetivo identificar elementos clave presentes en la matriz de referencia de desempeño (SipaeDF) y en el instrumento de sondeo de matemáticas de la SME-SP que puedan contribuir al desarrollo de herramientas evaluativas centradas en el análisis de las estrategias matemáticas empleadas por estudiantes de segundo grado de la educación primaria. Se trata de una investigación cualitativa basada en el análisis documental de dos instrumentos: la matriz de evaluación del Distrito Federal y el documento orientador de la red municipal de São Paulo. Mientras que la matriz SipaeDF utiliza ítems de opción múltiple para medir el rendimiento, la herramienta SME-SP permite registrar las estrategias de resolución de problemas. A partir del enfoque teórico de Buriasco, se argumenta que es posible articular la evaluación de rendimiento y la evaluación diagnóstica al considerar el proceso de aprendizaje.
Palabras clave: Evaluación en Matemáticas. Estrategias de Aprendizaje. Instrumento Diagnóstico. Educación Primaria. Evaluación Educativa.

[1] Mestranda em Educação pela Universidade de Brasília (UnB), Brasília, Distrito Federal, Brasil. E-mail: nina.segatto.2@gmail.com
[2] Doutora em Educação Matemática pela PUC-SP. Docente da Universidade de Brasília, Brasília, Distrito Federal, Brasil. E-mail: edvonete.alencar@unb.br

1. Introduction

This article analyzes two assessment instruments of distinct natures — the SipaeDF reference matrix and the SME-SP mathematics survey — to identify opportunities to link performance and diagnostic assessment. The analysis aims to support teachers' pedagogical work and educational managers' decision-making.

The discussion begins with the recognition of the role large-scale assessments play in public education systems, especially in verifying student performance and producing educational indicators. However, by predominantly using multiple-choice tests, these assessments do not allow for observing students' cognitive strategies for solving problem situations, especially in the mathematics component.

In the Federal District, the State Secretariat of Education (SEEDF) established, through Ordinance No. 38, of February 18, 2020, the performance reference matrix of the permanent educational assessment system of the Federal District (Sistema Permanente de Avaliação Educacional do Distrito Federal — SipaeDF). The purpose of this document is to guide large-scale district assessments in Portuguese language and mathematics, at both the elementary and secondary school levels. According to SEEDF (2020), the reference matrix is an excerpt from the Federal District's *Curriculum in Motion* [*Currículo em Movimento*], which aims to evaluate student performance in essential content related to the right to learning through standardized tests composed of multiple-choice items.

Given this context, the following question arises: What elements are highlighted by the SipaeDF performance reference matrix and the SME-SP mathematics survey for the development of assessment instruments that record the strategies students use in mathematics?

The objective of this article is to identify elements from the SipaeDF performance reference matrix and the SME-SP mathematics survey that contribute to the development of assessment instruments capable of verifying the mathematical strategies used by 2nd-grade elementary school students.

To achieve this goal, we conducted a comparative analysis. The comparative analysis adopted in this study follows Lima Júnior *et al.*'s (2021) methodological guidelines, which recommend this strategy to highlight similarities and contrasts between educational documents with different purposes. This article is structured into an introduction, a theoretical framework based on Buriasco (2000), a description of the methodological approach, based on the document analysis proposed by Lima Júnior *et al.* (2021), document analysis, final considerations, and references.

2. The evaluation system: theoretical framework

This article draws on Buriasco (2000) to discuss the social function of educational assessment, particularly in large-scale assessments. According to the author, this type of assessment measures student performance at a specific point in time using predefined instruments, typically objective tests. In this model, "achievement tests consist of placing the student at a certain point in the process" (Buriasco, 2000, p. 158), without it being possible to understand the procedures adopted by the student to obtain the result.

In this sense, Buriasco (2000) draws attention to the limitations of large-scale assessments, which do not allow for reflection on the results nor permit interventions in the teaching-learning process. The author argues that assessment should go beyond evaluating and measuring results, i.e., it should be a process that involves "defining principles based on objectives to be achieved; establishing instruments for action and choosing paths for that action, constantly verifying the progress" (Buriasco, 2000, p. 159). Curi, Santos, and Rabelo (2013) also note that results from that kind of evaluation have little impact on everyday teaching practice.

In addition to the contributions of the authors cited, this article engages with Fernandes (2015) in understanding formative assessment as a continuous process of collecting and interpreting evidence of learning, capable of providing feedback to pedagogical practice. It also considers Perrenoud's (1999) perspective, which highlights the importance of assessment as a regulator of learning rather than just a measure of results.

According to these authors, assessment should guarantee that the individuals involved have access to the objectives, criteria, and procedures used in the evaluation process. Buriasco (2000) considers this transparency as one of the fundamental principles of the evaluation process and recognizes it as an ethical axis.

According to the author, all performance evaluations can also serve as learning assessments, provided they are conceived as part of a continuous, reflective process. "If actions are planned before, during, and after the assessments, it will be possible to obtain relevant data on the development of student learning" (Buriasco, 2000, p. 160).

From this perspective, large-scale assessments can serve as instruments for diagnosing learning, provided they are conceived with a procedural logic that accounts for the pedagogical actions that precede and follow their application.

In the next section, we will present the methodology and its organization for this research.

3. Methodological paths

This article adopts a qualitative approach, focusing on document analysis, as proposed by Lima Júnior, Oliveira, Santos, and Schnekenberg (2021). According to the authors, qualitative research seeks to understand social phenomena by analyzing values, perceptions, meanings, and practices. Document analysis, in turn, consists of reading and interpreting documents that have not yet been subjected to third-party analytical treatment, such as laws, curriculum guidelines, opinions, institutional reports, and guiding documents. The analysis followed two methodological steps, as indicated by Lima Júnior *et al.* (2021): a preliminary reading, focused on contextualizing the documents and their objectives, and an in-depth reading, focusing on identifying elements that can support the construction of assessment instruments aimed at verifying the strategies used by students in solving mathematical problems.

Two documents were selected as *corpus* of analysis: (i) the SipaeDF performance reference matrix, as it is the document that guides large-scale assessments of the public schools of the Fede-

ral District[3], and (ii) the *Guiding Document for Mathematics Survey*[4] (Documento Orientador para a Sondagem de Matemática - SME/SP, 2018), for presenting a proposal for an already consolidated assessment instrument, which enables the recording of the strategies adopted by students during the resolution of problem situations.

The documents were selected for their institutional functions and for their potential for comparison. The focus on the 2nd grade of elementary school considers the phase of consolidating the number system and the beginning of the development of logical-mathematical reasoning, according to the BNCC (2017).

4. Dialoguing with the documents

Document reading followed two stages, as outlined in the methodological proposal of Lima Júnior *et al.* (2021): preliminary reading, aimed at understanding the general objectives of each text, and in-depth reading, focusing on identifying elements that can support the construction of assessment instruments in mathematics that allow the recording of strategies used by 2nd-grade elementary school students.

The SipaeDF performance reference matrix, developed by the State Secretariat of Education of the Federal District (SEEDF), guides the formulation of large-scale district assessments. The document defines that performance evaluation "is one indicator, among others, of educational quality and assesses the essentials that the student needs to know to have their right to learning guaranteed" (SEEDF, 2020, p. 10). The matrix organizes the objects of knowledge and skills into thematic units, following the *Curriculum in Motion* of the Federal District and the National Common Curriculum Base (BNCC).

This structure aims to construct objective tests composed of multiple-choice items. The purpose of the descriptors presented is to assess whether the student can mobilize specific skills in response to direct instructions, usually anchored in contextualized problem situations. However, as Buriasco (2000) points out, although it situates students at a specific point in their school trajectory, this evaluative model does not account for the cognitive strategies involved in the problem-solving process. Consequently, it indicates only whether the final answer is correct or incorrect, without highlighting the paths the student took to solve the task.

On the other hand, the *Guiding Document for Mathematics Survey*, prepared by the Municipal Department of Education of São Paulo (SME, 2018), proposes a different approach. The assessment is administered individually and aims to support the teacher's pedagogical work by focusing on monitoring student learning. According to the document, "the survey consists of carrying out challenging situations, involving reading, interpretation and problem solving, with a view to identifying the knowledge mobilized by the students" (SME, 2018, p. 6).

The instrument's structure allows recording of students' solutions, focusing on the analysis of the strategies used, regardless of whether the final answer is correct. The teacher records both

[3] Place of work for the author.

[4] Among the assessment tools for mathematics in the early years of elementary school, this one stands out for not adopting the structure of the matrices of the Basic Education Assessment System (Saeb).

the student's speech and the procedures they adopt while performing the task. This helps teachers identify hypotheses, regularities, and difficulties, which can guide pedagogical interventions that are more consistent with students' learning stage.

Although the SipaeDF performance reference matrix and the mathematics survey have distinct purposes — the first aimed at formulating public policies and the second at monitoring pedagogy — it is possible to identify elements that foster dialogue between the two documents. Both organize mathematical content into thematic units and describe expected skills for each school grade. However, only the survey includes mechanisms to record the strategies students use.

The convergence of these two instruments can contribute to the development of large-scale assessments that, without sacrificing the comparability of results, also consider aspects of the learning process. This possibility requires a review of the format of the instruments and application procedures, as well as an expansion of the purpose attributed to large-scale assessments, as proposed by Buriasco (2000), who argues that all performance assessment can also be a learning assessment, provided it is inserted in a continuous, dialogical and formative process (Buriasco, 2000, p. 160).

4.1. In-depth reading focused on teaching mathematics in the 2nd grade.

The in-depth reading phase focused on analyzing the skills and assessment proposals in the selected documents, specifically designed for 2nd grade in elementary school. This choice is justified by the importance of this stage in the mathematical literacy process, in which students build their first formal relationships with the number system, basic operations, and logical reasoning. According to the Brazilian National Common Curriculum Base (Brasil, 2017), during that period, students should develop the ability to solve addition and subtraction problems with understanding, as well as use personal strategies for calculation and mathematical argumentation.

In the matrix document (SEEDF, 2020, p. 47), the skills assigned to the 2nd grade include actions such as "solving problem situations involving the different ideas of addition and subtraction, involving numbers up to three digits" and "solving problem situations involving the ideas of multiplication with the support of images, involving numbers up to three digits." However, a detailed analysis shows that these skills are assessed through objective multiple-choice items that yield final performance data without access to the intermediate procedures students use. As already indicated by Buriasco (2000), this approach fails to capture students' reasoning strategies, thereby limiting the pedagogical value of the results obtained, especially in an early phase of mathematical development such as 2nd grade.

In contrast, the *Guiding Document for the Mathematics Survey* of the SME-SP (2018) proposes an approach that values the learning process and the reasoning developed by the student throughout the task resolution. Specifically in 2nd grade, the survey proposes open-ended problem situations that allow for multiple problem-solving strategies, such as progressive counting, use of concrete materials, number decomposition, or mental calculation. The detailed recording of students' speech, procedures, and hypotheses developed during the activities provides important qualitative data for understanding each student's learning stage.

This approach is consistent with what Luckesi (2011) advocates: that assessment should be a constitutive part of the educational process, rather than a control mechanism. For the author, evaluating is diagnosing, it is monitoring the movement of learning in its complexity, especially in the early years of schooling, when students are formulating their first cognitive strategies in the field of mathematics.

In-depth reading also revealed that the survey organizes learning expectations by numerical fields and expected strategies, allowing the teacher to identify not only what was learned but also how it was learned. For example, when proposing an activity involving the subtraction of two natural numbers, the tool guides the teacher in observing whether the student uses backward counting, decomposition, or the relationship between addition and subtraction. Each of these strategies represents a degree of refinement in mathematical thinking, as Curi, Santos, and Rabelo (2013) suggest in their studies on problem-solving procedures in the early years.

Comparative analysis shows that, for the 2nd grade, the survey is more detailed about learning processes than the matrix, as it considers the different paths students may take to arrive (or not) at the answer. Although the SipaeDF Matrix includes important skills aligned with the BNCC, its application through objective tests limits the identification of conceptual obstacles, learning gaps, and hypotheses under construction.

As pointed out by Buriasco (2000, p. 160), "Performance evaluation can also be a learning assessment, provided it is conceived as part of a continuous and reflective process." Thus, in-depth reading shows that, for the 2nd grade of elementary school, it is essential that assessment instruments allow access to students' reasoning, which requires more qualitative and open forms of recording, such as those proposed by the SME-SP survey.

Comparative table between the SipaeDF reference matrix (SEEDF, 2020) and the mathematics survey (SME-SP, 2018)

Chart 1: Reference Matrix and Mathematics Survey

Aspects	Reference Matrix – SEEDF (2020)	Mathematics Survey – SME-SP (2018)
Assessment focus	Verification of final performance through objective multiple-choice items.	Understanding the learning process through open-ended problem-solving situations.
Evaluation instrument	Standardized tests applied on a large scale (SIPAE-DF).	Diagnostic assessments, applied in educational contexts, with a formative focus.
Question type	Closed items, with alternatives.	Open-ended questions, which allow for multiple problem-solving strategies.
Access to student strategies	Restricted: Does not allow observation of students' procedures.	Broad: Observation and recording of procedures, hypotheses, and problem-solving strategies.
Focus on the BNCC (2018)	Skills aligned with the BNCC, such as problem-solving with addition, subtraction, and an introduction to multiplication.	Skills also aligned with the BNCC, with an emphasis on how the student learns and solves problems.

Student thought log	Not foreseen in the assessment instruments.	Encouraged and systematized through observations, note-taking, and analysis of students' speech and actions.
Diagnostic capacity	Limited: Focuses on right or wrong. It makes it difficult to identify obstacles or hypotheses under construction.	High: Allows one to identify gaps, advances, and cognitive strategies at different levels.
Assessment concept	Predominantly summative and classificatory.	Formative, diagnostic, and process-oriented, aligned with the concept of assessment as part of the teaching-learning process.
Theoretical framework	Buriasco (2000) criticizes the limitation of assessments that do not assess students' reasoning.	Based on authors such as Buriasco (2000), Luckesi (2011), and Curi et al. (2013), who advocate assessment as a diagnosis and continuous process.
Organization of skills	List of skills by school year, focusing on mathematical content.	Organization by numerical fields and expected strategies, facilitating the analysis of cognitive development.
Practical limitations	It produces comparable data, but without access to students' cognitive strategies.	It produces comparable data, but requires more time and teacher training for application and analysis of the results.

Source: Prepared by the author.

Both documents address fundamental content for the 2nd grade, such as basic operations and problem solving, aligned with the BNCC; however, while the SipaeDF matrix focuses on defining the skills to be developed and assessed objectively, the SME-SP mathematics survey emphasizes the learning process, valuing students' strategies to understand and solve problems, which reveals a significant difference in the assessment approach and in the understanding of mathematical content.

5. Final considerations

This article aimed to identify elements present in the performance reference matrix of SipaeDF (SEEDF, 2020) and the mathematics survey (SME-SP, 2018) documents that could contribute to the development of assessment instruments to verify the mathematical strategies used by 2nd-grade elementary school students. The qualitative and documentary analysis enabled a dialogue between the two documents and the formative assessment proposal presented in the mathematics survey of the Municipal Department of Education of São Paulo.

The SipaeDF reference matrix organizes knowledge and skills into objects based on the BNCC (Brazilian Core Curriculum Base) and the *Curriculum in Motion* of the Federal District. The instruments in this matrix use multiple-choice items to verify student performance. According to Buriasco (2000), this model limits students' access to the procedures and strategies they use, thereby reducing their contribution to the teaching process.

The mathematics survey by SME-SP proposes the use of open-ended problem situations to enable the observation and recording of students' reasoning processes. This approach allows us to monitor how students solve problems, helping to identify progress, difficulties, and hypotheses.

This approach considers assessment as part of the teaching and learning process, as advocated by authors such as Luckesi (2011), Curi *et al.* (2013), and Buriasco (2000).

A comparison of the two documents reveals differences in aspects such as the type of question, the pedagogical focus, the recording of students' thinking, and the organization of skills. While SipaeDF presents skills by school year based on content, the survey organizes them by numerical fields and strategies, which facilitates the analysis of students' procedures. Furthermore, the survey includes recording students' speech, actions, and decisions, allowing for more detailed monitoring of their learning.

The analysis reveals that it is possible to propose adjustments to large-scale assessments so that they also consider students' processes, rather than just the results. This integration between different approaches can broaden the pedagogical usefulness of assessments.

As a next step, we suggest developing blended assessment models that integrate quantitative and qualitative elements, reconciling the comparability of large-scale assessments with the diagnostic potential of pedagogical surveys. For public administrators, it is crucial to invest in continuing teacher education to ensure the effective adoption of assessment tools in daily school life. For future research, we recommended investigating the applicability of mixed assessment instruments across different networks and stages of basic education, focusing on the analysis of learning strategies and their correlation with performance outcomes.

6. References

BRASIL. Base Nacional Comum Curricular. Brasília: MEC, 2017. Disponível em: https://basenacionalcomum.mec.gov.br/. Acesso em: 17 set. 2025.

BURIASCO, Regina Célia Grando. Avaliação educacional: para além dos testes padronizados. In: CASTRO, Amélia Domingues de (Org.). Avaliação educacional: aspectos sociais e políticos. São Paulo: Autores Associados, 2000. p. 151–163.

CURI, Elisabete; SANTOS, Cláudia Andréa Batista dos; RABELO, Maria Helena Martins. Procedimentos de resolução de alunos de 5º ano revelados em itens do Saeb com relação ao sistema de numeração decimal. Revista Brasileira de Estudos Pedagógicos, Brasília, v. 94, n. 236, p. 211–231, 2013.

FERNANDES, Célia Maria Carolino Pires. Avaliação formativa: práticas e desafios. São Paulo: Cortez, 2015.LUCKESI, Cipriano Carlos. Avaliação da aprendizagem escolar: estudos e proposições. 22. ed. São Paulo: Cortez, 2011.

LIMA JUNIOR, Eduardo Brandão; OLIVEIRA, Guilherme Saramago de; SANTOS, Adriana Cristina Omena dos; SCHNEKENBERG, Guilherme Fernando. Análise documental como percurso metodológico na pesquisa qualitativa. Cadernos da Fucamp, v. 20, n. 44, p. 36–51, 2021.

PERRENOUD, Philippe. Avaliação: da excelência à regulação das aprendizagens – entre duas lógicas. Porto Alegre: Artmed, 1999.PIAGET, Jean. Equilibração das estruturas cognitivas: problema central do desenvolvimento. Rio de Janeiro: Zahar, 1975.

SÃO PAULO (Município). Secretaria Municipal de Educação. Sondagem de Matemática: documento orientador. São Paulo: SME, 2018. Disponível em: https://portal.sme.prefeitura.sp.gov.br/. Acesso em: 17 set. 2025.

SECRETARIA DE ESTADO DE EDUCAÇÃO DO DISTRITO FEDERAL. Currículo em Movimento do Distrito Federal – Ensino Fundamental. Brasília, DF, 2018. Disponível em: https://www.educacao.df.gov.br/wp-conteudo/uploads/2021/07/Curriculo_em_movimento_do_Distrito_Federal___Ensino_fundamental__2018_.pdf. Acesso em: 9 set. 2025.

SECRETARIA DE ESTADO DE EDUCAÇÃO DO DISTRITO FEDERAL. Matriz de desempenho (SipaeDF). Brasília, 2020. Disponível em: https://www.educacao.df.gov.br/wp-conteudo/uploads/2020/12/matriz_desempenho_sipaedf.pdf. Acesso em: 9 set. 2025.

Appendix – Editorial Details

Editorial History
Received: 30/03/2025.
Accepted: 24/08/2025.
Published: 26/11/2025.

How to cite – ABNT
SEGATTO, Janaina; ALENCAR, Edvonete Souza de. Development of mathematics assessments: dialogue between the SipaeDF matrix and the SME-SP diagnostic tool. **REVEMOP**, Ouro Preto/MG, Brazil, v. 7, e2025016, 2025. https://doi.org/10.33532/revemop.e2025016

How to cite – APA
Segatto, J., & Alencar, E. S. de. (2025). Development of mathematics assessments: dialogue between the SipaeDF matrix and the SME-SP diagnostic tool. *REVEMOP, 7*, e2025016. https://doi.org/10.33532/revemop.e2025016

Funding
Not applicable

Conflicts of Interest
The authors declare that there is no conflict of interest of a personal, commercial, academic, political, or financial nature regarding this article.

Contribuição dos Autores
Resumo/Abstract/Resumen: Janaina Segatto, Edvonete Souza de Alencar; **Introduction or First considerations:** Janaina Segatto, Edvonete Souza de Alencar; **Theoretical framework:** Janaina Segatto, Edvonete Souza de Alencar; **Methodology:** Janaina Segatto, Edvonete Souza de Alencar; **Data analysis:** Janaina Segatto, Edvonete Souza de Alencar; **Discussion of results:** Janaina Segatto, Edvonete Souza de Alencar; **Conclusion or Final considerations:** Janaina Segatto, Edvonete Souza de Alencar; **References:** Janaina Segatto, Edvonete Souza de Alencar; **Manuscript review:** Janaina Segatto, Edvonete Souza de Alencar; **Approval of the final published version:** Janaina Segatto, Edvonete Souza de Alencar.

CRediT – Contributor Role Taxonomy – https://credit.niso.org/.

Data Availability
Not applicable / These research data have not been published in the data repository; however, the authors are committed to sharing them if the reader is interested.

Review Process
Double-blind peer review

Reviewers
Two *ad hoc* reviewers evaluated this article and did not authorize the disclosure of their names.

Translation / Proofreading
This article was funded by the Minas Gerais State Research Support Foundation (Fundação de Amparo à Pesquisa do Estado de Minas Gerais–FAPEMIG), Project APQ-04960-23, Notice N. 008/2023–Support program for scientific and technological publications.