



**HAL**  
open science

# The algebraic learning of middle school students' evolution during a school year: a statistical large-scale study based on results in mathematics didactics

Brigitte Grugeon-Allys, Julia Pilet

## ► To cite this version:

Brigitte Grugeon-Allys, Julia Pilet. The algebraic learning of middle school students' evolution during a school year: a statistical large-scale study based on results in mathematics didactics. *Educational Studies in Mathematics*, 2024, 10.1007/s10649-024-10347-z . hal-04677377

**HAL Id: hal-04677377**

**<https://hal.science/hal-04677377v1>**

Submitted on 8 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



# The algebraic learning of middle school students' evolution during a school year: a statistical large-scale study based on results in mathematics didactics

Brigitte Grugeon-Allys<sup>1</sup> · Julia Pilet<sup>1</sup>

Accepted: 1 July 2024  
© The Author(s) 2024

## Abstract

This large-scale study characterizes the algebraic learning of a cohort of nearly 800 French students during the last year of middle school (9th grade, 14–15 years old) and examines their evolution in relation to the composition of the classes to which the students belong. Based on the Anthropological Theory of the Didactic (ATD), the study is founded both on a reference model of elementary algebra and on technological-theoretical levels characterizing the students' reasoning and their knowledge regarding institutional expectations. These are used to code the data and define the statistical variables on which multivariate descriptive analyses are carried out. Students' learning is analyzed through *Pépîte*, an algebra test given at the beginning and at the end of the year. A cluster analysis resulted in three distinct clusters of similar classes and of similar students both at the beginning and end of the grade 9 school year. They are interpreted didactically and could be useful for a teacher who has to manage the heterogeneity of learning in a class. Some classes and students move from one cluster to another between the beginning and the end of the year, showing a wide variety of ways in which students' algebra learning progresses or regresses over a school year.

**Keywords** Assessment · Students' learning · Algebra · Praxeology · Longitudinal study · Large-scale study

Algebra plays a crucial role in students' further studies, as it is involved in many science courses and therefore determines their access. Since the 1990s, mathematics education research on the teaching and learning of algebra (Bednarz et al., 1996; Chevallard, 1989; Kieran, 1992, 2007; Sfard, 1991; Vergnaud et al., 1987) has made a major contribution to understanding and characterizing the learning processes in algebra and the effectiveness of teaching practices in achieving them. However, students encounter many difficulties in

---

✉ Julia Pilet  
julia.pilet@u-pec.fr

Brigitte Grugeon-Allys  
brigitte.grugeon-allys@u-pec.fr

<sup>1</sup> Univ Paris Est Creteil, Université Paris Cité, CY Cergy Paris Université, Univ. Lille, Univ Rouen, LDAR, 94010 Creteil, France

international tests (e.g., PISA, TIMMS); in particular, in the TIMSS 2019, French Grade 8 students attained an overall score in mathematics of 483, below the international average of 511, and an even lower Algebra score of 468 (Le Cam & Salles, 2020).

Mathematical knowledge is highly dependent on the schools, classes and, more generally, the institutions in which it is lived, learned, and taught according to the didactic transposition process (Chevallard & Bosch, 2020a). Each student learns in several successive classes composed of different students, with teachers whose expectations and practices are often different, with potential impact on the student's learning. Teachers, who often say they have a "good" class or a "weak" class, may have to deal with significant heterogeneity in their students' knowledge and make teaching choices to help them progress toward what is expected. To illustrate this, consider the proof problem shown in Fig. 1 and the answers of three middle school students (Table 1) at the beginning of grade 9 (14–15 years old).

Students A and C both state that the assertion is true, but student A gives an incorrect proof by example, an arithmetic strategy corresponding to primary school practices, while student C uses an algebraic strategy that is expected at the end of middle school. Student B also uses an algebraic strategy, but incorrectly. Three levels of reasoning emerge from this example. Even after studying algebra for two years, students have already developed an algebraic activity that may differ from one student to another and be more or less adapted to their further schooling.

In addition, some researchers (Bressoux, 2012; Nye et al., 2004) are interested in the potential effect of the class environment on student learning. The class effect is related to the composition of the class, the average academic level of the students, their distribution between "good" and "not so good" and the resulting heterogeneity, and to the number of students in the class. Their studies show that the class effect needs to account for the classroom teacher's practices. To contribute to our understanding of the mechanisms underlying student learning, we aim to characterize the knowledge that students build in a mathematical domain (Grugeon-Allys et al., 2018) and the changes in this knowledge according to class membership over a school year.<sup>1</sup>

Therefore, we address the following questions: How can we account for the diversity of students' learning in algebra? How can we summarize the diversity of algebraic learning by student and by class? How can we account for the variations in students' learning over the course of a school year depending on the class effect, particularly the learning of other students in the class to which they belong?

To answer these questions, we look at student assessment. Assessing only success or failure in an item is insufficient. It is necessary to assess the students' knowledge and reasoning, to distinguish them according to whether they are arithmetic or algebraic, and to situate them according to curricular expectations.

To do so, we follow the research carried out in France since the 1990s (Artigue et al., 2001) on the assessment of students in algebra and the regulation of teaching. We use the *Pépité*<sup>2</sup> automated assessment (Chenevotot-Quentin et al., 2016; Grugeon, 1997; Grugeon-Allys et al., 2018, 2022), from which the task in Fig. 1 is taken, designed both by teachers and by researchers in mathematics education and in interactive learning environments. *Pépité* aims to study students' knowledge and reasoning in a holistic way, from an epistemological

<sup>1</sup> The relations with teaching practices are not included in this article, nor are those with socio-economic aspects.

<sup>2</sup> The *Pépité* test is available for 8th, 9th, and 10th grade levels (Chenevotot-Quentin et al., 2016; Grugeon-Allys et al., 2018).

**Item 6: Proof and calculation program**

A student says to a friend:

"Choose a number, add 4 to it, multiply the result by 3, subtract three times the original number and divide the result by 2".

This student says that the result is always the same.

Is this true for any number? Justify your answer.

**Fig.1** Example of a proof problem

**Table 1** Answers of three middle school students at the beginning of grade 9 to item 6

Student	Answer
A	For number 1: $1 + 4 = 5$ ; $5 \times 3 = 15$ ; $15 - 3 = 12$ ; $12 / 2 = 6$ . So, it is true
B	$x + 4 \times 3 - 3 \times x / 2 = x + 12 - 3x / 2 = -2x + 12 / 2 = -2x + 6$ . So, it is false
C	$((x+4) \times 3 - 3 \times x) / 2 = (3x + 12 - 3x) / 2 = 12 / 2 = 6$ . So, it is true

and institutional perspective. It associates the students' answers with knowledge and errors listed a priori. Previous studies using *Pépîte*, mainly qualitative, aimed at describing students' learning and helping teachers to manage heterogeneous classes. They highlighted recurring patterns in students' levels of reasoning in algebra. The quantitative study based on *Pépîte* in grade 9 presented in this article aims to characterize both the diversity of students' learning of algebra and its variation over the grade 9 school year, considering the class effect (in terms of learning).

The rest of the paper is organized as follows; we begin by presenting the theoretical background for characterizing students' learning of school algebra, followed by the methodology. The results are then provided in three sections, followed by the discussion of the findings.

## 1 Theoretical foundations

### 1.1 An anthropological approach

We use the anthropological theory of the didactic (ATD, Bosch et al., 2017; Chevallard & Bosch, 2020b), which is based on the hypothesis that mathematical objects do not exist per se but emerge from teachers' practices through the mathematical activities they develop in their classes, which may differ from one teacher to another. In ATD, all regularly performed human activity is modeled under a single model, called "praxeologies," in terms of types of tasks, techniques used to solve these tasks of a given type, "technological discourse" based on knowledge and reasoning developed to justify techniques, and "theories" that organize the local technological discourse into coherent structures. We describe the praxeologies learned by students in relation to the praxeologies to be taught at the end of middle school and at previous levels. Praxeologies are not isolated but structured in relation to each other: Praxeologies are aggregated into local praxeologies around a technology, then into regional praxeologies around a theory, and finally into global praxeologies around several theories. The task in Fig. 1 concerns a praxeology of proof, and the students' answers reveal three techniques involving three different technological discourses.

To describe learned praxeologies by students that are not always mathematically adequate, we link the techniques used by students, especially those that are erroneous or inappropriate, to the use of old knowledge or the incorrect use of new knowledge involving kinds of errors and reasoning already highlighted by research on school algebra.

## 1.2 Results in the didactics of algebra

A large body of research links students' difficulties to the discontinuities (Kieran, 2007; Vergnaud et al., 1987) that exist between arithmetic and algebraic thinking and to the specificity of algebraic semiotic practices, particularly with regard to the status of the letter, the status of equality, and the use of algebraic knowledge to solve problems. In the 1990s, teaching strategies were proposed to introduce algebraic thinking (Bednarz et al., 1996). Another approach, which emerged in the 2000s with the "Early Algebra" movement, aims to develop algebraic thinking earlier in the curriculum (Carraher et al., 2006; Kieran, 2018; Kieran et al., 2016; Radford, 2013) with a continuous trajectory from the beginning of primary school to the end of middle school.

We summarize findings on the teaching and learning of algebra according to Kieran's (2007) model of algebraic activity. This model is based on three types of activity. First, generative activity concerns the formation of algebraic objects (formulas, algebraic expressions, and equations) in order to solve problems (modelling, generalizing, proving). Solving these problems requires translation between different registers of representation (algebraic scripts, numerical scripts, graphical representations, geometric figures, natural language). Second, transformational activity involves transforming expressions and equations in a way that preserves their equivalence. This may involve substitution, development, factoring, simplifying algebraic expressions and solving equations. Transformational activity is based on the structure of objects (sum, product). Third, global/meta-level activity involves solving modeling, generalization, proof, and equation problems that involve analogy (Radford, 2013). It leads to algebraic justifications and proofs as opposed to arithmetic reasoning.

These three activities call on the different status of letters (variables, unknowns, as opposed to label letters), equality as a relation of equivalence as opposed to equality "to perform," the complementary procedural and structural characteristics (Sfard, 1991) of the objects of algebra, their denotation ("Bedeutung"), and their sense ("Sinn") (Drouhard, 1992; Frege, 1971), based on their structure.

Thus, algebra is both a tool and an object. It is a tool for solving different types of problems (Chevallard, 1985, 1989; Ruiz-Munzón et al., 2013) that involve generalization, proof, modeling, and equating. It is a structured set of objects—algebraic expressions, formulas, equations or inequations—with specific properties and semiotic representations associated with different registers and processing modes. The algebraic processing of these objects brings into play both their syntactic and semantic aspects, based on a fair balance between the technical and theoretical dimensions of processing.

This synthesis is the foundation of both the reference model of algebra presented next and the analysis criteria for distinguishing different levels of reasoning.

## 1.3 A reference model for elementary algebra

A reference model of a mathematical field (Bosch & Gascón, 2005; Ruiz-Munzón et al., 2013) is a possible way to describe the complexity of the knowledge to be taught, based

on praxeologies. Based on the previous synthesis, the epistemological reference model of elementary algebra that we adopted is structured into three regional praxeologies related to algebraic expressions, formulas, and equations, which in turn are structured into five local mathematical praxeologies (Grugeon-Allys et al., 2022). We present them taking into account the three objects, expressions, formulas, and equations:

1. “Modeling” praxeologies, noted M, aimed at solving problems of generalization (algebraic expressions), modeling (formulas) and equating (equations)
2. “Proving” praxeologies, noted P, aimed at proving properties
3. “Calculating Numerically” praxeologies, noted CN, aimed at calculating numerical expressions and recognizing equalities
4. “Calculating Algebraically,” noted CA, aimed at operating on algebraic expressions (substituting, recognizing, developing, factoring) and equations (testing, solving an equation)
5. “Representing” praxeologies, noted R, aimed at translating a relation between definite and indefinite elements from one register of semiotic representation to another, or at associating several representations of an expression or equation between different semiotic registers.

The reference model is used to analyze and link the praxeologies to be taught, taught, and learned. The five praxeologies structure the analysis. As presented in the introduction to the proof task, students construct praxeologies that are not always mathematically correct or to what is expected at the end of middle school. We will therefore try to describe the levels of reasoning on which they rely in the different praxeologies.

## 1.4 Assessment of students' algebraic learning

Assessing students' learning requires us to consider that students develop new praxeologies concerning algebraic expressions and equations in their training at middle school in interaction with the praxeologies taught by their teachers. We assess learning in a holistic way (Vergnaud, 2009). We characterize the students' learned praxeologies based both on their answers to tasks covering all the local praxeologies to be taught and on the way in which they solve them (techniques, knowledge and reasoning). For this purpose, we use the 24 tasks of the *Pépîte* test covering the praxeologies of the reference model in algebra (Table 2). One task can mobilize several local praxeologies.

The tasks are multiple-choice or open-ended. Figure 1 shows the sixth *Pépîte* task. Appendix and Grugeon-Allys et al. (2018) provide other examples. The analysis of *Pépîte*

**Table 2** Praxeologies involved in the 9th grade level *Pépîte* Test (24 items)

Local praxeologies	Number of items	Item number
Calculating numerically	3	1.1, 1.2, 1.3
Calculating algebraically	7	5.1, 5.2, 5.3, 7.1, 7.2, 7.3, 10.3
Modelling	3	3.1, 10.1, 10.2
Proving	7	2.1, 2.2, 2.3, 4.1, 4.2, 4.3, 6
Representing	4	3.2, 8.1, 8.2, 9

is fully automated. Students answer on the computer and an algorithm analyzes and codes their answers (closed and open) (Grugeon-Allys et al., 2018). In addition, *Pépité* not only assesses the validity of the answers produced by the students, task by task, but also assesses the validity of the answers produced by each student regarding what is expected in the curriculum. Indeed an a priori analysis of each task also enables an assessment of the knowledge and reasoning used by the student to justify his or her answer to each praxeology.

## 1.5 Technological-theoretical levels of students' algebraic learning

We assess the learned praxeologies according to four *technological-theoretical* levels (Grugeon-Allys, 2016) on each local reference praxeology, called  $\theta$ -levels. Based on the epistemological study of algebra presented above, we define these  $\theta$ -levels a priori by hierarchizing them as shown in Table 3. These  $\theta$ -levels enable us to distinguish the knowledge and reasoning used by a student for each task of a given type. These levels are then used to code the students' answers according to each local praxeology, as described in Table 4.

Given a school level, for each item, possible students' answers are listed and a priori coded in an analysis according to whether they are correct (V1 or V2 coding<sup>3</sup>) or not (V3 coding) and the  $\theta$ -level they involve in each praxeology (Grugeon-Allys, 2016; Grugeon-Allys et al., 2022). Tables 5 and 6 show an excerpt from this analysis for the item shown in Fig. 1. We have chosen to indicate the most frequent (non-exhaustive) techniques identified in the didactic analysis and in the students' recurrent answers. These techniques can be based on primary arithmetic or algebraic and depend on the correct or incorrect semiotic representations used. They are related to the corresponding  $\theta$ -level on the list.

The answer analysis of the three students presented in Table 1 is detailed in Table 7.

The definition of  $\theta$ -levels enables a macroscopic analysis on all tasks with the same coding to avoid sticking to a microscopic and task-by-task analysis. We can then identify the knowledge and reasoning that students predominantly use on all the tasks related to the same local praxeology. The student's learned praxeology in algebra is therefore described by his or her percentage of successful tasks and by five  $\theta$ -levels related to each of the five local praxeologies, described by a sextuplet.<sup>4</sup>

**Table 3** Description of the  $\theta$ -levels for algebra

Level	Description for the algebraic field
Adequate (A)	Appropriate algebraic justification with formulation of the expected knowledge
Weakly adequate (WA)	Appropriate algebraic justification without formulation of the expected knowledge
Under construction (UC)	Appropriate algebraic justification for direct application tasks, but incomplete or unsuitable justification for more complex tasks, which lets erroneous formal rules live or rules used outside their field of validity. The expected knowledge and the aggregation of the different praxeologies are under construction
Old (O)	Justification using primary school arithmetic, with errors linked to a failure to negotiate the epistemological break between arithmetic and algebra

<sup>3</sup> V2 means that the answer is correct but not expected at this grade level.

<sup>4</sup> A sextuplet could be noted (success (%),  $\theta$ -level on M,  $\theta$ -level on P,  $\theta$ -level on R,  $\theta$ -level on CA,  $\theta$ -level on CN).

**Table 4** Characterization of  $\theta$ -levels according to local praxeologies concerning algebra

Praxeology	$\theta$ -level	Description (code)
Modelling (M)	Adequate (A)	Use of letters to produce complex algebraical expressions and formulas consistent with relations between quantities (M_A)
	Weakly adequate (WA)	Use of letters to produce first degree algebraical expressions with parentheses, formulas consistent with relations between magnitudes (M_WA)
	Under construction (UC)	Use of letters to produce simple first degree algebraical expressions without parentheses that are often incorrect in more complex cases (M_UC)
	Old (O)	Little or no use of letters or in an abbreviated form. Use of arithmetic practices (M_O)
	Adequate (A)	Use of correct algebraic justification (especially the counterexample) (P_A)
Proving (P)	Weakly adequate (WA)	Use of correct algebraic justification (P_WA)
	Under construction (UC)	Use of erroneous algebraic justification or legal arguments (P_UC)
	Old (O)	Use of examples to prove (P_O)
Representing (R)	Adequate (A)	Correct translation with reformulation support for complex algebraic expressions (R_A)
	Weakly adequate (WA)	Correct translation with reformulation of simple first-degree expressions (R_WA)
	Under construction (UC)	Correct translation in simple cases (without reformulation), otherwise mistranslation (brackets forgotten) or step-by-step writing (R_UC)
Calculating algebraically (CA)	Old (O)	Incorrect translation with abbreviations (R_O)
	Adequate (A)	Correct substitution, correct use of algebraic properties, and rewriting of equivalent complex algebraic expressions (CA_A)
	Weakly adequate (WA)	Correct substitution, correct use of algebraic properties, and rewriting of equivalent simple algebraic expressions (CA_WA)
	Under-Construction (UC)	Considering operating priorities in simple calculations ( $a \times x + b$ ) but errors in algebraic calculations such as $a \times (b + c)$ or $a + b \times c$ . (CA_UC)
	Old (O)	No consideration of operating priorities and incorrect concatenative rewriting such as $a + b \rightarrow ab$ , $a^n \rightarrow na$ (CA_O)
Calculating numerically (CN)	Adequate (A)	Correct use of properties and parentheses for all numbers in expressions (CN_A)
	Weakly adequate (WA)	Correct use of properties and parentheses for all numbers in simple expressions (CN_WA)
	Under construction (UC)	Considering operating priorities in simple calculations but errors with decimals and rationales (CN_UC)
	Old (O)	No consideration of operating priorities (CN_O)



**Table 5** A priori analysis of primary arithmetic techniques

Example of answers when the student chooses 1 as the number	Analysis	Coding
$((1+4) \times 3 - 3 \times 1) / 2 = 6$	Arithmetic technique with proof by example and global expression that uses parenthesis	V3, M_O, P_O, CA_A, R_A
$1+4=5; 5 \times 3=15; 15-3=12; 12 / 2=6$	Arithmetic technique with proof by example and partial expressions	V3, M_O, P_O, CA_UC, R_A
$1+4 \times 3 - 3 \times 1 / 2 = 6$	Arithmetic technique with proof by example and an erroneous global expression that does not use parenthesis	V3, M_O, P_O, CA_O, R_O
$(1+4)3=15-3=12 \times 1=12 / 2=6$	Arithmetic technique with proof by example and erroneous step-by-step calculations	V3, M_O, P_O, CA_O, R_O

**Table 6** A priori analysis of algebraic techniques

Example of answers	Analysis	Coding
$\begin{aligned} & ((x+4) \times 3 - 3 \times x) / 2 \\ & = (3x + 12 - 3x) / 2 \\ & = 12 / 2 \\ & = 6 \end{aligned}$	Algebraic technique with generalization and correct global expression that uses parenthesis	V1, M_A, P_A, CA_A, R_A
$\begin{aligned} & (x+4) \times 3 = 3x + 12 \\ & 3x + 12 - 3 \times x = 12 \\ & 12 / 2 = 6 \end{aligned}$	Algebraic technique with generalization and correct step-by-step calculations	V2, M_WA, P_A, CA_A, R_A
$\begin{aligned} & x + 4 \times 3 - 3 \times x / 2 \\ & = x + 12 - 3x / 2 \\ & = -2x + 12 / 2 \\ & = -2x + 6 \end{aligned}$	Algebraic technique with generalization and erroneous global expression that does not use parenthesis	V3, M_UC, P_UC, CA_UC, R_UC
$\begin{aligned} & (x+4) \times 3 = 3x + 12 = 15x \\ & 15x - 3x = 12x \\ & 12x / 2 = 6x \end{aligned}$	Algebraic technique with generalization and erroneous step-by-step calculations	V3, M_UC, P_UC, CA_O, R_UC

More examples of *Pépite* items are given in Appendix according to the different local praxeologies at play in each item: *calculating numerically* for item 1, *proving* for item 2, and *modelling* for item 3

**Table 7** Analysis and coding of the answers of three students to item 6

Student	Analysis and coding
A	The answer is incorrect, coded V3, of an arithmetical nature without algebraic modeling, coded M_O, with a proof by example, coded P_O, with a correct representation of the numerical calculations step by step, coded R_A and CA_UC as algebraic calculations are expected at the end of secondary school (third row of Table 5)
B	The answer is incorrect, coded V3; the modelling is algebraic, but the representation of the calculation program is incorrect (no brackets indicating operating priorities), so it is coded M_UC and R_UC; the proof is algebraic but the calculations are incorrect (no development) so they are coded P_UC and CA_UC (fourth line of Table 6)
C	The answer is correct, so it is coded V1; the modelling and proof are algebraic, so are coded M_A and P_A; the calculation program is correctly translated (use of brackets), so is coded R_A; the development and reduction of the expression are correct, so are coded CA_A (first line of Table 6)

In projects using *Pépité*, an algorithm determined the dominant  $\theta$ -levels for each local praxeology (Grugeon-Allys et al., 2022), with thresholds set based on experience. In addition, we have defined three groups of students (Grugeon-Allys et al., 2018) to enable teachers to better organize their teaching according to the learning needs of their students. For each group, a type of reasoning is used predominantly, either an adequate algebraic reasoning, or one in the process of being constructed, or one taken from elementary school. In this study, this algorithm is not used as we perform statistical analyses on all the  $\theta$ -levels used by students in their answers.

## 1.6 Research questions

A teacher needs to have an overview of his or her class to know what his or her students' learning needs are. Even if the potential number of learned praxeologies (sexuplets) is large, there are certainly praxeologies learned by students who are close, that is, with many of the same  $\theta$ -levels on all tasks (for instance, a majority of "Adequate" per local praxeologies). The composition of one class may differ from another, depending on the praxeologies learned by the students who make up the class. Indeed, in any class, there may be a varying number of students with adequate, old, or under construction learned praxeologies, which does not offer the same learning conditions to the students. Therefore, we ask for classes (RQ1): Are there similar classes in terms of learning, that is, with most of the same  $\theta$ -levels of local praxeologies on all task answers of students composing a class? And for students (RQ2): Are there students with similar learned praxeologies? In addition, for similar classes and students, we want to know (RQ3) if there are any variations between the beginning and the end of the grade 9 year. And (RQ4) do these variations rely on class effect (in terms of learning)?

## 2 Methodology

This large-scale study involves a sample of 36 classes and 771 students and their teachers (36). We begin by presenting how the sample was constituted for this large-scale study. We then move on to the construction of the databases, showing how the reference model of algebra and the  $\theta$ -levels serve as a basis for structuring data and interpreting statistical analyses. Finally, we present the multivariate descriptive statistical methods used to obtain clusters of classes and students to establish learning similarities between classes (RQ1) and students (RQ2) at the beginning and at the end of the year (RQ3). The study of class and student belonging to clusters between the beginning and end of the year enables us to analyze variations over the school year in relation to class effect (RQ4).

### 2.1 Sampling

The sample was constructed with the help of the Limoges Académie French educational authority.<sup>5</sup> It is not representative of the population studied, but it was built according to specific criteria, presented in Table 8, to have a diversity of teaching contexts (rural or urban, public or private education, priority or not education) and teachers (experience and age) within the Académie.

The 771 grade 9 students were distributed in 36 classes, from the same academy. The students in the sample classes took the same *Pépîte* algebraic test twice during the school year, Test 1 at the beginning and Test 2 at the end of the 9th grade. Each test was administered during a 50-min class period. The administration of the tests in schools was managed by the DEPP and the Limoges Académie during the 2018–2019 school year, which was not a period of educational reform in France. We did not meet the teachers or the students and had no information on how the teachers planned their lessons or on what was taught.

**Table 8** Criteria used to select the sample

Variables	Choice
Education authority	The same: "Académie" of Limoges
Number of classes according to the population density of the three departments in the "Académie"	17 in Corrèze, 8 in Creuse and 16 Haute-Vienne
Diversity of the establishments to which the classes belong	4 classes in private establishments under contract, 3 classes in public establishments in priority education, 29 classes in public establishments outside priority education
Regional diversity	20 classes in rural areas, 16 classes in urban areas
Number of students per class	From 13 to 29 students
Age of teachers	From 26 to 62
Number of years' experience	From 1 to 30 years

<sup>5</sup> In France, an Académie is an administrative district of the Ministry of National Education and the Ministry of Higher Education and Research.

## 2.2 Organizing the data

We built a database of students' responses coded in relation to the  $\theta$ -levels for each local praxeology of algebra to study similar classes or similar students. Once the 771 students passed the tests, the collected data were anonymized, cleaned, and organized into databases in the form of individuals/variables tables. The data were cleaned to keep only the 25 classes in which at least 15 students passed Test 1 and the 454 students that took both tests. We constructed two classes databases C1 and C2 (unit of analysis is the class), one for each test to answer to RQ1 and RQ3. Respectively, we constructed two students databases S1 and S2 (unit of analysis is the individual student), one for each test to answer to RQ2 and RQ3.

The databases are structured by continuous quantitative statistical variables concerning the success rate and the  $\theta$ -levels of the local algebraic praxeologies. We built these databases on the initial coding of the students' answers in *Pépité* (Fig. 2), where each row captures the coding of a student's answer to an item and each column corresponds to the  $\theta$ -levels of the local praxeologies. A 1 is assigned when the answer corresponds to that  $\theta$ -level.

Class ID	Student ID	Item	Representing Adequate	Representing Weakly Adequate	Representing Under Construction	Representing Old	Calculating Algebraically Adequate	Calculating Algebraically Weakly Adequate	Calculating Algebraically Under Construction	Calculating Algebraically Old	Calculating Numerically Adequate Weakly Adequate	Calculating Numerically Under Construction	Calculating Numerically Old	Proving Adequate Weakly Adequate	Proving Under Construction	Proving Old	Modeling Adequate Weakly Adequate	Modeling Under Construction	Modeling Old	
1016269	22	1,1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
		1,2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
		1,3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
		2,1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
		2,2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
		2,3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
		3,1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		3,2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
		4,1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
		4,2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
		4,3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
		5,1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
		5,2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		5,3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		6	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1
		7,1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
		7,2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
		7,3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
		8,1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
		8,2	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
		9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		10,1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
		10,2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
		10,3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 2 Coding of the answers to the 24 items in *Pépité* from student 22 in class 1016269

Classes	Head-count	Success	Failure	Calculating Numerically			Calculating Algebraically			Representing				Modelling		Proving	
				WA	UC	O	WA	UC	O	A	WA	UC	O	WA	UC	O	WA
1016269	29	39	53	39	14	47	29	18	18	26	5	34	8	29	52	33	6
1682439	21	49	46	36	15	49	41	19	18	39	6	32	7	59	28	48	7
1789941	15	47	47	49	8	41	34	16	14	32	4	30	9	36	41	36	3
3318285	22	46	49	48	8	43	41	11	22	37	4	33	7	47	42	36	9
3714802	24	29	66	31	2	56	21	16	22	16	3	36	9	4	65	13	9
4050694	30	42	54	48	7	41	38	18	23	32	3	38	7	27	54	43	8
4309761	26	34	56	22	7	68	29	19	18	31	5	25	8	30	52	26	9
4395992	29	59	35	69	7	20	51	12	13	46	2	25	10	45	26	51	2

**Fig. 3** Extract from the classes database for Test 1, showing the proportion of success and failure and the various praxeologies

For the classes database, an extract of which is presented in Fig. 3, the values of the variables displayed in the columns are the results of several calculations, carried out on tasks answered by the students of each class:

- A Success rate: the percentage of tasks successfully answered by the students in the class.
- A Failure rate: the percentage of failed items. The sum of the Success and Failure rates does not necessarily equal 100%, because the calculation does not consider the tasks not processed and the answers not analyzed by the *Pépète* software.
- A rate on each  $\theta$ -level for each praxeology of algebra: percentage calculated from the sum of the “1” for each variable, for all the answers on the tasks processed by the students in the class.

Note that not all  $\theta$ -levels appear in Fig. 2. To obtain consistent statistical analyses, it was necessary to group certain  $\theta$ -levels because some students did not answer to the open-ended modeling and proof tasks. The same rates are calculated for each student on each test:

- A Success rate: the percentage of successful items completed by the student.
- A Failure rate: percentage of failed items.
- A rate on each  $\theta$ -level for each praxeology: percentage calculated from the sum of the 1 for each  $\theta$ -level, for all the student’s answers.

### 2.3 Multivariate descriptive analysis

Given our research questions and the large sample, we apply multivariate descriptive statistical analyses to these databases, in particular principal component analysis (PCA) and hierarchical agglomerative clustering (HAC), to determine clusters both of classes and students and the variables ( $\theta$ -levels for each local praxeology) that best identify them.

The objective of a PCA is to identify the structure of the data based on the most relevant dimensions using a factorial method to obtain an overview of the similarities between individuals, in this case the characterization of classes in terms of learning and then that of the

students. This method consists of constructing and selecting new variables, the principal components, obtained from the correlation matrix constructed on the data; these summarize the most important information in the database. The principal components enable us to determine one or more factorial plans that maximize the information in the cloud of points-individuals projected on these designs (Hahn & Macé, 2017). The quality of the representation of individuals and variables is obtained using the  $\text{Cos}^2$  method (Hahn & Macé, 2017). The representations used for PCA, such as correlation circles and individual scatterplots, are visual graphical summaries. Correlation circles indicate which of the linear combinations of variables are the most informative for each factorial axis. In this study, PCAs were performed on the classes and then the students databases for Tests 1 and 2. Studying the scatterplots of classes and students enables us to locate them in relation to each other and to situate classes or students according to the direction of the factorial axes we have interpreted.

HACs are then performed to further group classes or students into clusters that are similar in terms of learning. We chose this clustering method because the structure of the clusters is not known in advance, and our data have an intrinsic hierarchical structure (hierarchy in  $\theta$ -levels for each local praxeology). An HAC consists of performing a series of successive partitions of individuals, nested one within the other, with the groupings into clusters being made in relation to the greatest proximity of the individuals. HACs were performed on data reduced by PCAs. This choice was made to simplify the complexity of the data while preserving crucial information. Comparing the clusters obtained by the PCA and HAC of Test 1 with those of Test 2 for the classes and students enables us to study the variation of learning over a school year (RQ3) and the link with the composition of the classes in terms of learning (RQ4).

### 3 Results: similar classes in terms of learning and their variation over the school year

#### 3.1 A multivariate analysis on test 1 results per class

We identify similar classes and their characteristics on the C1 database. To facilitate PCA, certain statistical variables have been grouped together. For example, *Representing\_Adequate* and *Representing\_Weakly-Adequate* have been grouped together into *Representing\_Adequate\_Weakly-Adequate*.

The PCA identifies a first factorial plane (71% of the information) and the four variables correlated with Axis 1 (59% of the information) that best contribute to its formation, namely, *Failure*, *Success*, *Calculating-Algebraically\_Adequate*, and *Representing\_Adequate\_Weakly-Adequate* (Fig. 4). These variables have a very good representation quality of since their  $\text{cos}^2$  is greater than 0.8. *Calculating-Numerically\_Weakly-Adequate* and *Calculating-Numerically\_Old* variables are correlated with Axis 2 (12% of the information) but with a less good quality of representation.

In the first factorial plane (Fig. 5), the classes located farthest to the right in the first and fourth quadrants have a good *Success* rate, as well as *Calculating-Algebraically\_Weakly-Adequate* (correlated with axis 1) and *Calculating-Numerically\_Weakly-Adequate*, unlike the classes located farthest to the left. For several classes located in the middle, with  $\text{Cos}^2$  close to or less than 0.25, the quality of the representation is less good than those with a  $\text{Cos}^2$  greater than 0.5.

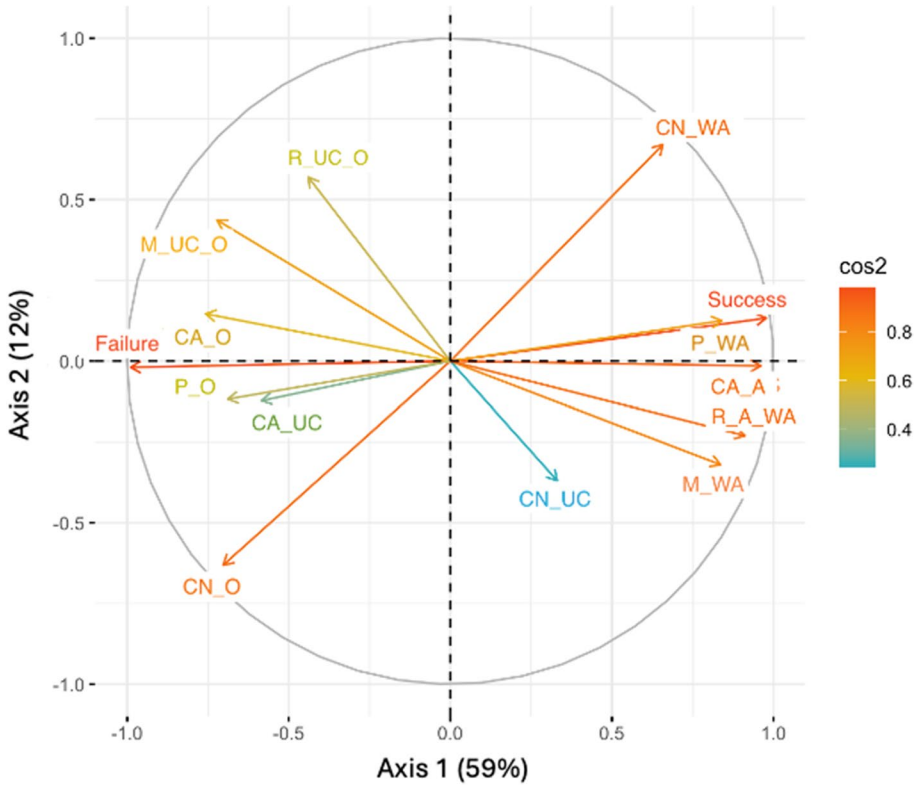


Fig. 4 Correlation circle in the first factorial plane for Test 1 ( $N=25$ )

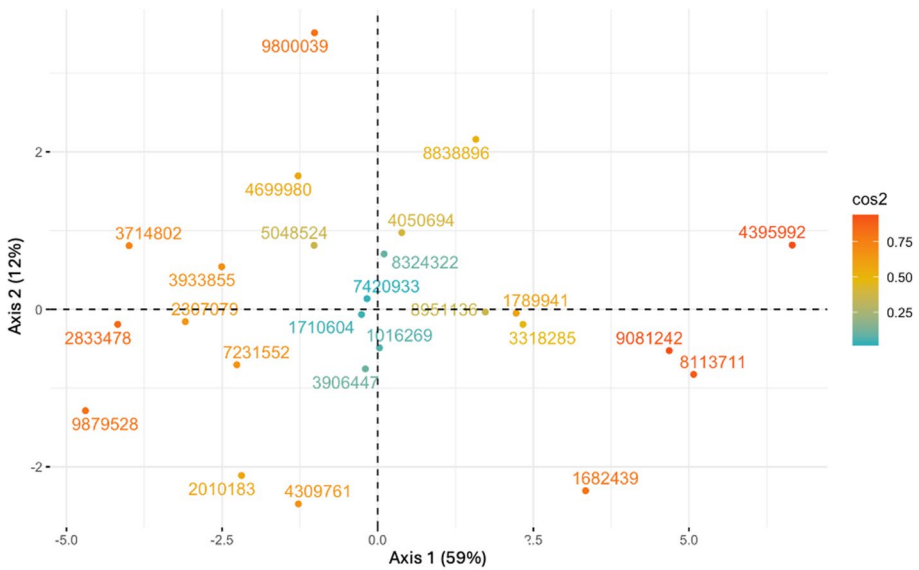


Fig. 5 Distribution of classes in Test 1 on the first factorial plane and quality of their representation ( $N=25$ )



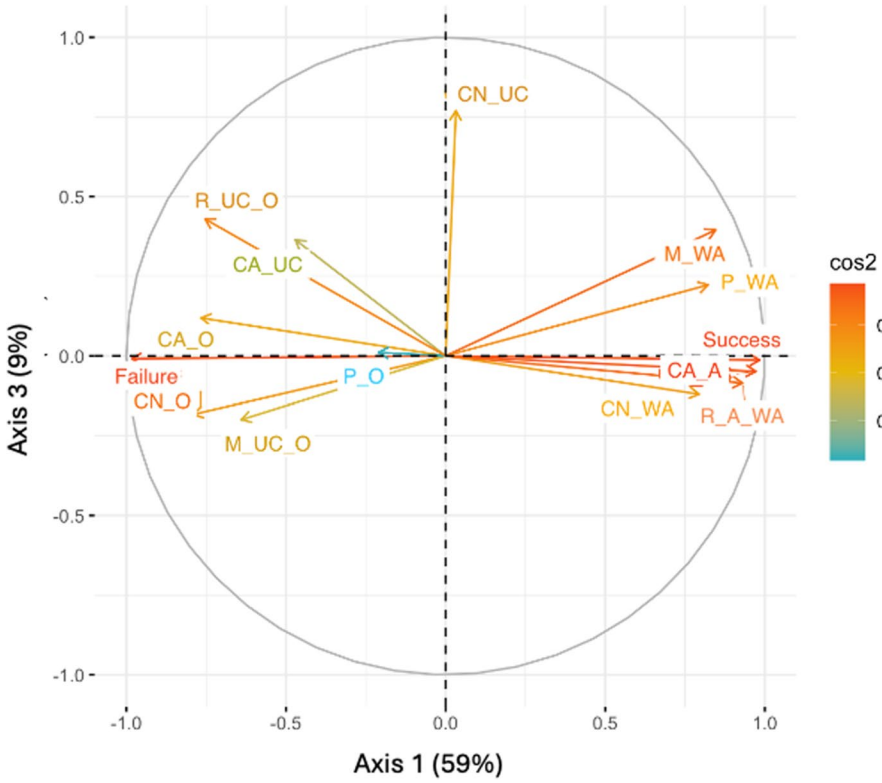


Fig. 6 Correlation circle of the second factorial plane for Test 1 (N=25)

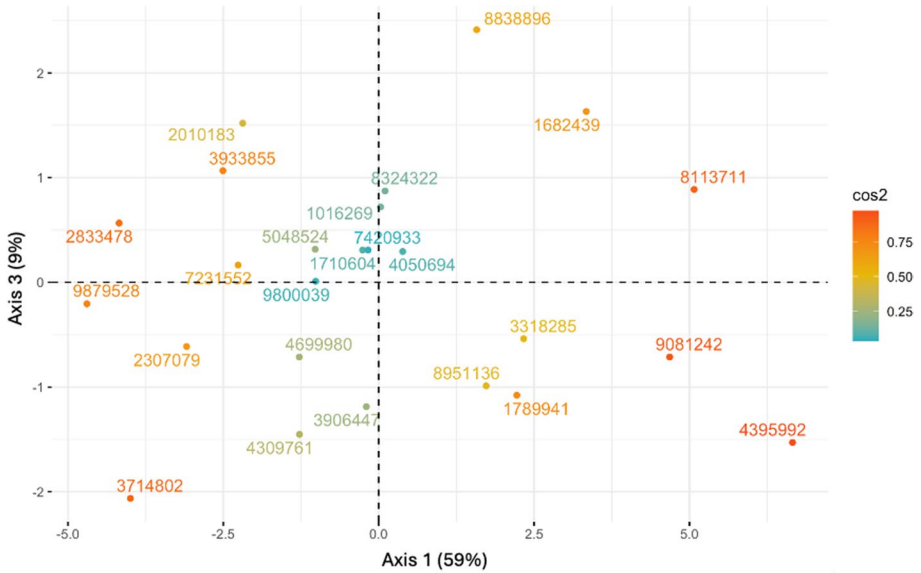


Fig. 7 Distribution of classes on the second factorial plane and quality of representation for test 1 (N=25)

The second factorial plane (68% of the information) shows a good quality of representation for the variable *Calculating-Numerically\_Under-Construction* on Axis 3 (9% of the information) (Fig. 6). Figure 7 shows the distribution of the classes on the second factorial plane.

We interpret these two factorial planes as relating the *Adequate* or *Old*  $\theta$ -levels of algebraic praxeologies and  $\theta$ -levels of *Calculating-Numerically* in order to study class similarity.

The HAC then identifies three clusters of similar classes (Fig. 8), called CI-A1 (4 classes), CI-B1 (14 classes), and CI-C1 (7 classes) and described in Fig. 9.

Cluster CI-A1 has a higher *Success* rate (53%) than *Failure* (40%), and the rates for the *Adequate* or *Weakly-Adequate*  $\theta$ -levels concerning each praxeology are close to

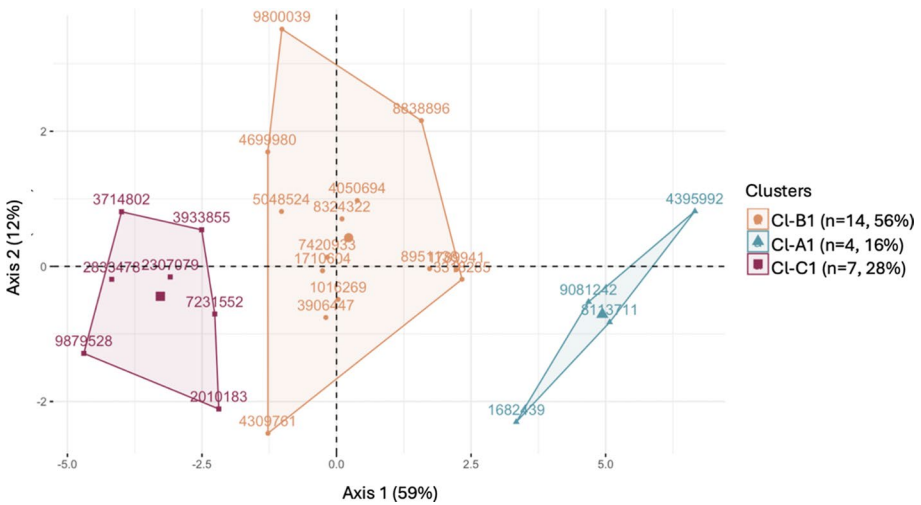


Fig. 8 Three clusters of similar classes on Test 1 (N=25)

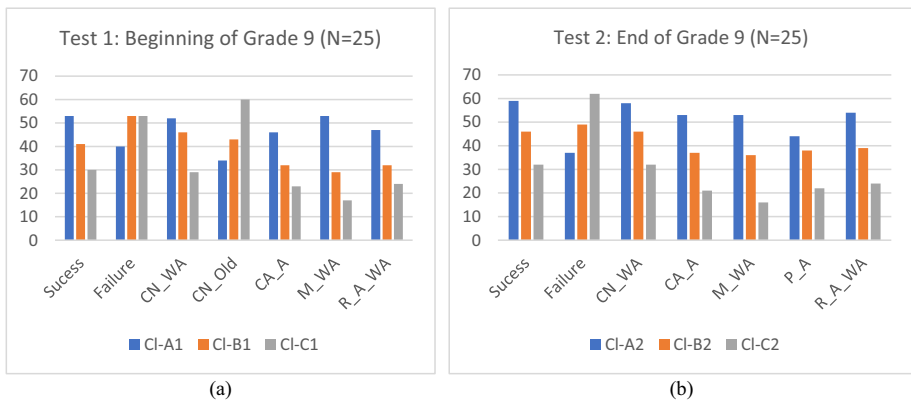


Fig. 9 Percentages for the variables that best contribute to the formation of the first factorial plane on Test 1 (a) and Test 2 (b)

50%. At the beginning of the school year, the CI-A1 classes contain many students who use the algebraic practices expected at this school level.

Cluster CI-B1 has a lower *Success* rate (41%) than the *Failure* rate (53%). These classes have rates for the *Adequate* or *Weakly-Adequate*  $\theta$ -levels concerning *Calculating-Algebraically*, *Modeling*, and *Representing* that are close to 30%. But the rate on *Calculating-Numerically\_Weakly-Adequate* is 46% indicating that these classes have a learning leverage for *Calculating-Numerically*. Having a rate of *Calculating-Numerically\_Weakly-Adequate* seems to be a lever at the beginning of the school year. This cluster brings together classes of rather heterogeneous composition.

Cluster CI-C1 has a very low *Success* rate (30%) and very low rates on the *Adequate* or *Weakly-Adequate* levels below 30%. Moreover, the rate on *Calculating-Numerically\_Old* is high (60%). At the beginning of the school year, the CI-C1 classes have a predominantly primary arithmetical practices to solving algebraic problems. The *Calculating-Numerically* praxeology is a not lever for classes that are far from what is expected by the school.

### 3.2 Variation during a school year

We performed the same analyses on the C2 database. The variables that best contribute to the formation of the first factorial plane are the same as in Test 1, except for *Calculating-Numerically\_Old* (absent) and *Proving\_Adequate* (new). We study how the classes are distributed in the clusters of the two tests (RQ1) and how the classes move from one cluster to another (RQ3).

The PCA and HAC analyses distinguish three clusters, called CI-A2 (9 classes), CI-B2 (13 classes) and CI-C2 (3 classes) (Fig. 10).

The characteristics of the three clusters in Test 1 (CI-A1, CI-B1, CI-C1) and those of the Test 2 (CI-A2, CI-B2, and CI-C2) are broadly comparable for all the variables that best contribute to the formation of the first factorial plane (Fig. 9), although the *Success* rate in Test 2 is higher than in Test 1 for each cluster (5 points higher for CI-A2 and CI-B2 and 2 points higher for CI-C2).

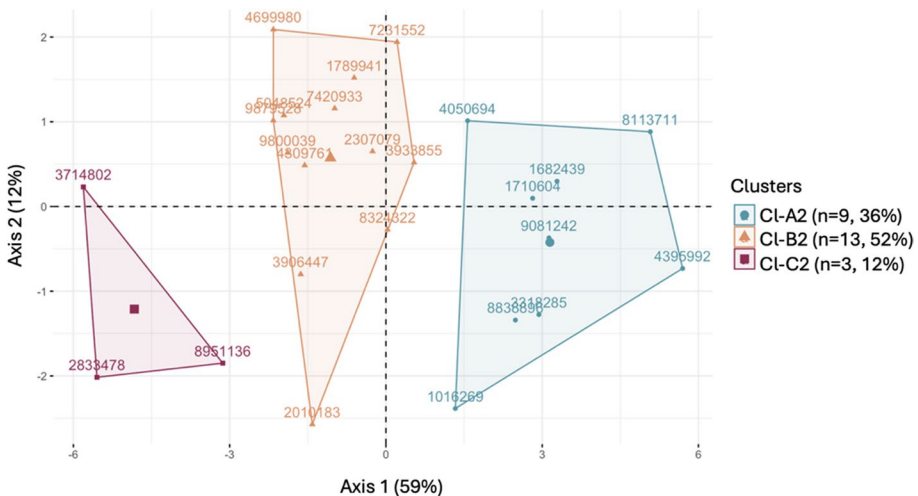


Fig. 10 Three clusters of similar classes on Test 2 (N=25)

**Table 9** Variation of the distribution by cluster of classes between Tests 1 and 2

Test	Frequency (%)			
	Test 1		Test 2	
Cluster	CI-A1	4 (16)	CI-A2 →	4 (16)
	CI-B1	14 (56)	CI-A2 ↗	5 (20)
			CI-B2 →	8 (32)
			CI-C2 ↘	1 (4)
	CI-C1	7 (28)	CI-B2 ↗	5 (20)
		CI-C2 →	2 (8)	
Total	Total of classes	25 (100)	Total of classes	25 (100)

Legend:

An arrow → indicates that classes remain in a cluster of equivalent level between Test 1 and Test 2 (e.g., classes in CI-A1 then in CI-A2)

An arrow ↗ indicates that classes progress to a higher-level cluster between Test 1 and Test 2 (e.g., classes in CI-B1 then in CI-A2)

An arrow ↘ indicates that classes regress to a lower-level cluster between Test 1 and Test 2 (e.g., classes in CI-B1 then in CI-C2)

**Table 10** Variation of the distribution by cluster of students between Test 1 and Test 2

Test	Frequency (%)			
	Test 1		Test 2	
Cluster	St-A1	116 (26)	St-A2 →	72 (16)
			St-B2 ↘	36 (8)
			St-C2 ↘	8 (2)
	St-B1	178 (39)	St-A2 ↗	38 (8)
			St-B2 →	73 (16)
			St-C2 ↘	67 (15)
	St-C1	160 (35)	St-A2 ↗	4 (1)
			St-B2 ↗	50 (11)
			St-C2 →	106 (23)
Total	Total of students	454 (100)	Total of students	454 (100)

Legend:

An arrow → indicates that students remain in a cluster of equivalent level between Test 1 and Test 2 (e.g., students in St-A1 then in St-A2)

An arrow ↗ indicates that students progress to a higher-level cluster between Test 1 and Test 2 (e.g., students in St-B1 then in St-A2)

An arrow ↘ indicates that students regress to a lower-level cluster between Test 1 and Test 2 (e.g., students in St-B1 then in St-C2)

The comparison of the number of classes in each cluster shows that the number of classes is greater in CI-A2 than in CI-A1 (4 CI-A1 and 9 CI-A2), less in CI-C2 than in CI-C1 (7 CI-C1 and 3 CI-C2), and closer in CI-B2 (13 classes) and CI-B1 (14 classes). Furthermore, as shown in Table 9, 10 classes advance by changing clusters, 5 from

CI-B1 to CI-A2 (especially for *Proving\_Adequate*), and 5 from CI-C1 to CI-B2. On average, by the end of the year, students in these five classes improve their arithmetic practices and begin to use the algebraic practices expected at this grade level to solve algebraic tasks, even if some algebraic techniques are still incorrect. Only one class regresses from CI-B1 to CI-C2 which means that, overall, this class uses more arithmetic practices than at the beginning of the year. The other classes progress within the same cluster, from CI-A1 to CI-A2 (4 classes) or CI-B1 to CI-B2 (5 classes) or CI-C1 to CI-C2 (2 classes). These classes perform better but with no significant change in their practices.

Statistical analysis has revealed three clusters of similar classes at the beginning and end of the year, answering to RQ1. The levels on the *Calculating-Numerically*, *Calculating-Algebraically*, *Modeling*, and *Representing* (i.e., 4 of the 5 local praxeologies) that best represent these clusters are consistent within each cluster, with rates at the *Adequate* or *Weakly\_Adequate* levels close to or above 50% for CI-A1 and CI-A2, between 30 and 50% for CI-B1 and CI-B2, and close to or below 30% for CI-C1 and CI-C2 (Fig. 9). These characteristics are consistent with the informal labels often used by teachers to designate “good” and “weak” classes, but they provide criteria for learning algebra and can guide teaching decisions to manage class heterogeneity.

To answer RQ3 about classes, comparing the overall distribution of classes within clusters shows that not all classes are moving in the same direction in terms of learning between the beginning and the end of the year. Classes in the CI-A1 cluster move to CI-A2, while a majority (5 out of 7) of CI-C1 classes move to CI-B2, marking an evolution in the *Calculating-Numerically* praxeology. CI-B1 classes evolve in very different ways.

These analyses are based on rates calculated by class, which do not provide any information about the students’ learning who make up these classes. This is why the following analyses focus on students (RQ2 and RQ3).

## 4 Results: similar students’ learned praxeologies in algebra and their variation during a school year

### 4.1 A multivariate analysis on test 1 results per student

We identify clusters of students with common learning characteristics on the S1 database. PCA identifies a first factorial plane (62% of information) with six variables that best contribute to its formations (Fig. 11): *Success*, *Calculating-Numerically\_Weakly-Adequate*, *Failure*, *Calculating-Numerically\_Under-Construction\_Old*, *Calculating-Algebraically\_Adequate*, and *Representing\_Adequate\_Weakly-Adequate*. As for classes, we interpret the first factorial plane as relating the *Adequate* or *Old*  $\theta$ -levels of algebraic learned praxeologies and  $\theta$ -levels of *Calculating-Numerically*. The HAC then leads to the identification of three clusters of students (Fig. 12), named St-A1 (26%), St-B1 (39%), and St-C1 (35%), and described in Fig. 13.

Cluster St-A1 includes students whose *Success* rate (60%) is very high compared to *Failure* rate (32%) and whose rates on the *Adequate* or *Weakly-Adequate*  $\theta$ -levels on three of the five local praxeologies (*Calculating-Numerically*, *Calculating-Algebraically*, and *Representing*) close to 60%. St-A1 students start the school year with already well-established algebraic practices.

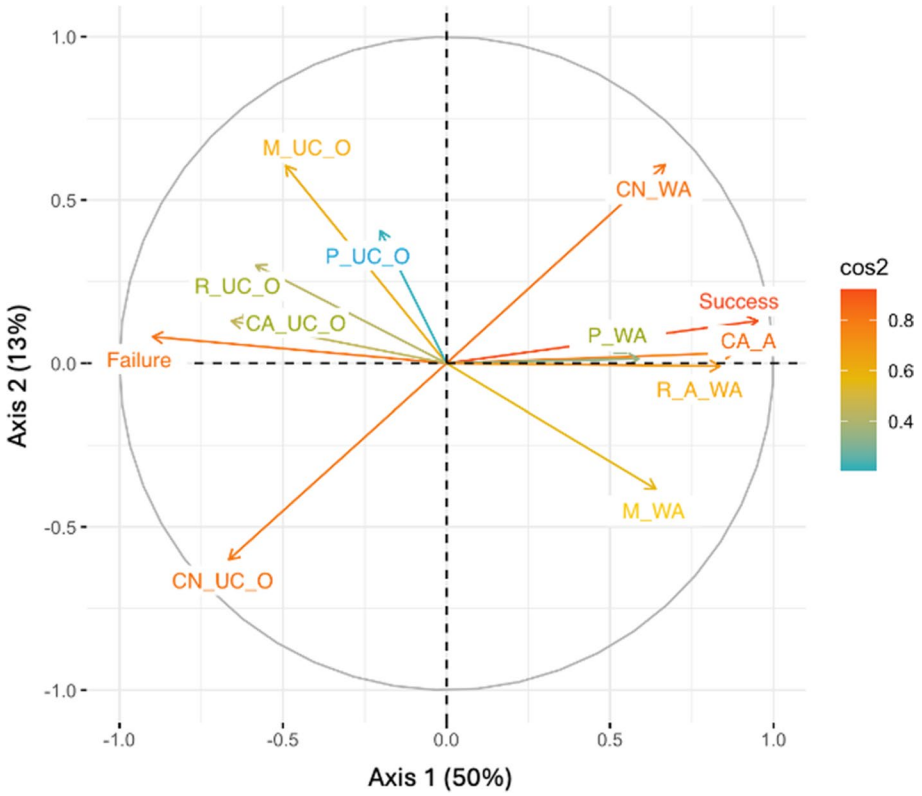


Fig. 11 Correlation circle in the first factorial plan for Test 1

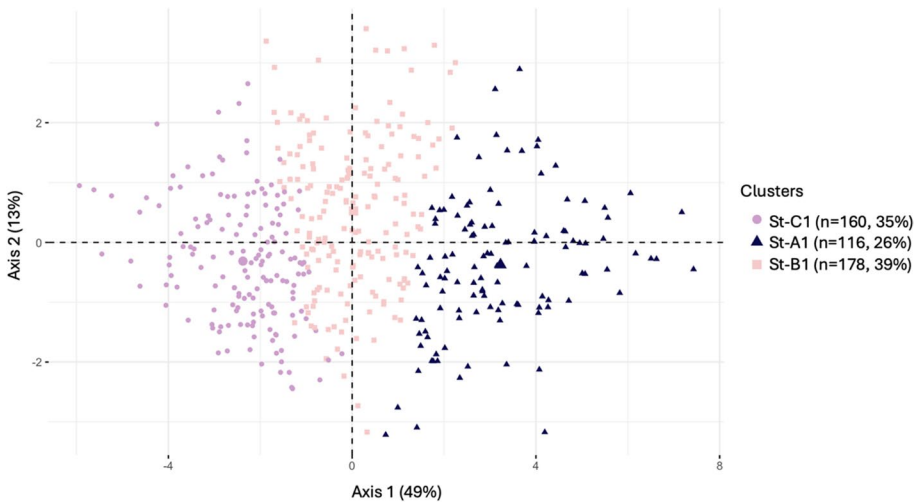
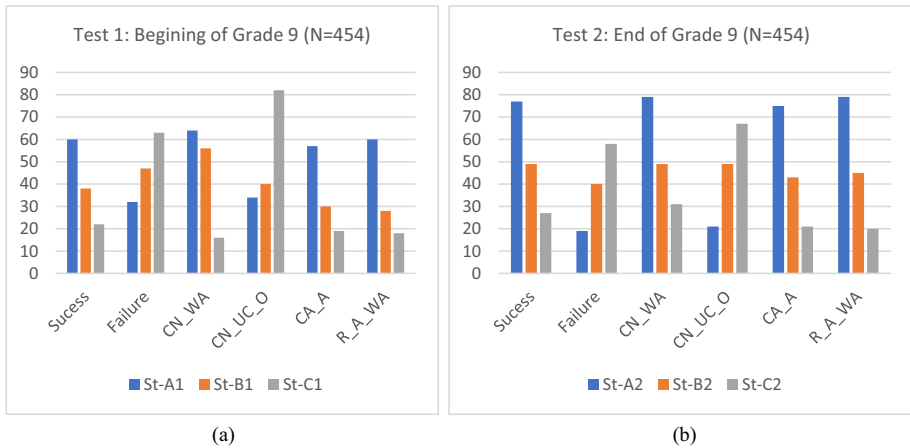


Fig. 12 Three clusters of similar students on Test 1 (N=454)



**Fig. 13** Percentages for the variables that best contribute to the formation of the first factorial plane on Test 1 (a) and Test 2 (b)

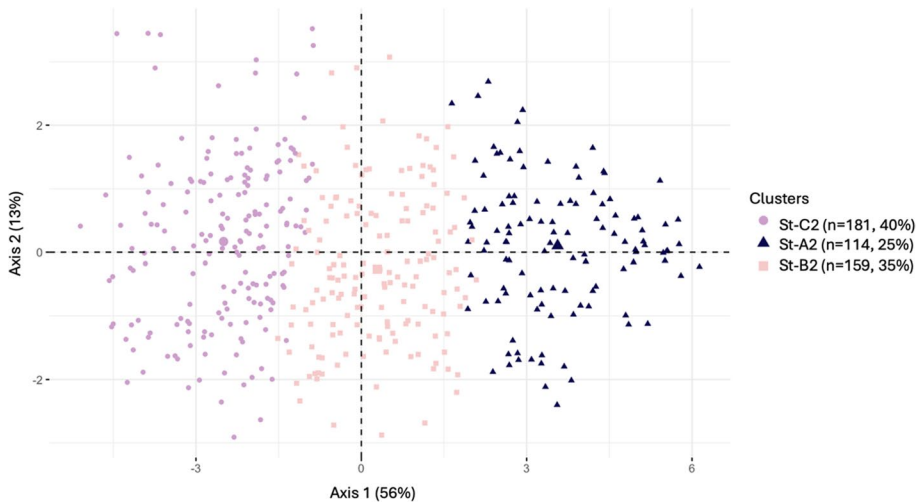
Cluster St-B1 is composed of students whose *Success* rate (38%) is lower than the *Failure* rate (47%) and whose rates on the *Adequate* or *Weakly-Adequate* are close to 30% on *Calculating-Algebraically* and *Representing*. The rate for *Adequate* and *Weakly-Adequate* (56%) in *Calculating-Numerically* is the highest among the *Adequate* and *Weakly-Adequate*  $\theta$ -levels of other praxeologies and is higher than for *Under-Construction* and *Old* (40%). Students in St-B1 start the year having built up *Calculating-Numerically* praxeologies, which are a possible support point for moving towards algebraic practices.

Cluster St-C1 is composed of students whose *Success* (22%) rate is very low compared to the *Failure* rate (63%) and whose rates on *Adequate* or *Weakly-Adequate* on three of the five local praxeologies (*Calculating-Numerically*, *Calculating-Algebraically*, and *Representing*) are less than 20%. *Calculating Numerically* is not a support for these students (82% on *Under-Construction* or *Old* levels against 16% on *Weakly-Adequate*). St-C1 students start the year well below school expectations.

## 4.2 Variation between tests 1 and 2

We performed the same analysis on S2 database. The variables that best contribute to the formation of the first factorial plane are the same. To answer RQ3 for students, we then compared how students are distributed in clusters on the two tests, and which cluster each student belongs to at the beginning and end of the year. Analogous analyses carried out on the S2 database led to three clusters, St-A2 (25%), St-B2 (35%), and St-C2 (40%) (Fig. 14).

The clusters for Test 1 and Test 2 are comparable across all variables (Fig. 13), but with more or less pronounced changes for certain variables depending on the clusters. The *Success* rate in Test 2 is higher than in Test 1 for each cluster (16 points higher for St-A2, 11 for St-B2, and 5 for St-C2). Students in St-C2 are making progress in *Calculating-Numerically*, with the rate on *Weakly-Adequate* almost doubling although still very low (31%), but not in *Calculating-Algebraically Adequate* and *Representing Adequate-Weakly-Adequate*. *Calculating-Numerically* praxeology is not always a point of support.



**Fig. 14** Three clusters of similar students for Test 2 ( $N=454$ )

The composition of the clusters varies slightly between the two tests. However, St-C2 has slightly more students than St-C1 (about 40% against 35%) and the reverse is true for St-B2 and St-B1 (35% versus 39%).

According to Table 10, 45% of the students progressed without changing cluster (16% from St-A1 to St-A2, 16% from St-B1 to St-B2, and 23% from St-C1 to St-C2), that is, there are no major advances in the  $\theta$ -levels they use to solve algebraic problems. Twenty percent of the students progressed with changing cluster (8% from St-B1 to St-A2, 1% from St-C1 to St-A2, 11% from St-C1 to St-B2). At the end of the year, these students belonged to clusters that had learned praxeologies closer to what is expected at the end of 9th grade. However, about 25% of students regressed and, at the end of the year, belonged to clusters that had built praxeologies that are different from what is expected (8% from St-A1 to St-B2, 15% from St-B1 to St-C2, and 2% from St-A1 to St-C2).

Taking these results together, we answer RQ2, indicating that the students' learned praxeologies fall into three clusters at the beginning and the end of the year with similar characteristics. They are consistent in terms of learning on the *Adequate* and/or *Weakly-Adequate* levels of the representative local praxeologies, *Calculating-Numerically*, *Calculating-Algebraically*, and *Representing*: rate above 50% for St-A1 and St-A2, rate between 30 and 50% for St-B1 and St-B2, and rate below 30% for St-C1 and St-C2 (Fig. 13). In the light of these results, we associate each cluster with a dominance described by the same level on these three local praxeologies: *Adequate* and/or *Weakly-Adequate* for St-A1 and St-A2, *Under-Construction* for St-B1 and St-B2, and *Old* for St-C1 and St-C2. Furthermore, we find that the *Modeling* and *Proving* praxeologies are not representative in the characterization of the clusters, which raises questions about the place attributed to these praxeologies in algebra teaching.

Between the beginning and the end of the school year, despite a higher success rate in Test 2, the proportion of students in each cluster remain fairly close (Table 10): A quarter of the students (St-A1 and St-A2) have learned the praxeologies expected at the end of middle school, between 35 and 40% of the students (St-C1 and St-C2) have old praxeologies, and between 39 and 35% of the students (St-B1 and St-B2) have praxeologies under



construction. However, some students change cluster (Table 10), with 20% progress and 25% regress. *Calculating-Numerically* is the central lever for moving from cluster St-C1 to St-B2. These results are comparable to those found in work using *Pépîte*, but they are richer because they indicate what is most representative of the students' learned praxeologies.

## 5 Variations in students' learning within the same class over the school year

We study links between variations in students' learning and the class effect (RQ4) during one year.

To do this, we go deeper into the characterization of similar classes for Test 1, particularly to improve the description of cluster CI-B1. For each class, we examine the distribution of students according to the cluster to which they belong.

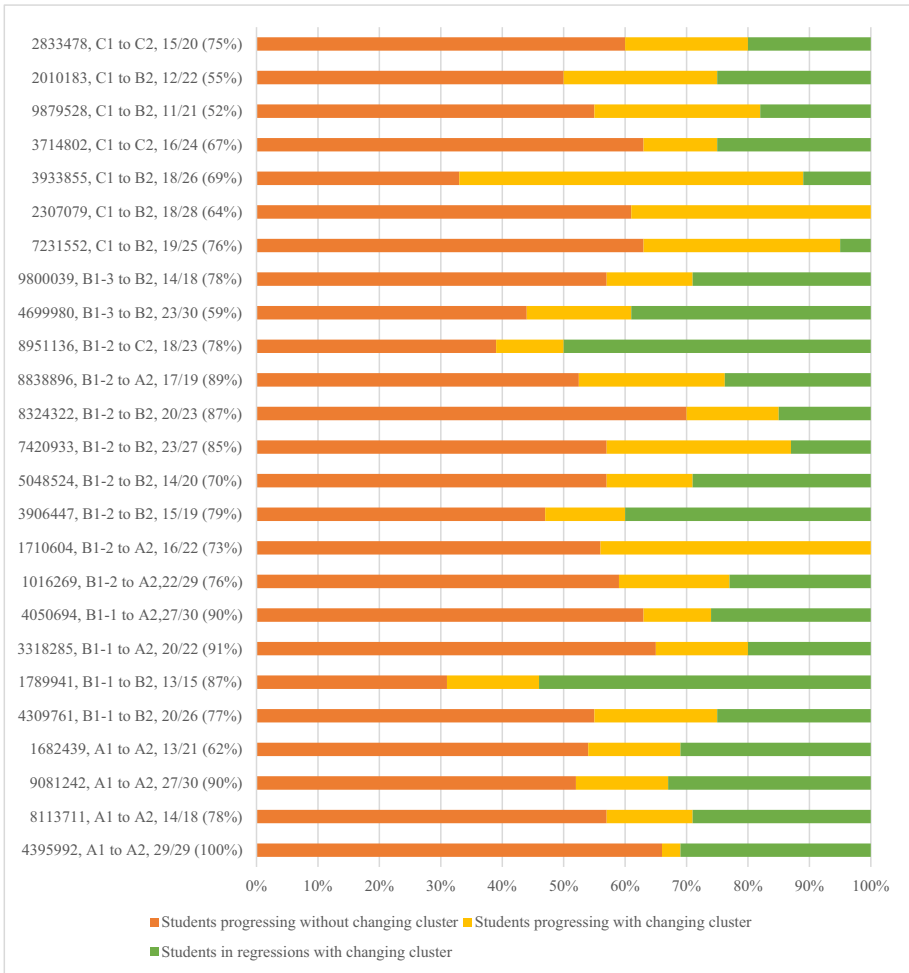
First, we combine multivariate descriptive analyses on classes and students to compare the distribution of students in similar classes on Test 1. In this way, we identify common characteristics in the distribution of students in classes of the same cluster, which we express in terms of threshold percentages of students (Table 11).

Two similar classes according to PCA and HAC have student distributions that follow the same trend for clusters CI-A1 and CI-C1 (Table 11), but with different trends for cluster CI-B1. For the classes in this cluster, we distinguish three intervals on St-A1 and St-C1 to better characterize the proximity of CI-B1 classes and to be more useful for interpreting learning variations.

Second, in order to further investigate the fact that similar classes do not evolve in the same way, Fig. 15 shows the variation of the distribution of students between Tests 1 and 2 within each class, distinguishing between students who progress without changing cluster (St-A1 to St-A2, St-B1 to St-B2, St-C1 to St-C2), with changing clusters (St-B1 to St-A2,

**Table 11** Class characterization for CI-A1, CI-B1 and CI-C1 clusters

Class clusters for Test 1	Cross-characterization with HCA student clusters on Test 1	New characterization	Number of classes
CI-A1	-St-A1 around half or more -St-A1 and St-B1 greater than 80%		4
CI-B1	-St-A1 less than 50% -St-A1 and St-B1 between around 50% and 80%	<b>Subgroup 1: CI-B1-1</b> -St-A1 between 40 and 50% -St-C1 less than 25%	4
		<b>Subgroup 2: CI-B1-2</b> -St-A1 between 20 and 40% -St-C1 between 25 and 50%	8
		<b>Subgroup 3: CI-B1-3</b> -St-A1 less than 20% -St-C1 between 25 and 50%	2
CI-C1	-St-C1 greater than 50% -St-A1 less than 10% -St-A1 and St-B1 less than 50%		7



**Fig. 15** For each class ( $N=25$ ), percentage of students progressing or regressing between the two tests, and percentage of students taking part in both tests

St-C1 to St-B2, St-C1 to St-A2) and those who regress (St-A1 to St-B2, St-A1 to St-C2, St-B1 to St-C2).

All but one class make progress, but not all classes show the same growth in student learning. To do this, we take into account the percentage of students present for both tests as new information about the class in Fig. 15. This one varies between 55 and 100% depending on the class. We comment on the most salient results.

The seven CI-C1 classes have a low rate of attendance (only 52% to 76% of students passing both tests), whereas CI-A1 classes have high rate of attendance (72% to 100% of students passing both tests). Students in CI-C1 classes seem to be progressing better overall than those in CI-A1 classes. Likewise, the students in two classes (1710604 in CI-B1-2, 2307079 in CI-C1) are all making progress, but around a third of the class is

missing. As absenteeism is very high, these results cannot be interpreted in relation to the initial class composition.

Students in five of the nine CI-A2 classes passed both tests at over 89%. Three of these classes progressed from CI-B1-1 or CI-B1-2 to CI-A2, with more students in these classes progressing (from 74 to 80%) than students in the other two classes (around 70%).

Although the progression is substantial from CI-C1 to CI-B2, for five of the seven classes, it is important to notice that only between half and three quarters of students are present for both tests. The progress of students who change clusters in these classes concerns 25% to 56% of them. However, the regression of students from one cluster to another is much lower than in the classes from CI-A1 or CI-B1, which may be related to absenteeism.

We cannot interpret these results further and answer RQ4 because absenteeism does not allow us to return to the initial composition of the class.

## 6 Discussion

The originality of this large-scale statistical study lies in its didactic foundations, in particular the characterization of elementary algebra by five local praxeologies and the hierarchy of students' reasoning and knowledge by four  $\theta$ -levels. This reference model of elementary algebra structures the databases and allows us to define similar classes and similar students, to study their variation for a year (RQ1, RQ2, and RQ3). The statistical analyses used, PCA and HAC, reflect the quality and relevance of the data coding. They are mutually consistent and interpretable with respect to the reference model (e.g., description of clusters).

Some of our choices could be questioned. Firstly, we grouped  $\theta$ -levels of the same praxeology across the analyses to address impossible to code or missing answers. This grouping is a limitation when it comes to characterizing clusters by praxeologies and the most representative levels associated with them. Secondly, we chose to build the class base by aggregating the coding of all students' responses to all items. Another option would have been to use student clusters, taking as a variable the number of students in each student cluster in each class. This would make it possible to interpret the similarity of the classes in terms of learning and their variation regarding the composition of the classes according to the number of students per cluster at the beginning of the year.

In addition, our study reveals a methodological limitation regarding the inclusion of student absenteeism in both tests which did not allow us to answer RQ4. Limiting absenteeism turns out to be an indispensable condition for studying the dependence, if any, between the evolution of students' learning and the class effect. These methodological comments led us to define the conditions for a new study, which could be a large-scale study on a representative sample of students and schools in France.

In addition to the composition of the class, according to student learning, we could take into account the teaching choices and practices of French teachers that influence the variation of students' learned praxeologies. The Praesco<sup>6</sup> study (Content-Specific Teaching Practices) (Coppé et al., 2021a, 2021b) has shown that there are four different clusters of teaching practices. They depend particularly on whether teachers

<sup>6</sup> PRAESCO: PRATIques d'Enseignement Spécifiques aux CONTENUS.

emphasize algebraic problem-solving or the technical aspect of algebraic calculation, and whether they take into account students' productions. To fully answer RQ4, we could examine the possible links between variations in similar classes over a school year and the teaching choices of their teachers according to the clusters of practices to which they belong.

## Appendix

### Examples of *Pépité* items

We present three items from *Pépité* (early 9th grade) and excerpts from their a priori analyses.

Item 1: Recognizing true numerical equalities.				
Tick the correct equalities				
1.1	<input type="checkbox"/> $10^3 \times 10^2 = 10^6$	<input type="checkbox"/> $10^3 \times 10^2 = 10^5$	<input type="checkbox"/> $10^3 + 10^2 = 10^5$	<input type="checkbox"/> $10^3 \times 10^2 = 10^{32}$
1.2	<input type="checkbox"/> $\frac{1}{2} + \frac{2}{3} = \frac{3}{5}$	<input type="checkbox"/> $\frac{1}{2} + \frac{2}{3} = \frac{3}{6}$	<input type="checkbox"/> $\frac{1}{2} + \frac{2}{3} = \frac{5}{6}$	<input type="checkbox"/> $\frac{1}{2} + \frac{2}{3} = \frac{7}{6}$
1.3	<input type="checkbox"/> $2 + 3 \times 4 = 14$	<input type="checkbox"/> $3 \times 4 + 2 = 14$	<input type="checkbox"/> $3 + 2 \times 4 = 20$	<input type="checkbox"/> $2 + 3 \times 4 = 20$

Choice	A priori analysis for item 1.2	Code
3/5	Addition of numerators and of denominators	V3 CN_O
3/6	Addition of numerators and product of denominators	V3 CN_UC
5/6	Incorrect rule for reducing to the same denominator	V3 CN_UC
7/6	Correct	V1 CN_A

Item 2: Determining whether a literal equality is always true.			
Indicate whether the following statements are true for all values of $a$ . Among the reasons offered, choose the most appropriate.			
	Equalities	True/False	Choose a reason
2.1	$a^2 = 2a$	<input type="radio"/> True <input type="radio"/> False	
2.2	$8 + 5a = 13a$	<input type="radio"/> True <input type="radio"/> False	
2.3	$4(3 + a) = 12 + a$	<input type="radio"/> True <input type="radio"/> False	

Item 2: Determining whether a literal equality is always true.		
Equalities	True/False	Choose a reason
2.1. $a^2 = 2a$	<input checked="" type="radio"/> True <input type="radio"/> False	None of these a squared if $a=2$ then 2 squared = 4, $2a$ is also written $2 \times a$ so, $2 \times 2 = 4$ $a \times a = a^2 \neq a + a = 2a$ $a$ squared is 2 times $a$ $2$ squared = $4 = 2 \times 2$ $a$ squared: multiplies $a$ by 2 and so does $2a$ Because $a$ is multiplied by its exponent
	<input type="radio"/> True <input checked="" type="radio"/> False	None of these $3^2 = 9$ et $2 \times 3 = 6$ or $9 \neq 6$ So $a^2 \neq 2a$ $a \times a = a^2 \neq a + a = 2a$ $a^n \neq a \times n$ When a number is squared, we multiply it by itself and not by the exponent Addition and multiplication must not be confused Because: $a^2 = a \times a$

A priori analysis of item 2.1 when a student selects "False"

Code

$a^2 = 2a$	Use of a counterexample to invalidate $a^2 = 2a$	V1 P_A
False	$(3^2 = 9 \text{ et } 2 \times 3 = 6 \text{ or } 9 \neq 6 \text{ So } a^2 \neq 2a)$	
	Algebraic justification by comparing definitions ( $a \times a = a^2 \neq a + a = 2a$ )	V1 P_A
	Algebraic justification by comparing general definitions ( $a^n \neq a \times n$ )	V1 P_A
	Justification by textual definition (When a number is squared, we multiply it by itself and not by the exponent)	V1 P_WA
	Justification by comparing operations (Addition and multiplication must not be confused)	V1 P_WA
	Incomplete algebraic justification (Because: $a^2 = a \times a$ )	V3 P_UC

**Item 3: Finding and recognizing the area of a rectangle.**  
*The student sees question 2 once he has answered question 1.*

**3.1/ Find the area of the large rectangle (in bold) for any value of x**

Draft	Result
<b>3.2/ Tick the expression(s) that give(s) the area of the large rectangle (in bold)</b>	
<input type="checkbox"/> $(x + 2)(x + 3)$ <input type="checkbox"/> $2x(x + 3)$ <input type="checkbox"/> $x + 2 \times x + 3$	<input type="checkbox"/> $x + 2(x + 3)$ <input type="checkbox"/> $2(2x + 5)$ <input type="checkbox"/> $x^2 + 5x + 6$

Possible solutions	A priori analysis of item 3.1	Coding
$(x+2)(x+3)$	Use the definition of the area of a rectangle (with commutativity)	V1, M_A, CA_A, R_A
$x^2+5x+6$	Expanded expression of area	V1, M_A, CA_A, R_A
$x+2(x+3)$	Use the definition of the area of a rectangle, but without using parentheses for the two factors	V3, M_A, CA_UC, R_UC
$x+2 \times x+3$	Use the definition of the area of a rectangle, but without using parentheses	V3, M_A, CA_UC, R_UC
$2(x+3+x+2)$	Confusion between area and perimeter	V3, M_O, CA_A, R_A
$2x \times 3x$	Abbreviated translation	V3, M_O, CA_O, R_O
$5x^2$		
$6x^2$		

**Acknowledgements** We would like to thank Muhammad Zubair Emritte, who carried out the statistical analyses during his end-of-study internship at Université Paris Cité. We would also like to thank the Department of Evaluation, Forecasting, and Performance (DEPP) of the Ministry of Education in France for their assistance with data collection.

**Funding** Open access funding provided by Université Paris-Est Créteil.

**Data availability** Data are not publicly available. In accordance with what has been communicated to families, and in agreement with the French Ministry of Education, students' answers to the tests are confidential.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Artigue, M., Grugeon, B., Assude, T., & Lenfant, A. (2001). Teaching and learning algebra: Approaching complexity through complementary perspectives. In H. Chick, K. Stacey, J. Vincent, & J. Vincent (Eds.), *The future of the teaching and learning of algebra. Proceedings of 12<sup>th</sup> ICMI Study Conference, The university of Melbourne, Australia, December 9–14, 2001* (pp. 21–32). Department of Science and Mathematics Education, University of Melbourne. <https://archive.org/details/proceedingsof12t0002unse/page/n3/mode/2up>
- Bednarz, N., Kieran, C., & Lee, L. (1996). *Approaches to algebra*. Kluwer. [https://doi.org/10.1007/978-94-009-1732-3\\_1](https://doi.org/10.1007/978-94-009-1732-3_1)
- Bosch, M., & Gascón, J. (2005). La praxéologie comme unité d'analyse des processus didactiques [Praxeology as a unit for analyzing didactic processes]. In A. Mercier & C. Margolinas (Eds.), *Balises pour la didactique : Cours de la 12<sup>ème</sup> école d'été de didactique des mathématiques* (pp. 107–122). La pensée sauvage.
- Bosch, M., Gascón, J., & Trigueros, M. (2017). Dialogue between theories interpreted as research praxeologies: The case of APOS and the ATD. *Educational Studies in Mathematics*, 95(1), 39–52. <https://doi.org/10.1007/s10649-016-9734-3>
- Bressoux, P. (2012). L'influence des pratiques enseignantes sur les acquisitions scolaires des élèves. [The influence of teaching practices on students' learning achievement]. *Regards Croisés Sur L'économie*, 12, 208–217. <https://doi.org/10.3917/rce.012.0208>
- Carraher, D. W., Schliemann, A. D., Brizuela, B. M., & Earnest, D. (2006). Arithmetic and algebra in early mathematics education. *Journal for Research in Mathematics Education*, 37(2), 87–115. <https://psycn.et.apa.org/record/2006-03239-003>
- Chenevotot-Quentin, F., Grugeon-Allys, B., Pilet, J., Delozanne, E., & Prévité, D. (2016). The diagnostic assessment Pépite and the question of its transfer at different school levels. In K. Krainer, & N. Vondrova (Eds.), *Proceedings of the 9th congress of the european society for research in mathematics education* (pp. 2326–2332). Charles University in Prague, Faculty of Education and ERME. <https://hal.science/CERME9>

- Chevallard, Y., & Bosch, M. (2020a). Didactic transposition in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 214–218). Springer. [https://doi.org/10.1007/978-3-030-15789-0\\_48](https://doi.org/10.1007/978-3-030-15789-0_48)
- Chevallard, Y., & Bosch, M. (2020b). Anthropological theory of the didactic (ATD). In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 53–61). Springer. [https://doi.org/10.1007/978-3-030-15789-0\\_100034](https://doi.org/10.1007/978-3-030-15789-0_100034)
- Chevallard, Y. (1985). Le passage de l'arithmétique à l'algébrique dans l'enseignement des mathématiques au collège – Première partie : L'évolution de la transposition didactique [The transition from arithmetic to algebra in middle school mathematics teaching – Part I: The evolution of didactic transposition]. *Petit x*, 5, 51–94. [https://irem.univ-grenoble-alpes.fr/medias/fichier/5x3\\_1570714298158-pdf](https://irem.univ-grenoble-alpes.fr/medias/fichier/5x3_1570714298158-pdf)
- Chevallard, Y. (1989). Le passage de l'arithmétique à l'algébrique dans l'enseignement des mathématiques au collège – deuxième partie : Perspectives curriculaires, la notion de modélisation [The transition from arithmetic to algebra in middle school mathematics teaching – part two: Curricular perspectives, the notion of modeling]. *Petit x*, 19, 43–72. [https://irem.univ-grenoble-alpes.fr/medias/fichier/19x5\\_1570440008367-pdf](https://irem.univ-grenoble-alpes.fr/medias/fichier/19x5_1570440008367-pdf)
- Coppé, S., Grugeon-Allys, B., Horoks, J., Pilet, J., Solnon, A., Raffaëlli, C., & Charpentier, A. (2021a). Premiers résultats de l'enquête sur les pratiques d'enseignement des mathématiques, Praesco en classe de troisième en 2019. [First results of the survey on mathematics teaching practices, Praesco in the ninth grade in 2019]. *Note d'Information de la Depp*, 21.11. <https://www.education.gouv.fr/media/74626/download>
- Coppé, S., Grugeon-Allys, B., Horoks, J., Pilet, J., Solnon, A., Raffaëlli, C., & Charpentier, A. (2021b). Document de travail: Premiers résultats de l'enquête sur les pratiques d'enseignement des mathématiques, Praesco en classe de troisième en 2019, correspondant à la Note d'information de la Depp n°21.11. [Working document "First results of the survey on mathematics teaching practices, Praesco in the third grade in 2019" corresponding to Note d'information de la Depp n°21.11.]. *Série Étude*, 2021-E02. <https://www.education.gouv.fr/media/74607/download>
- Drouhard, J-P. (1992). *Les écritures symboliques de l'algèbre élémentaire* [The symbolic writings of elementary algebra][Doctoral dissertation, Université Paris 7 Denis-Diderot]. Université Paris 7 Denis-Diderot. <https://theses.hal.science/tel-00925358>
- Frege, G. (1971). *Écrits logiques et philosophiques* [Logical and philosophical writings] (C. Imbert, Trans.). Éditions du Seuil (Original work published in 1892).
- Grugeon, B. (1997). Conception et exploitation d'une structure d'analyse multidimensionnelle en algèbre élémentaire [Design and operation of a multidimensional analysis structure in elementary algebra]. *Recherches en Didactique des Mathématiques*, 17(2), 167–210. <https://revue-rdm.com/1997/conception-et-exploitation-d-une/>
- Grugeon-Allys, B., Chenevotot-Quentin, F., Pilet, J., & Prévit, D. (2018). Online automated assessment and student learning: the PEPITE project in elementary algebra. In L. Ball, P. Drijvers, S. Ladel, H.-S. Siller, M. Tabach, & C. Vale (Eds.), *Uses of technology in primary and secondary mathematics education, ICMI-13* (pp. 245–266). Springer. <https://doi.org/10.1007/978-3-319-76575-4>
- Grugeon-Allys, B., Chenevotot-Quentin, F., & Pilet, J. (2022). Using didactic models to design adaptive pathways to meet students' learning needs in an on-line learning environment. In P. R. Richard, M. P. Vélez, S. Van Vaerenbergh (Eds.), *Mathematics education in the age of artificial intelligence—how artificial intelligence can serve mathematical human learning?* (pp. 157–188). Springer. <https://doi.org/10.1007/978-3-030-86909-0>
- Grugeon-Allys, B. (2016). Modéliser le profil diagnostique des élèves dans un domaine mathématique et l'exploiter pour gérer l'hétérogénéité des apprentissages en classe : une approche didactique multidimensionnelle [Modeling the diagnostic profile of students in a mathematical domain and using it to manage heterogeneous learning in the classroom: a multidimensional didactic approach]. *Evaluer - Journal International de Recherche en Education et Formation*, 2(2), 63–88. <https://journal.admee.org/index.php/ejiref/article/view/72/57>
- Hahn, C., & Macé, S. (2017). *Méthodes statistiques appliquées au management* [Statistical methods applied to management]. Pearson Education France.
- Kieran, C. (1992). The learning and teaching of school algebra. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). Macmillan.
- Kieran, C. (2007). Learning and teaching algebra at the middle school through college levels. Building meaning for symbols and their manipulation. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (vol. 2, pp. 707–762). I.A.P.
- Kieran, C. (2018). Teaching and learning algebraic thinking with 5- to 12-year-olds: The global evolution of an emerging field of research and practice. *Springer*. <https://doi.org/10.1007/978-3-319-68351-5>



- Kieran, C., Pang, S., Schifter, D., & Fong Ng, S. (2016). *Early algebra. Research into its nature, its learning, its teaching*. Springer. <https://doi.org/10.1007/978-3-319-32258-2>
- Le Cam, M., & Salles, F. (2020). TIMSS 2019 Mathématiques au niveau de la classe de quatrième : Des résultats inquiétants en France. [TIMSS 2019 Mathematics at the eighth grade level: Worrying results]. *Note d'Information de la DEPP*, 20(47), 1–4. <https://www.education.gouv.fr/timss-2019-mathematiques-au-niveau-de-la-classe-de-quatrieme-des-resultats-inquietants-en-france-307819>
- Nye, B., Konstantopoulos, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, 26, 237–257. <https://doi.org/10.3102/01623737026003237>
- Radford, L. (2013). The progressive development of early embodied algebraic thinking. *Mathematics Education Research Journal*, 26(2), 257–277. <https://doi.org/10.1007/s13394-013-0087-2>
- Ruiz-Munzón, N., Bosch, M., & Gascón, J. (2013). Comparing approaches through a reference epistemological model: the case of school algebra. In B. Ubuz, Ç. Haser & M. A. Mariotti (Eds), *Proceedings of the 8th Conference of the European Society for Research in Mathematics Education* (pp. 2870–2979). Middle East Technical University. [http://cerme8.metu.edu.tr/wgpapers/WG16/WG16\\_Bosch.pdf](http://cerme8.metu.edu.tr/wgpapers/WG16/WG16_Bosch.pdf)
- Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22, 1–36. <https://doi.org/10.1007/BF00302715>
- Vergnaud, G. (2009). The theory of conceptual fields. *Human Development*, 52(2), 83–94.
- Vergnaud, G., Cortes, A., & Favre-Artigue, P. (1987). Introduction de l'algèbre auprès de débutants faibles. Problèmes épistémologiques et didactiques [Introducing algebra to weak beginners. Epistemological and didactic problems]. *Actes du colloque de Sèvres : Didactique et acquisition des connaissances scientifiques* (pp. 259–288). La Pensée Sauvage.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.