Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Reliability and validity of a tool for analyzing resource potential for mathematical activity

Mickael Da Ronch

mickael.daronch@hepvs.ch

University of Teacher Education Valais

Marie-Line Gardes

University of Teacher Education Vaud

Ismaïl Mili

University of Fribourg

Research Article

Keywords: Reliability of tool, Validity of tool, Problem solving, Didactic and research variable, Resource analysis

Posted Date: April 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9538667/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Reliability and validity of a tool for analyzing resource potential for mathematical activity

Mickael Da Ronch[1*†], Marie-Line Gardes[2†] and Ismaïl Mili[3†]

[1*] University of Teacher Education Valais, Av. du simplon 13, Saint-Maurice, 1890, Switzerland.
[2]University of Teacher Education Vaud, Av. de Cour 33, Lausanne, 1007, Switzerland.
[3]University of Fribourg, Av. de l'Europe 20, Friboug, 1700, Switzerland.

*Corresponding author(s). E-mail(s): mickael.daronch@hepvs.ch;
Contributing authors: marie-line.gardes@hepl.ch; ismail.mili@unifr.ch;
†These authors contributed equally to this work.

Abstract

This paper presents a tool for effectively assessing the potential of resources to promote mathematical activity in problem-solving within textbooks. This tool is based on a theoretical framework grounded in the concepts of didactic variables and research variables derived from the Theory of Didactical Situations (TDS). The tool was tested through the analysis, using triple-blind coding, of a large corpus of 312 problems from Cycle 2 (pupils aged 8 to 12), taken from the official textbook used in French-speaking Switzerland (MER). The calculation of Fleiss's kappa coefficient combined with a Z-test showed that the criteria used to evaluate the resources in this corpus are reliable ($\boldsymbol{\kappa \approx 0.69}$, $\boldsymbol{z \approx 23.09}$, p-value $<$ **0.001**). Furthermore, the validity of our tool was confirmed by functional *a priori* analyses carried out on prototypical problems belonging to each of the four problem classes identified in this corpus. These results therefore show that our tool is reliable and valid and can therefore be reused for other international research into problem-solving in textbooks.

Keywords: Reliability of tool • Validity of tool • Problem solving • Didactic and research variable • Resource analysis

1 Introduction

Many contemporary mathematics curricula advocate the integration of inquiry-based activities in the classroom, aiming to simulate authentic mathematical practices. This curricular orientation, observed across several education systems (e.g., France, the United States, Singapore, Turkey), positions problem solving as a central competence involving analysis, interpretation, solution development, and reflection. It is also a key component of international large-scale assessments such as PISA and TIMSS, where mathematical literacy is defined as the ability to reason, formulate, and apply mathematics in a variety of contexts (OECD, 2023). Furthermore, numerous international studies highlight the prominent role of problem solving within curricula and its development across different countries. These studies point to a convergence in educational approaches, particularly in Australia, Singapore, USA and UK (Burkhardt & Bell, 2007; Schoenfeld, 2007; Stacey, 2005), as well as in Mexico (Santos-Trigo, 2007) for instance, where problem solving is considered a structuring element of mathematics education.

These requirements were coupled with the development of many educational resources, whether public, institutional, or offered by private publishers, which were made available to teachers to support the implementation of such activities. Some of these resources focus specifically on this type of activity (Da Ronch, Gandit, & Gravier, 2020; Gardes, 2013; Gravier & Ouvrier-Buffet, 2022; Grenier & Payan, 1998; Liljedahl, Santos-Trigo, Malaspina, & Bruder, 2016; Ouvrier-Buffet, 2011; Schoenfeld, 1985, 1992a), while others offer a mix of activities, accompanied by a categorization identifying those likely to involve research activity (Boaler, 1998; Borasi, 1986; Jäder, Lithner, & Sidenvall, 2020). Others finally restrict themselves to a more implicit presentation of this aim in the form of challenges described by Schoenfeld (1985) or explorations present in NCTM (2000). This heterogeneity raises the issues about the criteria used to classify a statement as a "research" activity, and about how these criteria are explained in the proposed materials.

In the French-speaking part of Switzerland, the *Plan d'Etudes Romand* (thereafter PER) is an official program that focuses significantly on problem solving (CIIP, 2023). It is presented as a means of "promoting a research attitude [...] through trial and error, generalization, conjecture and validation" (authors' translation). The stated objective is to bring mathematical activity, and specifically problem solving, closer to a scientific approach. Questions must be identified and situations problematized. Tools and approaches must be brought into play (CIIP, 2023). Problem formulation and problem solving thus appear to be recurring elements of institutional requirements. In this context, the *Moyens d'Enseignement Romands* (hereinafter MER, understood as the only officially recognized resource, subject to teachers' pedagogical freedom) are designed in line with the PER's recommendations. As such, they propose strategies and tasks aimed at promoting research activity, through internal categorizations intended to guide teachers' work.

These claims raise a key question: how can we identify, among the activities offered by these resources, those that effectively develop a research mindset in students? This is the perspective taken by our research questions, in line with the work of Da Ronch, Gardes, and Mili (2023) and Mili, Da Ronch, and Gardes (2025) : what theoretical and

methodological tools would enable us to identify, among the resources made available to teachers, those activities likely to stimulate genuine research activity? Implicit in this is the question of the validity of the categorizations proposed by the resources themselves: do activities labeled as research-based fulfill this ambition?

2 Conceptual framework

2.1 Mathematical research activity

For mathematicians (Garey & Johnson, 1979; Hadamard, 1945; Halmos, 1980; Polya, 1945) and mathematics educators (Brousseau, 1997; Da Ronch, 2022; English & Gainsburg, 2016; Gardes, 2013; Lesh & Zawojewski, 2007; Lester & Kehle, 2003; Liljedahl & Santos-Trigo, 2019; Schoenfeld, 1985, 1992b) mathematical research is closely related to problem solving. According to Garey and Johnson (1979), mathematical problems can be understood as a pairing of a general question[1] with a set of instances.

> For our purposes, a problem will be a general question to be answered, usually possessing several parameters, or free variables, whose values are left unspecified. A problem is described by giving: a general description of all its parameters, and a statement of what properties the answer, or solution, is required to satisfy. An instance of a problem is obtained by specifying particular values for all the problem parameters (Garey & Johnson, 1979, p. 4).

Therefore, doing mathematics means solving problems and accepting scientific responsibility in the pursuit of truth and proof in relation to a given problem (Da Ronch, 2022; Deloustal-Jorrand, Gandit, & Mesnil, 2025). This activity should therefore facilitate the implementation of various problem-solving processes, including experimentation, formulation and validation (Brousseau, 1997). These processes are essential for learning institutional mathematical knowledge. They enable the mobilization of abilities specific to the practice of mathematical activities. For instance, studying specific cases when approaching problem solving is a form of knowledge related to experimentation. The formulation of conjectures based on the study of these cases and the way in which mathematical objects are represented — which can sometimes lead to modelling premises — are examples of knowledge linked to the formulation process. Finally, mathematical proofs rely on logical rules, such as the law of the excluded middle and the principle of non-contradiction, as well as mathematical reasoning, such as direct implication, reductio ad absurdum, induction and counterexample. These enable us to argue and prove whether a mathematical statement is true or false.

2.2 What theoretical framework can be used to consider how this research approach might be applied to pupils, and how this might be achieved?

A central issue in contemporary mathematics education concerns how research-oriented activities, as promoted in curricula, can be effectively translated into

[1] A problem can be described either as an imperative sentence making a request, or as a mathematical statement of conjecture. In any case, it can always be rephrased as a question (Da Ronch, 2022).

classroom practice. This process of didactic transposition is strongly informed by theoretical frameworks in mathematics education, which provide tools to analyse and design teaching situations (Santos-Trigo, 2020). In the French context (Artigue & Houdement, 2007), two major frameworks have been identified as particularly influential for modelling mathematical research activity in teaching: the Theory of Didactical Situations (TDS) developed by Brousseau (Brousseau, 1997) and the Anthropological Theory of Didactics (ATD) developed by Chevallard et al. (Chevallard et al., 2022). Other traditions also contribute to this reflection. For instance, the Dutch approach of Realistic Mathematics Education emphasizes modelling grounded in real-world contexts (Santos-Trigo, 2020), while in China, problem solving tends to focus more on experiential engagement than on the explicit analysis of cognitive processes (Cai & Nie, 2007).

Beyond these theoretical traditions, a substantial body of international research has examined mathematical problem solving as a process (Liljedahl et al., 2016; Rott, Specht, & Knipping, 2021; Schoenfeld, 1985, 1992b, 2015; Selden, Mason, & Selden, 1989). From this perspective, learning mathematics is not limited to mastering definitions or applying theorems to obtain correct answers. Instead, emphasis is placed on the processes involved in inquiry, particularly heuristic strategies (Rott, 2014). These strategies must gradually evolve: they become decontextualized, depersonalized, and institutionalized as shared mathematical knowledge. Consequently, knowledge is not reduced to individual or local experience but acquires a broader and more general status. In this view, the final solution to a problem becomes secondary to the development of inquiry processes (Chevallard, 1998).

These processes are closely related to the dialectics identified by Brousseau 1997, namely action, formulation, and validation. Engaging in "doing mathematics" involves experimenting, conjecturing, modelling, proving, and interacting with others, while progressively constructing and appropriating mathematical concepts and tools (Brousseau, 1997). For example, exploring simplified cases of a problem reflects experimentation; formulating conjectures or changing representations relates to formulation; and constructing proofs relies on logical reasoning, such as necessary and sufficient conditions (Da Ronch, 2022; Deloustal-Jorrand et al., 2025).

Within this perspective, TDS provides a particularly relevant framework for designing classroom situations that foster inquiry. Didactic situations are intentionally constructed to confront students with a challenging environment in which they must take responsibility for validating their own actions (Brousseau, 1997). Rather than simply following teacher instructions, students are expected to develop, test, and refine their own strategies. This creates strong parallels between students' activity in such situations and that of mathematical researchers. In both cases, individuals face problems without immediate solutions, formulate conjectures, explore different approaches, interpret feedback, and revise their reasoning. Knowledge emerges progressively through interaction with the problem, rather than being directly transmitted.

A key element in this analogy is the role of the environment. In mathematical research, reasoning develops through interaction with definitions, representations, examples, and counterexamples. Similarly, in adidactic situations, the environment is carefully designed to provide meaningful feedback on students' actions, helping them

regulate their problem-solving processes. Moreover, this environment evolves through successive interactions: certain variables become salient, new patterns emerge, and initial approaches may be abandoned. This dynamic closely mirrors the evolving nature of research activity, where the problem itself may be redefined over time.

Several teaching models explicitly build on these principles. Research Situations for the Classroom (RSC) aim to place students in the role of researchers, engaging them in processes such as conjecturing, modelling, defining, and proving (Da Ronch, 2022; Da Ronch & Hocquard, 2026; Gravier & Ouvrier-Buffet, 2022; Ouvrier-Buffet, 2011, 2020). These situations are often connected to genuine research questions, encouraging movement between specific cases and generalization. Similarly, the Problem-Based Learning Situation (PBLS) model developed by the DREAM group (Front, 2015; Gardes, 2013; Loisy et al., 2019) focuses on fostering a dialectic between heuristic development and knowledge acquisition.

A common feature of these approaches is the emphasis on an experimental relationship to mathematics (Gardes, 2018), where students navigate between concrete cases, mathematical objects, and emerging theoretical ideas. Central to this design is the choice of problem: tasks must be sufficiently open to allow multiple strategies and to support the articulation between the particular and the general. Such openness is governed by what can be termed research variables, whose values influence the types of inquiry processes students engage in, similarly to didactic variables in TDS (Brousseau, 1997)

Overall, these frameworks and models highlight the importance of designing learning environments that support authentic mathematical activity. By engaging students in non-routine problems that require exploration, conjecture, and validation, they foster deeper mathematical understanding and align classroom practices with the nature of mathematical research (Selden et al., 1989).

2.3 Didactic variable and research variable

Brousseau (1997) introduces the concept of didactic variables as a specific type of situational variable. These are variables that are relevant to a given situation and whose values can influence the hierarchy of problem-solving strategies (e.g. cost, validity and complexity), as well as the strategies themselves. Teachers can adjust these variables in line with their learning objectives.

For example, in the following problem — "Determine the sum of p consecutive numbers starting from n as quickly as possible" — the value of p is a teaching variable whose relevant values are: 10; a number other than 10 that is not a multiple of 10; and a multiple of 10 (Durand-Guerrier, 2010, 2025). If the value of p is 10, several methods are possible: (1) breaking down each number into tens and units, leading to the identification of the calculation method "multiply the first number by 10 and add 45"; (2) adding 5 to the fifth number in the list; (3) breaking down the numbers starting from the first number in the series $(p, p+1, p+2, \ldots)$, leading to the formula $10p+45$; (4) grouping the numbers into pairs of equal sums, in the manner of Gauss; and (5) taking the median value and multiplying by 10. If p is a multiple of 10, all these methods can be used to quickly determine the sum of p consecutive integers starting from n.

Conversely, if p is not a multiple of 10, the methods relying on the properties of the decimal number system (methods 1 and 2) cannot be generalised. However, the methods relying on the concept of the successor can be (methods 3, 4 and 5). Further details on this situation can be found in Barallobres (2007); Barallobres and Giroux (2008); Durand-Guerrier (2010) and DREAM (2020).

Within the framework of RSC, Godot (2005) introduces the concept of research variables. These are the didactic variables in the task for which the pupil is responsible for choosing the values. According to Gravier and Trouvé (2025), research variables are specific didactic variables that enable pupils to grasp the particular-general dialectic. Figure 1 illustrates the interplay between didactic variables and research variables.

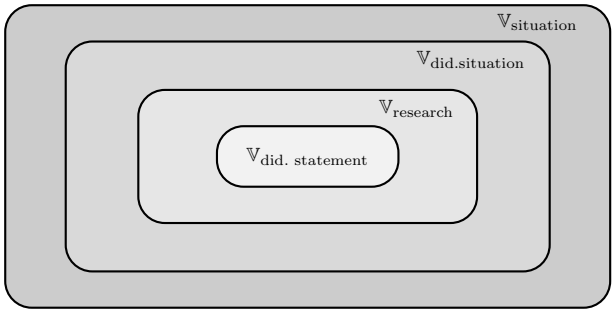


Fig. 1: Interplay between didactic variables and research variables – $\mathbb{V}_{\text{situation}} \supset \mathbb{V}_{\text{did.situation}} \supset \mathbb{V}_{\text{research}} \supset \mathbb{V}_{\text{did. statement}}$

Gravier and Trouvé (2025) classify research variables into two types:

- restricting the possible values or specifying a value to focus on a particular instance of the problem under consideration,
- extending the possible values to consider the problem under consideration as a sub-problem of a more general problem (Gravier & Trouvé, 2025, p. 116).

To encourage pupils to develop a particular-general dialectic in their mathematical work, teachers can vary the values of these research variables. These authors identify two possibilities:

- If pupils are allowed to choose the value themselves, it is referred to as a variable whose value is not fixed.
- If pupils are allowed to change the value initially specified in the problem, however, we refer to it as a variable whose value is fixed but can be altered (Gravier & Trouvé, 2025, p. 116).

Returning to the previous example (determine the sum of p consecutive numbers starting from n as quickly as possible), the teaching variable "value of p" is a research variable since it allows pupils to explore the relationship between the specific and the general. Studying the specific value of 10, for example, enables pupils to identify initial invariants, such as realizing that the sum of the units is always 45 and that the sum of the tens is always the first number in the list, or finding the median and multiplying it

by 10. These invariants can be identified by studying several series of ten consecutive integers, *i.e.* by varying n (the first given number). Therefore, the value of n is also a variable to be investigated. Taking a value of p that is not a multiple of 10 — for example, 8 — leads pupils to apply their method again in an attempt to generalize it. For instance, 8 times the first number plus the sum of the units always equals 28 or taking the median and multiplying by 8. By varying the value of p in this way, depending on their level, pupils can derive the following general formulas: p times the first number plus the sum of the units from 1 to $(p-1)$, or calculating the median (the sum of the first number plus $\frac{p-1}{2}$ and multiplying by p. Thus, in this problem statement, the two variables n and p are unfixed search variables.

This example demonstrates how the concepts of the didactic and research variables can be used to develop teaching situations that explore the specific-general dialectic inherent in mathematical activity. Returning to our research question, we will now examine the importance of these two concepts to identify the statements in a teaching resource that encourage pupils to engage in mathematical research.

2.4 Research variable: a useful tool for investigating a problem's potential to support mathematical activity

Gravier and Trouvé (2025) note that the presence of a research variable in a problem statement ensures the transposition of the particular-general dialectic. We have therefore adopted this concept in designing our tool for analyzing the potential of a resource for mathematical activity. To this end, we identified the following criterion: the presence and status of a research variable within the problem statement. We have developed four indicators to help us determine whether the criterion is met.

- The problem statement contains either an unspecified search variable or a specified variable that can be changed; the choice of values for this variable is left to the pupils. Examples include the problem statement "Quickly determine the sum of p consecutive integers starting from n", which contains two unfixed search variables (n and p).
- The problem statement contains a fixed-but-adjustable search variable, for which different values are given in the statement or suggested in the scenario provided to the teacher. For example, in Statements 2 and 3 below, one or other of the variables is fixed but adjustable, and different values are suggested.
  - Statement 2. Find the sum of the ten consecutive integers in the following sequences as quickly as possible:
    - Series A: 17, 18, 19, 20, 21, 22, 23, 24, 25, 26.
    - Series B: 23, 24, 25, 26, 27, 28, 29, 30, 31, 32.
    - Series C: 447, 448, 449, 450, 451, 452, 453, 454, 455, 456.

    In this problem, the value of p is fixed and the value of n varies. Certain values are suggested to encourage pupils to identify patterns (e.g. the sum of the units is always 45 for any value of n).
  - Statement 3. Determine the sum of ten consecutive integers starting from 17 as quickly as possible. Determine the sum of eight consecutive integers starting from

41 as quickly as possible. Determine the sum of 125 consecutive integers starting from 23 as quickly as possible.
In this question, n is set to "not a multiple of 10" and p varies, encouraging pupils to generalize their calculation methods as discussed in the previous section.

- The problem statement contains a didactic variable, some values of which could be considered research variables. However, these values are neither specified nor identified in the statement or the resource. It is therefore up to the teacher to identify them. This is the case with Statement 4, for example, which asks pupils to determine the sum of the ten consecutive integers starting from 17 as quickly as possible. The values of n and p are fixed in this statement, which does not allow pupils to engage in a particular-general dialectic. The values provided are too small to prompt a search for patterns or generalization. Similarly, if the problem of the 99 squares were reduced to 7 squares, the particular-general dialectic would not be necessary to solve the problem.
- The problem statement does not contain any instructional variables that could be characterized as research variables. The following example illustrates this: Camille has a bag that weighs 2.8 kg, and she adds 1.3 kg of fruit that she has just bought from the market. What is the weight of her bag now? The choice of numbers is an instructional variable whose values can influence pupils' calculation procedures. For instance, an understanding of decimal number rules is required to calculate $2.8+1.3$, whereas this is not the case for $2+1$. However, no value of this variable will enable pupils to engage in dialectical thinking, moving from the particular to the general.

These four indicators, together with this criterion, constitute an educational tool suitable for exploring potential issues when addressing the particular-general dialectic in mathematical activities. This approach does not require a comprehensive *a priori* analysis such as those described in Artigue (2015, 2020). It is just limited to the impact of variations in the problem's parameters on the procedures.

2.5 Detailed research questions

We have just shown the theoretical validity of this teaching tool. The next step is to test it against empirical data by analyzing a corpus. This leads us to the following research questions.

- **RQ1 – Validity:** Is the mere presence or absence of a research variable sufficient to identify issues that might trigger the dialectic of experimentation, formulation and validation inherent in mathematical research (and mirroring that of the researcher)? In other words, is this tool valid?
- **RQ2 – Reliability:** Is it reliable?

3 Method

3.1 Overview of the corpus

The French-speaking Swiss corpus has been selected for the purpose of evaluating the validity and reliability of the tool that has been developed. There are several

reasons for this decision. Firstly, the French-speaking Swiss Curriculum (PER) explicitly promotes a dialectic of experimentation–formulation–validation, and developing a research-oriented attitude is closely linked to problem-solving, which aligns perfectly with our theoretical framework. Secondly, in this context, a unique set of institutional teaching resources exists that are explicitly designed to align with the PER. These resources are structured around thematic areas, one of which is specifically dedicated to problem-solving.

As mentioned in the introduction, Switzerland's federal structure means that all matters relating to education have been delegated to the provinces, known as "cantons". In 2006, a new curriculum known as the "Plan d'Etudes Romand" (hereafter referred to as the PER) was introduced. The CIIP (a multi-provincial conference) was then tasked with developing teaching materials (hereafter referred to as MER : *Matériel d'Enseignement et de Ressources*[2]) in accordance with the PER. From 2016 onwards, the various provinces were able to introduce these new resources at each stage of schooling according to their own timetable.

Teachers can adapt this reference tool to suit their needs, while retaining their pedagogical freedom. However, these MERs are mandatory at an institutional level, as they have been developed, structured and drafted to address the subject-specific content of the PER. Regarding Cycle II (pupils aged 8 to 12), the structure of the MERs follows that of the PER, with thematic areas relating to numbers, operations, geometry and measurement. However, emphasis is placed on cross-curricular elements, such as problem-solving.

This quest for compliance has led the drafters of the MERs to include an entire section devoted to problem solving assistance (*Aide à la résolution de problème* in French, hereafter ARP). Like the other thematic axes, this ARP section is subdivided into four chapters (appropriating a mathematical problem, using research strategies, verifying the answer to a problem, communicating the result of one's research) with the particularity, however, of proposing both specific statements and links to statements from the other thematic axes. The stated issue of this ARP section is to set up a problem-solving teaching:

> Problem solving, which is at the heart of the priority aims of mathematics teaching from cycle 1 to cycle 3 (students aged 4 to 15), is also a source of difficulty for many students. Clearly, it is not enough to let students look for problems and then correct them using the students who have succeeded in doing so, so that everyone learns to solve problems. It is necessary to set up a real teaching of problem solving. That's what the ARP part is all about (author's translation, CIIP, 2023).

Our analysis focuses on the MER resources for Cycle II, which covers pupils aged 7 to 11. This choice is explained by two factors. Firstly, the resources for Cycle III were being rewritten at the time of our study and were therefore unavailable. Secondly, it is in Cycle II that the PER introduces cross-curricular components explicitly formulated under the heading "Elements of problem-solving". For this reason, Cycle I has been excluded from this phase of the research project.

[2]In the Swiss education system, MER means in French *Moyens d'Enseignement Romands*, but for an international audience we translate it into French within text as: *Matériel d'Enseignement et de Ressources*.

We are focusing primarily on the ARP section, as it comprises tasks explicitly designed to develop problem-solving skills. Within Cycle II, the ARP section is divided into four sub-sections: "understanding a mathematical problem", "solving a problem", "checking the answer to a problem" and "communicating the solution to a problem".

However, our corpus is not limited to problems strictly classified under this section. We also include statements from other thematic areas, such as numbers, operations, geometry, quantities and measurements, provided they are explicitly linked to the ARP by the resources themselves.

The corpus thus compiled comprises a total of 312 problems submitted for analysis.

3.2 Data coding tool

The tool we used to analyze the scenarios presented in the MER is the one we outlined in the previous section, based on the criterion : presence and status of a research variable in the problem statement. We have assigned a code to each of the four indicators.

- **Code [3]**: The problem statement contains at least one research variable that is either unfixed or fixed but modifiable, and the choice of values for this variable is left to the pupils.
- **Code [2]**: The problem statement contains at least one fixed but modifiable research variable for which different values are given in the statement or suggested in the provided scenario.
- **Code [1]**: The problem statement contains at least one didactic variable (fixed and modifiable), certain values of which could transform it into a research variable. However, these values are neither specified in the statement nor identified by the resource. It is therefore the teacher's responsibility to identify them.
- **Code [0]**: The problem statement does not contain any instructional variables that could be developed into research variables.

The coding of the situations therefore involved, in accordance with our definition of the research variables and in light of our research question, assigning one of these four codes to each problem statement, whilst examining its wording and the suggested adjustments to the way the activity is managed for the teacher.

3.3 Data analysis tool

We analyzed and coded the entire ARP corpus using the coding scheme outlined above. This corpus comprises 312 statements. The coding was carried out in parallel and in a triple-blind manner by the three authors of this paper. An initial sample of ten statements per level (*i.e.* forty statements across four levels) was used to calibrate the criteria.

Inter-coder agreement was then measured using Fleiss's Kappa coefficient (Fleiss, 1971) to assess reliability, which determines the reliability of the criteria by measuring the degree of agreement in the coding. This was combined with a Z-test to determine whether this agreement is statistically significant.

Furthermore, having four categories available strengthens confidence in the interpretation of the obtained coefficient. Finally, testing on a large dataset of 312 statements further reinforces the reliability (Fleiss, 1971; Landis & Koch, 1977; Sim & Wright, 2005).

Ultimately, to assign a single code to each statement, we opted for a discussion which will be held to reach a consensus if the coders propose different codes containing different triplets such as $(1, 1, 0)$ or $(1, 0, 2)$ for instance. In the first case, this means that two coders assigned the value 1 and a third assigned the value 0.

With regard to the validity of our criteria, we will show that the *a priori* analysis of prototypical problems drawn from each category – limited to variations in the parameters of the task and their influence on pupils' procedures (carried out through coding) – ensures that the problem prompts pupils to engage with the action-formulation-validation dialectic.

The aim is to show qualitatively that our criteria do indeed enable us to identify the potential of the statements to generate mathematical activity, as defined in the theoretical framework section.

4 Analysis and results

4.1 Reliability: are our criteria reliable ? (RQ2)

The inter-rater agreement among the coders during the triple-blind coding process was measured using Fleiss's kappa coefficient noted κ (Fleiss, 1971; Landis & Koch, 1977). This triple-blind coding is presented in the Table 1 where for each triplet coded $(x, y, z) \in \{0, 1, 2, 3\}^3$, the number of occurrences $\#(x, y, z)$ is given. Only the occurrences present in the encoding are listed. The order within the triplet is irrelevant (*i.e.* permutation allowed). For example, the following triplet are considered identical : $(1, 0, 1)$ or $(1, 1, 0)$ or $(0, 1, 1)$, because there are always two values of 1 and one value of 1 assigned by the encoders.

Table 1: Distribution of coded triplets by the three coders by grade – unordered triplets

	School grade				
# Coded triplet (x, y, z)	5H	6H	7H	8H	Total
$\#(0, 0, 0)$	58 (77.3%)	49 (71%)	70 (76.1%)	54 (71%)	231 (73.4%)
$\#(1, 0, 0)$	7 (9.3%)	9 (13%)	5 (5.4%)	4 (5.3%)	25(8.1%)
$\#(1, 1, 0)$	5 (6.7%)	–	7 (7.6%)	6 (7.9%)	18 (5.9%)
$\#(1, 1, 1)$	5 (6.7%)	9 (13%)	6 (6.5%)	11 (14.5%)	31 (10.0%)
$\#(2, 1, 1)$	–	1 (1.4%)	–	–	1 (0.3%)
$\#(2, 2, 1)$	–	–	–	1 (1.3%)	1 (0.3%)
$\#(2, 2, 2)$	–	1(1.4%)	3 (3.3%)	–	4 (1.3%)
$\#(3, 3, 3)$	–	–	1 (1.1%)	–	1 (0.3%)
Total	75	69	92	76	312

The overall result, presented in the Table 2 (last column), indicates a high degree of agreement among the coders : $\kappa \approx 0.688 \in [0.61, 0.80]$ ($\overline{p} \approx 0,905$; $\overline{p_e} \approx 0,698$), according to Landi and Kosh's table (Table 3). Furthermore, the results of the associated Z-test are highly significant ($z \approx 23.09$, p-value < 0.001), indicating that this agreement is significantly greater than would be expected by chance.

Consequently, our criteria for the various codes are considered reliable. It should be noted that no coding discrepancies were identified among the three coders and that, in each instance where a difference was observed, there was only a single level of deviation (see Table 1). In these conditions, it was not deemed necessary to use another reliability measure, such as Krippendorff's alpha (Krippendorff, 2011).

Table 2: Fleiss's Kappa (κ) and Z-test scores (z), by grade, as well as their overall score

	School grade				
	5H	6H	7H	8H	5H-8H
$\overline{p}$	0.89333333	0.90338164	0.91304348	0.91228076	0.90598291
$\overline{p_e}$	0.75600988	0.66944386	0.69559966	0.63923515	0.68805017
κ	0.56282383	0.70770969	0.714335	0.75685187	0.6986147
z	8.371	10.986	14.439	11.325	23.090
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Table 3: Interpretation of Kappa stastistic coefficient (Landis & Koch, 1977)

Kappa statistic (κ)	Strength of agreement
< 0	Poor
$0.00 - 0.20$	Slight agreement
$0.21 - 0.40$	Fair agreement
$0.41 - 0.60$	Moderate agreement
$0.61 - 0.80$	Substantial agreement
$0.81 - 1.00$	Almost perfect agreement

The table below (Table 4) presents the final coding results obtained for each code according to primary school year group. The majority of statements fall under codes [0] (76%) and [1] (22.4%), while those coded as [2] and [3] are in the minority (1.3% and 0.3% respectively). This trend is also confirmed at each level of education.

Table 4: Final distribution of codes for every grade after agreement

School grade	Code [X]				Total
	Code [0]	Code [1]	Code [2]	Code [3]	
5H (8–9 y.d)	62	13	0	0	75
6H (9–10 y.d)	50	18	1	0	69
7H (10–11 y.d)	70	18	3	1	92
8H (11–12 y.d)	55	21	0	0	76
Total	237	70	4	1	312
Percentage	76%	22.4%	1.3%	0.3%	100%

4.2 Validity : is our tool valid ? (RQ1)

To ensure the validity of our tool, we carried out a preliminary analysis of four prototypical problems drawn from each of the coded categories. This analysis is limited to variations in the parameters of the task and their influence on pupils' procedures (as identified through the coding process). We thus highlight the problem's potential to generate work on the action-formulation-validation dialectic. In short, we show that the presence of a single research variable is enough to prompt work on the action-formulation-validation process. In light of our conceptual framework, the validity process is presented as follows.

- **Step 1** – Identify all the didactic variables in the statement, and among these, those that can be characterized as research variables (see Figure 1).
- **Step 2** – Justify the coding carried out.
- **Step 3** – Examine the potential procedures.
- **Step 4** – Determine whether the exploration of the identified research variables involves a didactic process of experimentation, formulation and validation.

These different steps are presented in the Fig. 2 below.

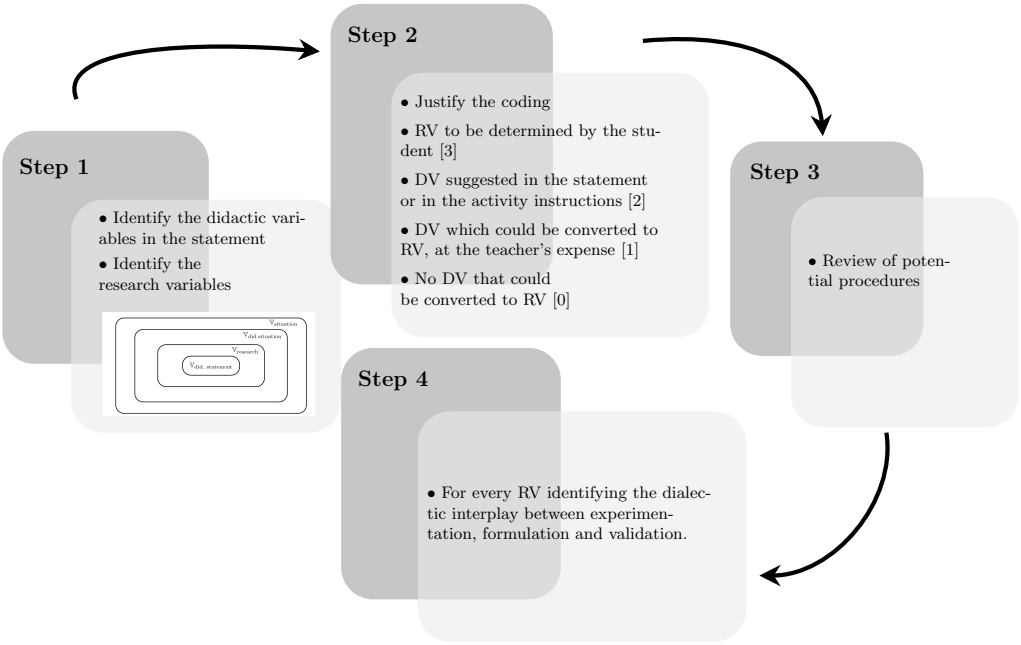


Fig. 2: Different steps about validity process

4.2.1 Analysis of a prototypical problem in category [0]

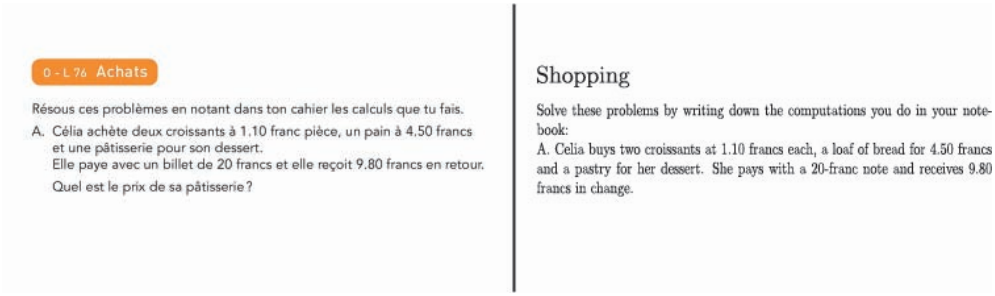


Fig. 3: Problem statement 'Shopping' from MER 8H and its English translation on the right

Step 1 of the problem in category [0]

In this case (Fig. 3), we can identify two didactic variables: the value of the numbers and the number of numbers given. The values of these two variables are fixed in the statement, and their variation is not suitable for a particular-general dialectic.

Step 2 of the problem in category [0]

Thus, in this statement, there is no research variable. The code assigned is therefore [0].

Step 3 of the problem in category [0]

A more detailed initial analysis highlights three possible approaches to this mathematical problem.

- The first approach consists of determining the price of the pastry through a process of trial and error. One sets a price, for example CHF 3, computes the total cost, and compares it to the difference between CHF 20 and CHF 9.80. If the two amounts do not match, the price is adjusted (for example to CHF 4) and the process is repeated until the exact amount is found.
- The second type of procedure consists of directly calculating the expression using different notations, based on the idea of determining the remainder when subtracting CHF 10.20 from the total expenditure. For example: $(20-9.80)-(2\times 1.10+4.50) =$ CHF 3.50. This calculation can also be carried out step by step: *a*) $1.10 \times 2 = 2.20$ then *b*) $4.50 + 2.20 = 6.70$ then *c*) $20 - 9.80 = 10.20$ and finally *d*) $10.20 - 6.70 = 3.50$. Alternatively: *a*) $20 - 9.80 = 10.20$ then *b*) $10.20 - 4.50 = 5.70$ and thus *c*) $5.70 - 2.20 = 3.50$.
- Finally, a third algebraic method is possible by letting x represent the price of the pastry and solving the following equation: $2 \times 1.10 + 4.50 + x = 20 - 9.80$.

Step 4 of the problem in category [0]

Within these procedures, the knowledge drawn upon relates to the choice of operations (to accurately reflect the real-life situation), the structure and syntax of the written expression (use of brackets, order of operations), and the performance of calculations (addition, subtraction and multiplication). The choice of integers and the small number of numbers involved may lend themselves to a trial-and-error approach, which involves a form of experimentation but lacks the dialectical process of formulating and validating a conjecture. This confirms the limited potential of this task to engage pupils in a specific 'particular–general' dialectical process.

4.2.2 Analysis of a prototypical problem in category [1]

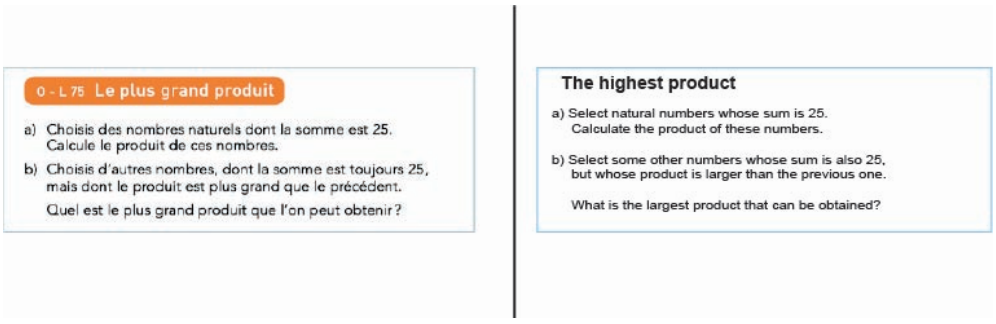


Fig. 4: Problem statement "The highest product" from MER 7H and its English translation on the right

Step 1 of the problem in category [1]

In this statement (Fig. 4), we can identify a key variable: the value of the target number. If this value is small (for example, less than 10), it is possible to find all the additive decompositions, calculate the corresponding products and compare them. If this value is higher (for example, 213), the analysis of specific cases and the identification of invariants (e.g. it is not appropriate to have more than three 2; it is not appropriate to have any 4; there are many 3, *etc.*) are prioritized in the search for a solution. In this way, a particular-general dialectic can be explored: this didactic variable is therefore a research variable.

Step 2 of the problem in category [1]

The code [1] has been assigned to this statement because the value of the variable is set to 25, and no other value for this variable is proposed or suggested in the statement or in the activity instructions.

Step 3 of the problem in category [1]

A more detailed preliminary analysis identifies three two ? possible approaches to this problem (assuming a target number of 25).

- One method involves selecting additive decompositions of 25, calculating the corresponding products – for example, $25 = 11 + 11 + 3$ ($11 \times 11 \times 3 = 363$) and $25 = 3 + 4 + 5 + 10 + 3$ ($3 \times 4 \times 5 \times 10 \times 3 = 1800$) – and comparing them. This computation relies on numerical calculations and comparisons of numbers.
- A second procedure involves selecting additive decompositions of 25, calculating their corresponding products, comparing them, and considering what makes one product larger than the other. This leads to conjectures, for example, that there must be no 0s, that 4 and $2+2$ yield the same product, that $3+3$ is more interesting than 6 (since $3 \times 3 = 9$), and so on.

Step 4 of the problem in category [1]

This analysis shows that choosing 25 is appropriate for favoring the second procedure. Indeed, the first procedure has its limitations, as the number 25 is large enough to prevent the search for an exhaustive list of additive decompositions of 25. The second approach allows for an exploration of the general-particular dialectic in the search for the reasons behind the maximum product, involving phases of experimentation (producing, studying and comparing different products), formulation (conjecturing that $2+2$ equals 4, and that the maximum is 3) and validation ($2+2$ equals 4 because $2 \times 2 = 4$). However, as this analysis is not included in the resource, it is unlikely that pupils will explore this dialectic in class.

4.2.3 Analysis of a prototypical problem in category [2]

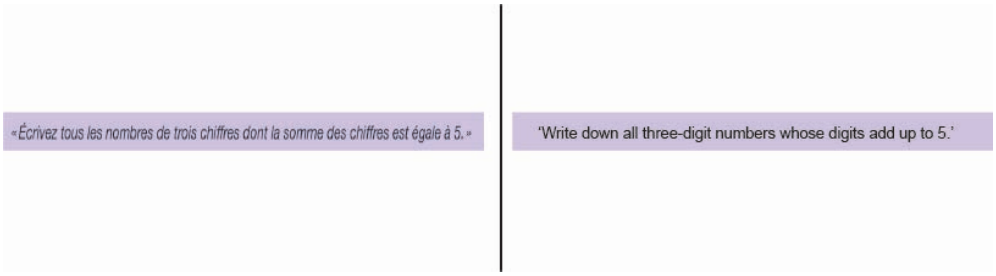


Fig. 5: The problem statement "All numbers" taken from MER 6H and its English translation on the right

Step 1 of the problem in category [2]

In this statement (Fig. 5), we can identify three didactical variables. The first is the number of digits, and this value influences the number of possible combinations. Thus, a small value may make it possible to find all the combinations by trial and error. A larger number will require an algorithm to identify and count all the combinations. The second teaching variable is the sum of the numerical values: choosing a relatively small integer allows for the generation of several combinations by trial and error, whereas a larger value requires the identification of a pattern (for example, if the target number is 9, there will be 2 combinations starting with 8, 3 combinations starting with 7, and so on). Finally, the way the question is phrased is a third pedagogical variable, as it influences the nature of the task. Asking pupils to 'write down all the numbers' encourages them to carry out an exhaustive and systematic search for all possible cases. The phrase 'how many are there?' directs the expected approach towards counting, without requiring pupils to produce a complete list. Among these didactic variables, we have identified two research variables:

1. the number of digits. This is a fixed but modifiable value, and its variation is not specified in the problem statement or in the instructions given to the teacher ;
2. the value of the target number. This is a fixed but modifiable value, and its variation is suggested by the comments accompanying the activity.

Step 2 of the problem in category [2]

As there is at least one research variable suggested by the comments on the activity, we have coded this statement with the code [2].

Step 3 of the problem in category [2]

A more detailed initial analysis highlights three possible approaches to this problem.

- The first approach is an empirical method involving going through all the numbers from 100 to 999 and testing the sum of their digits. Although this method is not very efficient, it may serve as an initial intuitive attempt.

- The second approach involves systematic testing. It involves determining the hundred digit, then distributing the remaining sum between the other two digits. For example, if the hundred digit is 1, there are 4 left to distribute between the tens and units, resulting in the possibilities $(4, 0)$, $(3, 1)$, $(2, 2)$, $(1, 3)$, $(0, 4)$. Next, either all the possibilities are listed by examining every possible value for the hundreds digit, or a pattern is identified during this process: when the hundreds digit is 1 (or 2), there are 5 possibilities (or 4), and we conclude that the total number of possibilities is equal to $1 + \ldots + 5 = 15$.
- A third method involves listing the triplets whose sum is 5, then counting the possible combinations involving these three digits: five triplets are identified – $(5, 0, 0)(4, 0, 1)(3, 2, 0)(3, 1, 1)(2, 2, 1)$ – then, for each triplet, there are 6 possible combinations, with the exception of triplets containing a 0 (as the hundreds digit cannot be 0) and triplets containing two identical numbers. We therefore obtain $56–(5 + 2 + 2 + 3 + 3) = 15$

Step 4 of the problem in category [2]

Analysis of the procedures shows that the statement (with the variable values set at 3 and 5) triggers elements characteristic of the action–formulation–validation dialectic.

- Action – carrying out tests (whether organized or not), breaking down the number 5 into three parts, organizing the search according to the hundreds digit;
- Formulation – devising a hypothesis regarding the number of possible occurrences for each value of the hundreds digit (for example: 5 occurrences if the hundreds digit is 1, 4 if it is 2, and so on);
- Validation – verifying that the resulting list is exhaustive, using a completeness argument, and justifying the plausibility of the conjecture.

Thus, this statement can serve as a starting point for generalizations and statements such as 'the sum of the digits equals n for numbers with k digits'. It should also be noted that solving this problem requires a firm grasp of place value, and in particular the distinction between hundreds, tens and units, as well as an understanding of and ability to apply the principles of organized listing, thereby avoiding duplication and omissions (*i.e.* a certain ability to structure the search in a systematic manner).

4.2.4 Analysis of a prototypical problem in category [3]

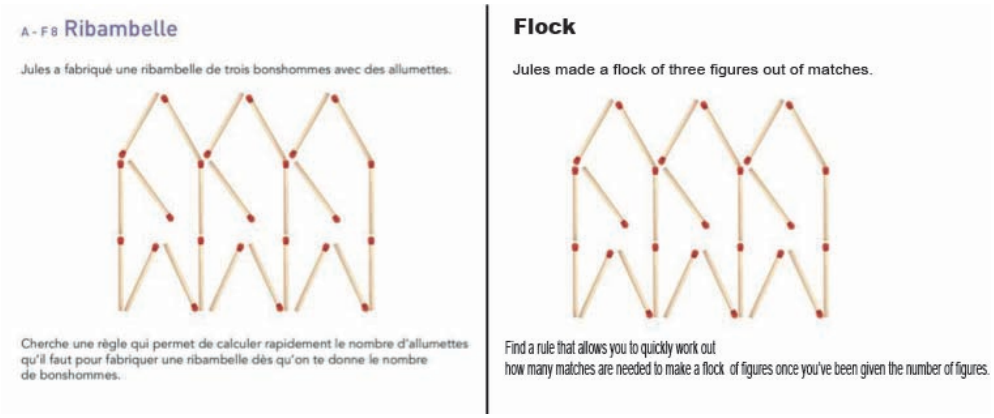


Fig. 6: Problem statement "Flock" from MER 7H and its English translation on the right

Step 1 of the problem in category [3]

In this problem (Fig. 6), we can identify two variables. The first is the number of shapes (stick figures, as specified in the problem). A small value allows the matches to be counted, while a large value, or an unspecified value, requires the discovery and identification of a pattern (for example, adding seven matches to move from one pattern to the next). The second teaching variable is the content of the illustration that accompanies the statement. If the illustration shows a sequence of several steps, such as 1, 2, and 3, a pattern of recurrence can be identified by comparing them, *i.e.* by adding 7 to the previous figure. If the illustration shows only one step, as in this example, the functional aspect can be emphasized (*i.e.* $f(n) = 7n + 2$, where n represents the number of steps).

Step 2 of the problem in category [3]

The question does not suggest a specific number of patterns, so this variable can be chosen at random by the student. It is therefore a research variable, not a fixed one. This problem is coded as 3.

Step 3 of the problem in category [3]

A more detailed preliminary analysis of the possible approaches reveals three types, bearing in mind that a counting method based on diagrams is not feasible, as the question asks us to find a rule for the general case.

- The first procedure involves comparing a pattern with its successor, which results in the identification of a recurrence relation of the form $a_n = a_{n-1} + 7$. The general term of this relation is $a_n = 7(n-1) + 9 = 7n + 2$.
- The second method involves listing a series of consecutive numbers, such as 9, 16, 23, and so on, and identifying the arithmetic sequence with a common difference

of 7 resulting from this. Given the initial conditions, this sequence yields $7n + 2$. Examining specific examples, with or without illustrations, can make identifying these two sequences easier.

- The third method differs from the first two in that it examines the overall structure of the figure. For example, each sub-figure consists of seven matches, and two additional matches are added to "close off the final pattern". Therefore, there are $7n + 2$ matches for n figures.

Step 4 of the problem in category [3]

This analysis reveals a dialectical process involving experimentation with specific cases. This leads to the identification of patterns, such as the ratio of an arithmetic sequence or the structure of a shape, and the formulation of a conjecture regarding the number of matches required for n figures. Validation focuses on establishing a correspondence between two domains: the algebraic rule and the figurative pattern structure.

In summary, a preliminary analysis of the four prototypical problems in our coding scheme indicates that the presence of at least one research variable establishes an interplay between experimentation, formulation, and validation processes. This interplay ensures the application of a particular-general dialectic, enabling us to validate our tool for identifying a problem's potential to stimulate mathematical research activity among pupils.

5 Conclusion and outlook

This study has enabled us, on the one hand, to establish the validity and reability of the tool developed and, on the other hand, to show that the research variable is sufficient to account for the potential of the statements to engage with the experimentation–formulation–validation dialectic (and thus a problem-solving activity akin to that of a researcher, as advocated by the curricula). In this respect, the *a priori* analyses conducted confirm the existence of a link between the coding assigned to the statements and their presumed capacity to support work falling within this dialectic.

However, the tool's external validity is yet to be establishedand will require the conduct of *post-hoc* analyses based on the actual observation of pupils' activities in real-life situations. In this regard, it would be appropriate to consider two or three in-depth case studies for code [0], given the diversity of the problems grouped within this category.

Furthermore, the reliability of the tool was established using Fleiss's Kappa coefficient (Fleiss, 1971), indicating substantial inter-rater agreement ($\kappa \approx 0.69$) which was statistically significant ($z \approx 23.09$, p-value < 0.001). This method, which was developed to be applicable to large corpora, stands out for its practicality. This is particularly apparent when it is compared with other existing frameworks, such as the one proposed by Georget (2009). It has also provided relevant answers to the initial research questions.

Furthermore, the design of this tool also contributes to the theoretical development of TDS (Theory of Didactical Situations) (Brousseau, 1997) by refining the notions of didactic variables and research variables. We thus provide reliable and valid operational criteria for analyzing a large corpus, thereby offering a new approach within the field of TDS.

Beyond the initial objectives, the results highlight a mismatch between, on the one hand, the resources made available to teachers and, on the other, the institutional expectations set out in the curriculum guidelines. This tension is particularly evident in the low number of items with a high level of complexity, even though these are intended to encourage mathematical activity that fosters an inquiry-based approach. These findings have several implications.

Firstly, while the developed tool is part of a research project, the question arises as to how it could be adapted for training purposes to help teachers identify and select statements that could underpin exploratory mathematical activities.

Secondly, the discrepancy between the PER's expectations and the proposals put forward by Romandie Teaching Resources suggests that new teaching scenarios should be designed to address this shortfall.

The third issue to consider is the progressive nature of the selection of problems for learning the general-particular dialectic. This is in line with recent research that was initiated as part of a LéA project (2021–2024) on proof from nursery school to higher education (Gandit, 2024; Gandit, Mossuz, & Gravier, 2022).

Finally, a fourth perspective involves extending this work to other corpora for comparison and generalization on an international scale.

Declaration

Availability of data and materials. The data that support the findings of this study are available from the corresponding author upon request.

Competing interests. The authors declare no competing interests.

References

Artigue, M. (2015). Perspectives on design research: The case of didactical engineering. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: Examples of methodology and methods* (pp. 467–496). Dordrecht: Springer Netherlands. Retrieved from https://doi.org/10.1007/978-94-017-9181-6_17

Artigue, M. (2020). Didactic engineering in mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 202–206). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-15789-0_44"

Artigue, M., & Houdement, C. (2007). Problem solving in france: didactic and curricular perspectives. *ZDM–Mathematics Education*, *39*(5), 365–382, https://

doi.org/https://doi.org/10.1007/s11858-007-0048-x

Barallobres, G. (2007). Introduction à l'algèbre par la généralisation : problèmes didactiques soulevés. *For the Learning of Mathematics*, *27*(1), 39–44, Retrieved from http://www.jstor.org/stable/40248560

Barallobres, G., & Giroux, J. (2008). Carences et régulations des milieux en situation de validation. N. Bednarz & C. Mary (Eds.), *L'enseignement des mathématiques face aux défis de l'école et des communautés.* Éditions du CRP. (Actes du colloque EMF 2006, cédérom)

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education*, *29*(1), 41–62, https://doi.org/10.2307/749717

Borasi, R. (1986). On the nature of problems. *Educational Studies in Mathematics*, *17*(2), 125–141, https://doi.org/https://doi.org/10.1007/BF00311517

Brousseau, G. (1997). *Theory of didactical situations in mathematics: Didactique des mathématiques, 1970–1990* (N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield, Eds. & Trans.). Dordrecht: Kluwer Academic Publishers. Retrieved from https://doi.org/10.1007/0-306-47211-2

Burkhardt, H., & Bell, A. (2007). Problem solving in the united kingdom. *ZDM–Mathematics Education*, *39*(5-6), 395–403, https://doi.org/10.1007/s11858-007-0026-z

Cai, J., & Nie, B. (2007). Problem solving in chinese mathematics education: Research and practice. *ZDM – Mathematics Education*, *39*(5), 459–473, https://doi.org/https://doi.org/10.1007/s11858-007-0042-3

Chevallard, Y. (1998). Analyse des pratiques enseignantes et didactique des mathématiques: l'approche anthropologique. R. Noirfalise (Ed.), *Analyse des pratiques enseignantes et didactique des mathématiques – Actes de l'Université d'Eté de la Rochelle – Charente-maritime* (pp. 91–118). Retrieved from https://www.univ-irem.fr/IMG/pdf/analyse_des_pratiques_univ_d_ete_la_rochelle.pdf

Chevallard, Y., et al. (Eds.). (2022). *Advances in the anthropological theory of the didactic* (Vol. 2). Springer. Retrieved from https://doi.org/10.1007/978-3-030-76791-4

Conférence intercantonale de l'instruction publique de la Suisse romande et du Tessin (CIIP) (2023). *Domaines disciplinaires.* Retrieved 2026-04-06, from https://portail.ciip.ch/per/domains/2

Da Ronch, M. (2022). *Pratique de l'activité mathématique en médiation: modèles didactiques et conception d'ingénieries* (Doctoral dissertation, Université Grenoble Alpes). Retrieved from https://cnrs.hal.science/tel-04089443/

Da Ronch, M., Gandit, M., Gravier, S. (2020). Du problème de wang vers une nouvelle situation de recherche pour la classe. *Repères Irem*, *121*, 77–105, Retrieved from https://hal.science/hal-03030861/

Da Ronch, M., Gardes, M.-L., Mili, I. (2023). Study of the potential of problems to practice a research activity in mathematics at elementary school in french-speaking switzerland. P. Drijvers, C. Csapodi, H. Palmér, K. Gosztonyi, & E. Kónya (Eds.), *Proceedings of the thirteenth congress of the erme (CERME13)* (pp. 96–103). Alfréd Rényi Institute of Mathematics and ERME. Retrieved from https://hal.science/hal-04408292v1

Da Ronch, M., & Hocquard, H. (2026). De l'expérimentation à la preuve en mathématiques : étude de cas d'une situation issue de la recherche en optimisation combinatoire. *Repères Irem*, *143*, 5–36, Retrieved from https://hal.science/hal-05434743 (In press)

Deloustal-Jorrand, V., Gandit, M., Mesnil, Z. (2025). Spécificités des connaissances en logique et conséquence sur la double discontinuité de klein. *Recherches en didactique des mathématiques*, *45*(3), 239–269, https://doi.org/10.46298/rdm.14819

DREAM (2020, July 30). *La somme de 10 entiers consécutifs.* https://math.univ-lyon1.fr/dream/wp-content/uploads/2022/03/La-somme-des-10-entiers-consecutifs.pdf.

Durand-Guerrier, V. (2010). La dimension expérimentale en mathématiques. Enjeux épistémologiques et didactiques. G. Aldon et al. (Eds.), *Expérimenter des problèmes innovants en mathématiques à l'école.* INRP. Retrieved from http://educmath.ens-lyon.fr/applet/exprime/texteintroress.pdf

Durand-Guerrier, V. (2025). Divers usages du concept de variables didactiques. C. Derouet, V. Durand-Guerrier, C. Lemrich, & A.-C. Mathé (Eds.), *Pré-actes du colloque international en hommage à l'œuvre de Guy brousseau* (pp. 104–109). IREM d'Aquitaine. Retrieved from https://hal.science/hal-05184913/

English, L.D., & Gainsburg, J. (2016). Problem solving in a 21st-century mathematics curriculum. L.D. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (3rd ed., pp. 313–335). New York: Routledge. Retrieved from https://eprints.qut.edu.au/87756/24/87756.pdf

Fleiss, J.L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382, https://doi.org/10.1037/h0031619

Front, M. (2015). *Émergence et évolution des objets mathématiques en Situation Didactique de Recherche de Problème : le cas des pavages archimédiens du plan* (Doctoral dissertation, Université Claude Bernard - Lyon I). Retrieved from https://theses.hal.science/tel-01312266

Gandit, M. (2024). Chercher–débattre–prouver à partir du cycle 1 : des résultats issus d'un léa. *Actes de la 22e école d'été de didactique des mathématiques.* Retrieved from https://hal.science/hal-04819108/

Gandit, M., Mossuz, L., Gravier, S. (2022). L'enseignement et l'apprentissage de la preuve en mathematiques du cycle 1 au cycle 3: Premiers outils et premiers resultats. *Représenter et modéliser en mathématiques. de l'activité des élèves à la formation des professeurs des écoles. 48ème colloque de la COPIRELEM. Toulouse 2022* (pp. 431–448). Retrieved from https://hal.science/hal-04818426v1

Gardes, M.-L. (2013). *Étude de processus de recherche de chercheurs, élèves et étudiants, engagés dans la recherche d'un problème non résolu en théorie des nombres* (Doctoral dissertation, Université Claude Bernard-Lyon I). Retrieved from https://theses.hal.science/tel-00948332/

Gardes, M.-L. (2018). Démarches d'investigation et recherche de problèmes. C. Editions (Ed.), *Le Rallye mathématique, un jeu très sérieux !* (p. 73-96). Retrieved from https://hal.science/hal-03533121

Garey, M.R., & Johnson, D.S. (1979). *Computers and intractability* (Vol. 29). wh freeman New York.

Georget, J.-P. (2009). *Activités de recherche et de preuve entre pairs à l'école élémentaire: perspectives ouvertes par les communautés de pratique d'enseignants* (Doctoral dissertation, Université Paris-Diderot-Paris VII). Retrieved from https://shs.hal.science/tel-00426603/

Godot, K. (2005). *Situations recherche et jeux mathematiques pour la formation et la vulgarisation. Exemple de la roue aux couleurs.* (Doctoral dissertation, Université Joseph-Fourier - Grenoble I). Retrieved from https://theses.hal.science/tel-00102171

Gravier, S., & Ouvrier-Buffet, C. (2022). The mathematical background of proving processes in discrete optimization—exemplification with research situations for the classroom. *ZDM–Mathematics Education*, *54*(4), 925–940, https://doi.org/10.1007/s11858-022-01400-3

Gravier, S., & Trouvé, T. (2025). Les variables de recherche: un outil pour transposer la dialectique particulier-général. C. Derouet, V. Durand-Guerrier, C. Lemrich, & A.-C. Mathé (Eds.), *Pré-actes du colloque international en hommage à l'œuvre de Guy Brousseau* (pp. 115–119). IREM d'Aquitaine. Retrieved from https://hal.science/hal-05325431v1

Grenier, D., & Payan, C. (1998). Spécificités de la preuve et de la modélisation en mathématiques discrètes. *Recherches en didactique des mathématiques*, *18*(1), 59–100,

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. Princeton University Press. Retrieved from https://worrydream.com/refs/Hadamard_1945_-_The_psychology_of_invention_in_the_mathematical_field.pdf

Halmos, P.R. (1980). The heart of mathematics. *The american mathematical monthly*, *87*(7), 519–524, https://doi.org/10.1080/00029890.1980.11995081

Jäder, J., Lithner, J., Sidenvall, J. (2020). Mathematical problem solving in textbooks from twelve countries. *International Journal of Mathematical Education in Science and Technology*, *51*(7), 1120–1136, https://doi.org/10.1080/0020739X.2019.1656826

Krippendorff, K. (2011). *Computing krippendorff's alpha-reliability* (Tech. Rep.). University of Pennsylvania. Retrieved from http://repository.upenn.edu/asc_papers/43

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174, https://doi.org/10.2307/2529310

Lesh, R., & Zawojewski, J.S. (2007). Problem solving and modeling. F.K. Lester (Ed.), *The second handbook of research on mathematics teaching and learning* (pp. 763–804). Charlotte, NC: Information Age Publishing.

Lester, F.K., & Kehle, P.E. (2003). From problem solving to modeling: The evolution of thinking about research on complex mathematical activity. *Beyond constructivism* (pp. 501–517). Routledge.

Liljedahl, P., & Santos-Trigo, M. (Eds.). (2019). *Mathematical problem solving*. Springer. Retrieved from https://doi.org/10.1007/978-3-030-10472-6

Liljedahl, P., Santos-Trigo, M., Malaspina, U., Bruder, R. (2016). Problem solving in mathematics education. In *Problem solving in mathematics education* (pp. 1–39). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-40730-2

Loisy, C., Sabra, H., Courtney, S.A., Rocha, K., Glasnović Gracin, D., Aldon, G., ... Robutti, O. (2019). Analyzing teachers' work with resources: Methodological issues. In L. Trouche, G. Gueudet, & B. Pepin (Eds.), *The 'resource' approach to mathematics education* (pp. 257–321). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-20393-1_10

Mili, I., Da Ronch, M., Gardes, M.-L. (2025). Combining a theoretical framework and a statistical measure to assess the reliability of didactic criteria in the analysis of large corpora. M. Bosch, G. Bolondi, S. Carreira, C. Spagnolo, & M. Gaidoschik (Eds.), *Proceedings of the fourteenth congress of the european society for research in mathematics education (CERME14)* (pp. 2832–2839). Free University of Bozen-Bolzano and ERME. Retrieved from https://hal.science/hal-05283880v1

National Council of Teachers of Mathematics (NCTM) (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

OECD (2023). *Pisa 2022 results (volume i): The state of learning and equity in education.* Retrieved from https://doi.org/10.1787/53f23881-en

Ouvrier-Buffet, C. (2011). A mathematical experience involving defining processes: in-action definitions and zero-definitions. *Educational Studies in Mathematics*, *76*(2), 165–182, https://doi.org/10.1007/s10649-010-9272-3

Ouvrier-Buffet, C. (2020). Discrete mathematics teaching and learning. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 227–233). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-030-15789-0_51

Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton university press.

Rott, B. (2014). Rethinking heuristics – characterizations and examples. A. Ambrus & E. Vasarhelyi (Eds.), *Problem solving in mathematics education – proceedings of the 15th promath conference* (pp. 176–192).

Rott, B., Specht, B., Knipping, C. (2021). A descriptive phase model of problem-solving processes. *ZDM–Mathematics Education*, *53*(4), 737–752, https://

doi.org/10.1007/s11858-021-01244-3

Santos-Trigo, M. (2007). Mathematical problem solving: An evolving research practice domain. *ZDM–Mathematics Education*, *39*(5-6), 523–536, https://doi.org/10.1007/s11858-007-0030-3

Santos-Trigo, M. (2020). Problem-solving in mathematics education. S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 686–693). Cham: Springer. Retrieved from https://doi.org/10.1007/978-3-030-15789-0_129

Schoenfeld, A.H. (1985). *Mathematical problem solving.* Elsevier.

Schoenfeld, A.H. (1992a). Learning to think mathematically: Problem solving, metacognition, and sense-making. *Handbook of Research on Mathematics Teaching and Learning*, , https://doi.org/10.1177/002205741619600202

Schoenfeld, A.H. (1992b). On paradigms and methods: What do you do when the ones you know don't do what you want them to? issues in the analysis of data in the form of videotapes. *The Journal of the Learning Sciences*, *2*(2), 179–214, https://doi.org/10.1207/s15327809jls0202_3

Schoenfeld, A.H. (2007). Problem solving in the united states, 1970–2008: Research and theory, practice and politics. *ZDM –Mathematics Education*, *39*(5-6), 537–551, https://doi.org/10.1007/s11858-007-0038-z

Schoenfeld, A.H. (2015). How we think: A theory of human decision-making, with a focus on teaching. S.J. Cho (Ed.), *The proceedings of the 12th international congress on mathematical education* (pp. 229–243). Cham: Springer. Retrieved from https://doi.org/10.1007/978-3-319-12688-3

Selden, J., Mason, A., Selden, A. (1989). Can average calculus students solve nonroutine problems. *Journal of Mathematical Behavior*, *8*(1), 45–50, Retrieved from https://www.researchgate.net/publication/256445046_The_Journal_of_Mathematical_Behavior_article_Can_average_calculus_students_solve_nonroutine_problems

Sim, J., & Wright, C.C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, *85*(3), 257–268, https://doi.org/10.1093/ptj/85.3.257

Stacey, K. (2005). The place of problem solving in contemporary mathematics curriculum documents. *The Journal of Mathematical Behavior*, *24*(3), 341-350, https://doi.org/10.1016/j.jmathb.2005.09.004

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESMDaRonchetal2026.zip