Educational Studies in Mathematics (2023) 113:207–227
https://doi.org/10.1007/s10649-023-10220-5

Pre-service teachers' flexibility and performance in solving Fermi problems

Carlos Segura[1] · **Irene Ferrando**[1]

Accepted: 19 February 2023 / Published online: 23 March 2023
© The Author(s) 2023

Abstract
Fermi problems are real-context estimation tasks that are suitable for introducing open-ended problems in primary school education. To ensure their effective introduction in the classroom, teachers must have adequate proficiency to deal with them. One of the key aspects of problem-solving proficiency is flexibility, but there are few studies on flexibility in solving real-context problems. This study, based on an analysis of the errors made by 224 prospective teachers when solving a Fermi problem sequence, establishes performance levels. In addition, we define levels of flexibility in using multiple solutions across the sequence, which allows us to address the main objective: to study the relationship between performance and flexibility. We found that there are significant relationships between flexibility levels and the number and severity of errors made. Encouraging flexibility in prospective teachers may be an efficient way to improve their performance in solving real-context problems.

Keywords Estimation · Real-context · Flexibility · Problem-solving · Errors · Pre-service primary teachers · Fermi problems

The flexible use of multiple strategies is an important aspect that could enable differentiating levels of problem-solving performance. However, studies addressing the relationship between flexibility and performance in the framework of real-context problems are scarce and qualitative in nature (Govender, 2020; Schukajlow et al., 2015). Fermi problems are real-context problems that require solvers to make assumptions about a situation and to estimate relevant quantities before engaging in often simple calculations (Ärlebäck, 2009). These tasks are particularly suitable to introduce real-context open-ended problem-solving in primary school education (Albarracín & Gorgorió, 2019) because they require relatively simple arithmetic procedures and they are therefore accessible. Ferrando et al. (2020) found that, when solving Fermi problems, pre-service teachers provide different solution strategies to similar problems depending on features of the problems. This is what Elia et al.

✉ Irene Ferrando
irene.ferrando@uv.es

Carlos Segura
carlos.segura@uv.es

[1] Universitat de València, Valencia, Spain

(2009) call flexibility in problem-solving. Heinze et al. (2009) highlighted that the flexible use of strategies enables individuals to solve problems quickly and accurately; indeed, it is a characteristic of teachers' problem-solving proficiency (Chapman, 2015). Actually, Schoenfeld (1982) stressed that successful performance in problem-solving depends on selecting suitable strategies and discarding inappropriate approaches. The present study explores the relationship between pre-service teachers' performance and flexibility when solving Fermi problems. This is an important issue that can contribute to providing insight into real-context problem-solving proficiency.

1 Theoretical framework

In order to study the relationship between flexibility and performance in real-context problem-solving, we use Fermi problems. We begin with a description of the characteristics of these open-ended problems and we outline the previous studies that lead us to identify and classify the multiple strategies used to solve them. Characterising Fermi problems as multiple solution tasks allows us to address flexibility in problem-solving, which we explain in the second subsection. Finally, in the third subsection, we address the notion of performance in problem-solving and, in particular, we focus on its analysis for Fermi problems through the study of errors. The theoretical framework will allow us to pose the research questions on the relationship between flexibility and performance for Fermi problems.

1.1 Fermi problems

Fermi problems are non-numerical problems, in which the only information provided is the element whose number we want to estimate and the real context in which these elements are located. In order to solve Fermi problems, it is necessary to make an argument that explains an estimation based on a real situation (Ärlebäck, 2009). According to Sriraman and Knott (2009), Fermi problems foster solvers to make educated guesses. The physicist Enrico Fermi gave a classic example of this type of problem. He asked his students, "How many piano tuners are there in the city of Chicago?" (Efthimiou & Llewellyn, 2007). To answer this question, it is first necessary to clearly identify the relevant variables of the problem (inhabitants of the city of Chicago, proportion of families who own a piano, etc.) based on an interpretation of the real context in which it is formulated. Next, a solver should establish a mathematical strategy to obtain a numerical solution. Finally, the solver should interpret and validate that solution in order to, if needed, tackle a more complex resolution to obtain a more accurate estimate.

Fermi problems are open-ended problems, enabling different approaches that can lead to different solutions. Achmetli et al. (2019) identified three ways to differentiate the solutions of a real-context problem: the first is to establish different hypotheses that usually lead to different outcomes; the second is to apply different mathematical strategies, which usually led to the same mathematical outcome; and the third is the combination of the previous two. Relying on the third perspective, Albarracín et al. (2021) studied the solution of activities such as those used in the present work, obtaining the solution spaces (Leikin & Levav-Waynberg, 2008) that allow classifying different solution strategies. The solution strategy categorisation presented by Albarracín et al. (2021), based on the productions of secondary school students, was extended for pre-service teachers (Ferrando et al., 2021). A categorisation of all possible solution strategies of these Fermi problems allows us to

consider them as multiple solution tasks (Levav-Waynberg & Leikin, 2012), which makes it possible to monitor which solutions are suitable for a particular task and to measure whether the problem solvers know and use more than one solution strategy when they face a sequence of Fermi problems.

Our research employs sequences of Fermi problems, following Ärlebäck and Doerr's (2015) idea of using them to facilitate the development of problem-solving proficiency. In order to study flexibility, the design of the sequence promotes changes in the solution strategies across four problems. The activities included in the sequence of Fermi problems require an argued estimate of the number of elements in a bounded enclosure. To design the sequence, and following the theory of variation (Ko & Marton, 2004), we rely on the contrast between relevant contextual features that had been studied in previous works. In prior work, Ferrando et al. (2020) found that some characteristics of the context in which the problem is formulated (namely the size of the enclosure, the size of the elements, and their distribution in the enclosure) influence the solution strategies proposed by prospective teachers.

1.2 Flexibility in problem-solving

Researchers on problem-solving strategies use the term "flexible" with different meanings (Heinze et al., 2009, p. 536). Taking the broadest definition, flexible strategy use refers to individuals being able to choose between different solution strategies when dealing with a mathematical activity. Flexibility is an important mathematical skill; indeed, it is necessary for students to acquire the ability to adapt their solution strategies to the characteristics of the task or context (Heinze et al., 2009). In the problem-solving framework, studies on the flexible use of multiple solution strategies found it essential for building deep and connected knowledge (Levav-Waynberg & Leikin, 2012; Star & Rittle-Johnson, 2008). Most studies on the influence of the development of multiple solution strategies and their flexible use have focused on intra-mathematical tasks (Levav-Waynberg & Leikin, 2012; Star & Rittle-Johnson, 2008; Threlfall, 2002). Elia et al. (2009) developed a study about flexibility with primary school students who solved a sequence of three non-routine intra-mathematical problems. In their work, these authors defined two types of flexibility: inter-task flexibility (strategy switching between tasks) and intra-task flexibility (strategy switching within a task). These authors found that students who demonstrated inter-task flexibility were more successful than those who persevered with the same strategy. However, these authors did not find relationship between intra-task flexibility and success. Subsequent studies confirmed these findings (Arslan & Yazgan, 2015; Keleş & Yazgan, 2021).

In contrast, there are few empirical studies linking the development of multiple solution strategies and performance in real-context problem-solving (Achmetli et al., 2019; Schukajlow & Krug, 2014; Schukajlow et al., 2015). There is also little research on pre-service or in-service teachers' flexibility (Berk et al., 2009; Ferrando & Segura, 2020; Lee, 2017; Leikin & Levav-Waynberg, 2007). It is important to learn more about prospective teachers' flexibility because it could provide information about their proficiency in problem-solving. Following Elia et al.'s (2009) definition of inter-task flexibility, in this paper, we consider that a solver shows flexibility across a Fermi problem sequence when switching strategies in any of them (Elia et al., 2009). As we will explain in the methodology section, the number of strategy changes throughout the sequence allows us to define flexibility levels. We will study the relationship between this type of flexibility and prospective teachers' performance in problem solving.

1.3 Teachers' problem-solving performance

From a systematic review of the literature on problem-solving, Chapman (2015) highlighted the importance of teacher knowledge of problem-solving. Some researchers suggest that teachers should experience problem-solving from the problem solver's perspective before they can adequately approach teaching it (Thompson, 1985). In many educational programmes, the importance of teaching mathematics through real-context problems is emphasised (Borromeo Ferri, 2018; Cevikbas et al., 2022; Kaiser & Sriraman, 2006; MEFP, 2022; Schukajlow et al., 2021). It is therefore essential to study prospective teachers' competence as real-context problem solvers in order to assess their specialised content knowledge about these kinds of tasks (Ball et al., 2008).

Schoenfeld (1982) indicated that successful problem-solving performance relies on two conditions. On the one hand, it is necessary to have a knowledge of the basic techniques of problem-solving; in addition, it is necessary for the solver to have a "management strategy" for selecting appropriate approaches and discarding unsuccessful ones. In their study with in-service teachers, Copur-Gencturk and Doleck (2021a) analysed performance in solving verbal problems. In order to measure problem-solving performance (which they refer to as "strategic competence for word problems"), these authors defined three levels based on the ability to devise a valid strategy and to execute this strategy without mathematical errors in order to obtain a correct answer.

In contrast to verbal problems, Fermi problems are open-ended tasks and their solution depends on the assumptions and simplifications made in developing a model of the situation. Consequently, it is not easy to define the solver's performance based on the accuracy of the estimation because, in contrast to the verbal problems analysed by Copur-Gencturk and Doleck (2021a), there is no correct answer for Fermi problems. In the framework of real-context problems, Moreno et al. (2021) developed a qualitative study of performance. In their work, these authors did not carry out the analysis by comparing the productions with a given solution, but by analysing their internal coherence, and identifying errors in the solution strategy. Studying errors made by pre-service teachers when solving real-context problems can be useful to measure their performance level (Klock & Siller, 2020). Based on a review of previous studies (Klock & Siller, 2020; Moreno et al., 2021), Segura and Ferrando (2021) established a system of errors specific to Fermi problems that is the basis for the categorisation of errors used in this study. Thus, within this system of errors, it is possible to differentiate between errors that impede devising a valid strategy (strategic errors) and mathematical errors. Following Copur-Gencturk and Doleck's (2021a) approach, the analysis of the errors made throughout a sequence of Fermi problems will allow us to define three levels of performance.

Drawing on the aspects discussed in the theoretical framework, we focus on studying the relationship between flexibility and performance in solving Fermi problems. Specifically, we aim to address four research questions related to pre-service teachers' behaviour when solving a sequence of four Fermi problems:

Research question 1. What is the level of performance (according to whether or not they make strategic or mathematical errors) of prospective teachers in solving Fermi problems?
Research question 2. What level of flexibility do future teachers show when faced with a sequence of Fermi problems?

Research question 3. Is there a significant correlation between prospective teachers' flexibility and performance across a sequence of Fermi problems?
Research question 4. Focusing on those prospective teachers who do not make strategic errors across the sequence, do those who demonstrate flexibility make fewer mathematical errors?

This is an exploratory study. The answers to the first two questions are a descriptive analysis, while we will approach the last two from an inferential analysis that will allow us to identify the associations between flexibility and performance and between flexibility and number of errors.

2 Methodology

2.1 Description of the participants

The study was carried out at the Faculty of Education of the University of Valencia (Spain); all the participants were pre-service teachers in the Bachelor's in Primary Education. In Spain, teacher training is a four-year university degree during which students receive theoretical and practical training that enable them to work as primary school teachers (with students between the ages of 6 and 12) after graduation. During the first three academic years of the program, pre-service teachers learn mathematics in a 90 hour course that includes content in arithmetic, geometry, statistics, probability, and algebra. In addition, they complete a 60 hour course on the teaching of arithmetic and problem-solving. At the time of data collection, all the participants had completed these two components and were starting the last component related to the didactic-mathematical content. Participation was compulsory, because the activity was part of a mandatory course. The sample consisted of 224 students in their last year of their program; their average age was 23.9 years and 72% were female. This sample represents 25% of the students in this course at the university where we carried out the research. While this is a convenience sample drawn from six different groups of students in the program, it is representative of the population of future teachers who are about to complete their studies in this program because the groups of students in this faculty are heterogeneous in terms of gender, social origin, or academic level.

2.2 Data collection

This is an observational study (Lodico et al., 2010). The two authors were also the teachers of the participating pre-service teachers, and were responsible for data collection. Data were collected in two consecutive years (113 participants in 2017 and 111 in 2018), replicating each time the same data collection with different groups of students. All participants in the study have the same background even though they belong to different groups, because the design of initial teacher training in Spain requires that all future teachers receive the same training in subjects with didactic-mathematical content.

2.2.1 Design of Fermi problem sequence

The instrument used in the study for data collection was a sequence of Fermi problems. The sequence consisted of four tasks requiring the estimation of a large number of elements

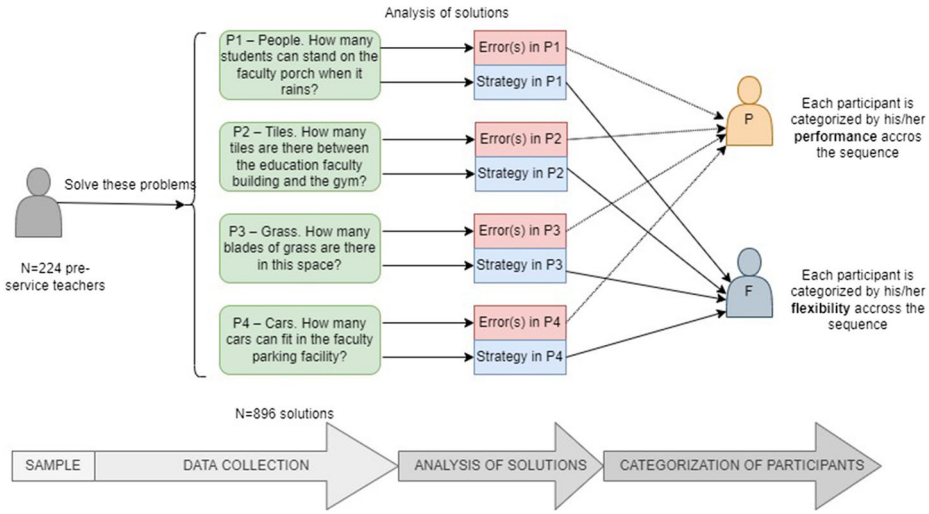


Fig. 1 Description of methodological design of the study

in a rectangular area, large enough for the solver not to obtain the estimate directly. The problems posed in the sequence were set in real locations in the Faculty of Education. Although the context of the problems was familiar to all participants, a picture accompanied each problem statement. To design the sequence, we followed the theory of variation (Ko & Marton, 2004) and the contrast between the contextual variables had been studied with similar problems in previous works (Ferrando et al., 2020, 2021). The contrast in a sequence of problems helps the solver to discern a new aspect of the real situation a problem poses through comparison with another problem that has not changed.

In these problems, students were asked to estimate the number of people (P1), the number of tiles (P2), the number of blades of grass (P3), and the number of cars (P4), in rectangular enclosures with different dimensions (see Fig. 1). The problems differed in the size of the elements (people, tiles, blades of grass, and cars), the total space, the regularity or not of the elements' shape (people and tiles are considered regular), and the order or disorder in the arrangement of the elements (tiles and cars are ordered in rows and columns). Ferrando et al. (2020) justified the value of these variables (dimensions, regularity, and shape) for the Fermi problems used in this sequence. Figure 1 outlines the methodological design of the study. In the following sections, we provide details about the procedure for data collection and analysis of the solutions, and the criteria and categorization of the participants according to their flexibility and performance across the Fermi problem sequence.

2.2.2 Data collection procedure

We conducted data collection during a class session in the last week of September (at the beginning of the academic semester). We presented the sequence of activities to the students as an activity in the Didactics of Geometry and Measurement course. Because the data collection took place at the beginning of the semester, it ensured that the participants were not familiar with estimation strategies in a real context. However, we asked them to

indicate whether they had previously solved similar problems. None of the participants in the study reported prior knowledge of Fermi problems.

During a 90 minute class session, we provided each participant with the written statements of the four problems. During the first ten minutes, we explained to the participants that they were going to face a sequence of four tasks, emphasising that (a) in each problem, they should propose a solution, clearly indicating their strategy and the measures they would need to obtain the estimation; (b) that they should work individually using only paper and pencil to explain their procedures in written form and may use drawings or diagrams; and (c) that they did not need to obtain a numerical solution but rather to explain how to obtain the requested estimate. Once the data collection from the 224 participants was completed, we reviewed the solutions to verify that all students had given an answer to the four problems in the sequence. We collected 896 solutions and scanned them to facilitate the analysis (see Fig. 1).

2.3 Analysis of solutions

The analysis of the 896 solutions was done in two phases. For each problem, we first analyzed and categorized the errors made by the participants. Second, we analyzed and categorized the strategies proposed by the participants.

2.3.1 Error analysis

We relied on Segura and Ferrando's (2021) error system to code each of the solutions. We coded two categories of errors: errors that impeded the devising of a valid strategy, and mathematical errors. To understand the error analysis, we relied on participants' descriptions of the strategies and measurements needed to obtain an estimate, which we requested of students. That is, we did not expect a numerical solution, but rather their explicit reasoning. Thus, in identifying errors, we limited ourselves to analysing the internal consistency of the solution written by each participant and did not attempt to assess the quality of their thinking. Table 1 presents the description of the two categories of error used in the analysis of the collected solutions. To ensure consistency in the analysis and to avoid missing errors, the two authors independently coded the errors in 224 solutions. After pooling the results, the two authors discussed the discrepancies and reached a consensus and, finally, aggregated the results of the analysis.

2.3.2 Analysis of solution strategies

Following the categorisation of strategies set out in Albarracín et al. (2021), we established four categories: linearisation, base unit, density, and incomplete. Table 2 describes the categories and illustrates them with an example.

Once we fixed the categories, three researchers conducted the coding of solutions by strategy. In order to avoid discrepancies and to warrant a reliable analysis, we follow the procedure described by Denzin (2009): two researchers independently coded the solutions of the 224 participants. To analyse the reliability, we made a concordance table and calculated Cohen's kappa (Landis & Koch, 1977), obtaining $\kappa = 0.81$. This is a good value. To clarify the coding criteria in cases of discordance between the two researchers, we discussed each case with a third researcher expert in this topic. Discussions during the writing of this paper led us to revise our initial coding, leading us to code as incomplete those solutions based on exhaustive counting strategy (Albarracín & Gorgorió, 2014).

Table 1 Description of types of errors used in solution analysis

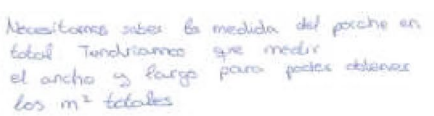
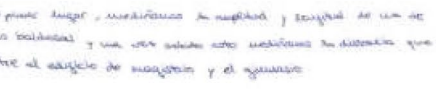
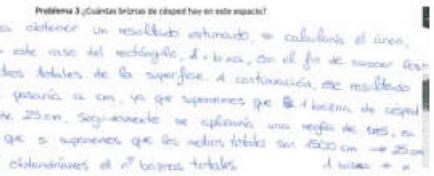
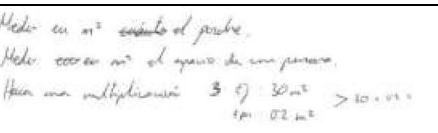
Type of error	Example	Description
Errors that impede devising a valid strategy (strategic errors) can appear during simplification or the structuring phase of the real situation. That is, they appear when the solver does not identify all the variables or all the relevant aspects of the real situation that are necessaries for obtaining a valid strategy. We also include in this category those solutions that only propose counting because this is not a valid strategy since it cannot lead to a solution.	 Solution proposed by a participant to problem *P1 – People*. Translation: "We need to know the total size of the porch. We would need to measure the width and length to get the total m^2."	The strategy described does not allow obtaining an estimate of the number of people that can fit in the porch. The participant only proposes to calculate the area of the enclosure.
	 Solution proposed by a participant to problem *P2 – Tiles*. Translation: "First, we would measure the width and length of one of the tiles, and once we knew this, we would measure the distance between the faculty of education building and the gymnasium."	The solver has identified the two variables needed to develop a strategy to solve the problem, but has not indicated that to obtain the estimate it is necessary to make a measurement division, so the strategy has not been fully developed in the written solution.
Errors during mathematical work (mathematical errors) can relate to deficiencies in measurement skills: an error of perception of the quantity involved, for instance, confusing length and area. We identify these errors by the improper use of measurement units, such as centimetres instead of square centimetres. Errors related to lack of skill in calculation procedures also occur during mathematical work.	 Solution proposed by a participant to problem *P3 – Grass*. Transcription: "To obtain an estimate, we would calculate the area, in this case of the rectangle, $A = b \times a$, to know the total surface metres. Next, this result would be turned into a cm, as we assume that 1 blade of grass measures 25 cm. Following that, a rule of three would be applied, in which if we assume that the total metres are 1500 cm → 25 cm, 1 blade → x. This would give us the total number of blades.	The solver makes an error in reasoning from the measurement of the space occupied by a blade of grass using centimetres.
	 Solution proposed by a participant to problem *P1 – People*. Transcription: "- Measure, in m^2, the porch. - Measure, in m^2, the space of a person. - Make a multiplication. E.g., 30 m^2, 1pers: 0.2 m^2 > 30 x 0.2"	In this solution, the participant confuses multiplication and division.

Table 2 Description of solution strategies

Coded strategy	Example	Description
Linearisation	*P2—Tiles* solution: "First of all I would need to count the number of tiles on one side of the area occupied by these tiles [of the rectangular area between the gymnasium and the teacher training college] and then those on the other perpendicular side, and finally multiply them". [*En primer lugar, tendría que contar el número de baldosas de un lado del área ocupada por estas baldosas y luego las del lado perpendicular, y finalmente multiplicarlas.*]	The solver establishes a model of the elements organised in rows, following a grid distribution model (Albarracín & Gorgorió, 2014). Once the solver obtains the linear estimate of the elements that fit in a row/column, Cartesian product (rows by columns) leads to the estimate of the total number
Base unit	*P4—Cars* solution: "First we would calculate the surface area of the car park by multiplying the width by the length. Then we would calculate how much space a car occupies. Finally, knowing the area occupied by one car, we would divide the total area by the area occupied by one car". [*Primero calcularíamos la superficie del aparcamiento multiplicando ancho por largo. Luego calcularíamos el espacio que ocupa un coche. Por último, conociendo la superficie que ocupa un coche, dividiríamos la superficie total por la superficie de un coche.*]	The process behind this solution strategy is how to cover the total surface area using the area of the element as the unit of measurement: to do this, the solver has to divide the measurement of the total area by the measurement of the area occupied by the element
Density	*P3—Grass*: "Firstly, I would measure one cm long and one cm wide and mark it out, so that we can count how many blades are in one square centimetre. Secondly, I would measure the width and length of the whole lawn to find out the total area in square centimetres. Finally, I would multiply the number of blades in 1 cm^2 by the total area". [*En primer lugar, mediría un centímetro de largo y uno de ancho y lo marcaría, así contaré cuántas hojas hay en un centímetro cuadrado. En segundo lugar, mediría la anchura y la longitud de todo el césped para averiguar la superficie total en centímetros cuadrados. Por último, multiplicaría el número de hojas en 1 cm^2 por la superficie total.*]	This solution strategy bases on delimiting a unit area in order to estimate easily the number of elements per unit area, and then reasoning from proportionality
Incomplete	*P3—Grass* solution: "*I could do the same, but here I cannot calculate the area of a blade because they have very different shapes and sizes.*"[*Podría hacer lo mismo, pero aquí no puedo calcular el área de una brizna porque tienen formas y tamaños muy distintos]*	There are those solutions that are in one of two cases: they do not answer the problem question; or they do not develop a valid solution strategy. In the latter case, this solution necessarily includes an error that impedes devising a valid strategy. In the solution of the example we observe that the participant attempts to apply the Base unit strategy used in problems P1 and P2, but the irregular size and shape of the elements prevents completing a strategy that involves estimating their area

2.4 Categorization of performance and flexibility across the Fermi problem sequence

The analysis of the solutions described previously allowed us to categorize participants' performance and flexibility across the sequence.

2.4.1 Analysis of performance across the sequence

Error analysis of participants' solutions made it possible to define performance across the Fermi problem sequence. Following Copur-Gencturk and Doleck (2021a), we first considered whether the participants devised a valid solution strategy, and then we evaluated the mathematical work, which generated three levels of performance for coding the participants' performance across this sequence.

- Performance across the sequence is low when the participant proposes at least one solution with an error that impedes devising a valid strategy, that is, an incomplete solution. In this case, participant's performance level is coded as 0.
- Performance across the sequence is basic when the participant does not make errors that impede devising a valid strategy in any of the solutions, but there is at least one solution containing a mathematical error. This participant's performance level is coded as 1.
- Performance across the sequence is high when the participant makes no errors in any of the solutions to the four tasks. This participant's performance level is coded as 2.

2.4.2 Analysis of flexibility across the sequence

After the classification of solution strategies (linearisation, base unit, and density), to analyse flexibility across the sequence of Fermi problems, we used the number of strategy changes throughout the sequence. Based on the qualitative analysis of the solutions of each of the 224 participants, we assigned each solution a code of up to four digits according to the proposed solution strategy: incomplete = 1, linearisation = 10, base unit = 100, and density = 1000. By adding the codes of the four solutions of each participant, we obtained a four-digit number that identified the different strategies used by each participant and quantified their strategy changes. Thus, we established the following levels of flexibility across the sequence:

- Non-flexible. Includes prospective teachers who proposed the same solution strategy in all completed tasks
- Moderately flexible. Includes those who proposed two different valid solution strategies, but switched solution strategy only on one problem
- Very flexible. Includes those who proposed two or more different valid solution strategies and switched solution strategy in two or more problems

2.5 Variables involved in the study and statistical analyses performed

In this study, we first analysed each participant's solutions to the four problems in the sequence. This categorisation led to two nominal variables that refer to the solutions: types of errors and types of solution strategies. In addition, from this analysis of errors in the solutions, we also

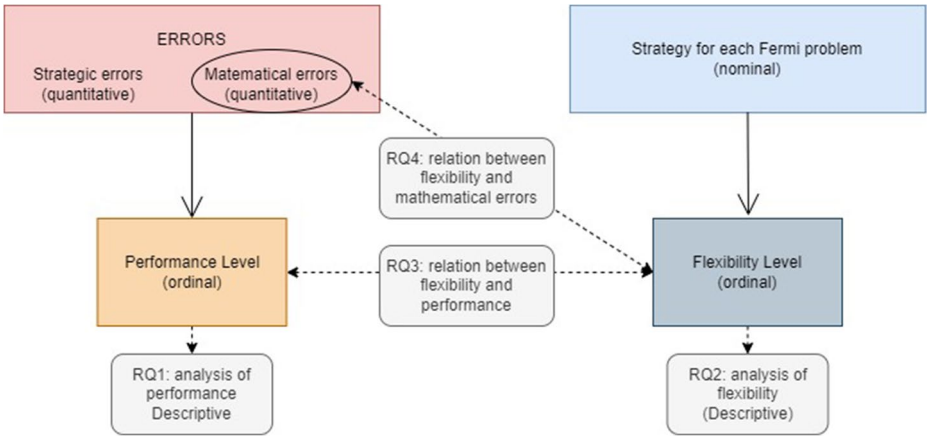


Fig. 2 Description of variables and research questions addressed in this study

obtained a quantitative variable, the number of errors made by each participant throughout the sequence. Using these nominal variables, we categorised participant's performance and flexibility across the sequence by constructing two ordinal variables: performance (coded as 0, 1, or 2) and flexibility (coded as non-flexible, moderately flexible, and very flexible). To answer the first two research questions, we conducted a descriptive analysis. To study performance, we relied on the results relating to performance level as these depend directly on the error analysis. To study flexibility across the sequence, we relied on the ordinal variable flexibility.

To answer to the other two questions, we used inferential analysis. We used a Chi-square test and measured its strength with the Spearman's rank correlation coefficient because we wanted to identify whether there is a significant relationship between two ordinal variables levels of flexibility and performance. Finally, to answer the fourth question, we conducted two parallel analyses, one based on the data from participants with low performance (those who did not complete the sequence) and the other with the data from the rest of the participants. In both cases, we compared whether the distribution of mathematical errors made by participants was different according to their flexibility level. Because the distribution is not normal, we used the non-parametric Kruskal–Wallis test and Dunn's multiple comparison test to determine which specific means were significant with respect to the others.

Figure 2 shows an outline of the variables analysed and the research questions posed in this study.

3 Results

The results are organised into four sections to respond to each of the research questions.

3.1 Pre-service teachers' performance across the sequence

To study performance, we analysed the solutions and identified for each one the type of errors made and their frequency. When analysing the solutions, we counted and coded each and every error made; so we were not actually quantifying wrong solutions, but

Table 3 Distribution of strategic and mathematical errors by problem

Problem	Strategic errors		Mathematical errors	
	Identification of variables	Relation between variables	Measurement procedures	Calculation procedures
P1-People	13	27	62	29
P2-Tiles	8	21	74	17
P3-Grass	36	22	44	28
P4-Cars	11	17	24	28
TOTAL	68 + 87 = 155		204 + 102 = 306	

rather overall errors. From this analysis, we identified 461 errors across all the solutions, 155 (34%) of which were errors that impeded devising a valid strategy (strategic errors) and 306 (66%) were errors during the mathematical work (mathematical errors). Within the strategic errors, we found 68 out of 155 (44%) errors related to not identifying some relevant variable of the real situation. The other 87 (56%) strategic errors corresponded to solutions not including neither quantifying the relationships between variables (see Table 1). Regarding mathematical errors, we found that most, 67% (204 out of 306), corresponded to deficiencies in measurement procedures. In particular, 177 of these 204 errors were related to shortcomings in the perception of quantities (e.g., confusing area and length), and 25 were related to an inadequate use of measurement units. The remaining 102 out of 306 (33%) were related to incorrect calculation procedures (e.g., confusing multiplication and division). Table 3 shows the results of the error analysis for each of the four problems.

In terms of prospective teachers' performance across the Fermi problem sequence, we found that 90 out of 224 (40%) participants demonstrated low performance level, as they had solutions with errors that impeded devising a valid strategy. We found that 77 out of 224 (35%) prospective teachers demonstrated basic performance level, as they only made mathematical errors but managed to develop their strategies. Finally, 57 out of 224 (25%) participants did not make any errors; that is, they demonstrated high performance level.

3.2 Pre-service teachers' flexibility across the sequence

In terms of the strategies proposed by pre-service teachers, we found that the most frequently used strategy was base unit with 45% (407 out of 896), closely followed by density and linearisation, each at 19% (170 and 166 out of 896, respectively). In addition, 17% (153 out of 896) of solutions were incomplete. Regarding flexibility, we found that 66 out of 224 (30%) of the participants demonstrated a very flexible use of strategies across the Fermi problem sequence and 89 out of 224 (40%) demonstrated a moderate flexible use of strategies. Thus, less than one-third of the participants (69, that represent 31%) showed a non-flexible use.

3.3 Relationship between flexibility and performance level across the sequence

Table 4 shows the distribution of the participants by their level of flexibility and their performance level exhibited in their solutions. In the first row, we show the

Table 4 Distribution of the $N=224$ pre-service teachers according to flexibility and performance levels of their solutions

Performance Level	Flexibility Level		
	Non-flexible	Moderately flexible	Very flexible
Level 0. Low	45 (65%)	42 (47%)	3 (5%)
Level 1. Basic	15 (22%)	31 (35%)	31 (47%)
Level 2. High	9 (13%)	16 (18%)	32 (48%)

Note. Level 0. Low: student made errors that impeded devising a valid strategy (then, leading to at least one incomplete solutions). Level 1. Basic: student succeeded in completing all the tasks but made mathematical errors. Level 2. High: students completed all the tasks without errors

number (and the percentage relative to flexibility level) of prospective teachers who made errors that impede devising a valid strategy and therefore had some incomplete solution (low performance) by flexibility level. In the second row, we show the number (and the percentage) of pre-service teachers who succeeded in completing all the tasks but made mathematical errors (basic performance) by the flexibility level. In the third row, we show the number (and the percentage) of prospective teachers who completed all the tasks without errors (high performance) by the flexibility level.

The relationship between flexibility and performance was statistically significant (χ^2 (4, 224) = 58.86, $p < 0.001$) suggesting that both ordinal variables are related. The Spearman's rank correlation coefficient ($r(222) = 0.47$, $p < 0.001$) indicated that there was a significant moderate positive relationship between flexibility and performance. That is, the deviation of the frequency of solvers who were very flexible with high performance was much higher than expected (+ 90.5%), and the deviation of the frequency of very flexible solvers with basic performance was higher (+ 36.6%).

3.4 Flexibility and number of mathematical errors

Overall, we identified a total number of 461 errors; however, when we crosschecked these data with the results of the flexibility analysis, we found differences in the number and type of errors among the flexibility levels. In Table 5, we show the distribution of errors by type (strategic or mathematical) and by performance and flexibility levels.

Because errors that impede devising a valid strategy (strategic errors) determine low performance level, the significant and positive relationship between flexibility and performance levels indicates that strategic errors are related to low levels of flexibility. However, errors during mathematical work (mathematical errors) give us more information about

Table 5 Distribution of errors made by participants depending on their performance and flexibility levels ($N=461$)

		Non-flexible	Moderately flexible	Very flexible
Level 0. Low	Strategic errors	91	61	3
	Mathematical errors	83	67	6
Level 1. Basic	Mathematical errors	35	69	46

how pre-service teachers solve the sequence of Fermi problems. Thus, a focus on those pre-service teachers who completed the sequence (i.e., their performance is basic or high; they have not made strategic errors) allows us to investigate whether flexibility is associated with a decrease in mathematical errors.

We found that, among these $n = 134$ solvers, the average error rate per solver according to flexibility level was 1.46 errors per non-flexible solver, 1.47 errors per moderately flexible solver, and 0.73 errors per very flexible solver. The Kruskal–Wallis H test was significant ($H(2, n = 134) = 6.61$, $p = 0.037$) and the post hoc Dunn's test indicated that the mean rank was statistically significantly different between very flexible and non-flexible solvers ($p < 0.05$) and between very flexible and moderately flexible solvers ($p < 0.05$). There were no significant differences between non-flexible solvers and moderately flexible solvers regarding the number of mathematical errors. Therefore, the distribution of mathematical errors among very flexible prospective teachers with performance level basic or high was significantly lower than in the case of moderately flexible and non-flexible ones. A similar analysis with the $n = 90$ prospective teachers with low performance level revealed no significant differences (H (2, $n = 90$) = 0.58 and $p = 0.75$).

4 Discussion

We found that three-fourths of the pre-service teachers who participated in this study made errors when solving Fermi problems; that the solutions of more than two-thirds of the participants suggested that they were moderately or very flexible in solving Fermi problems; that there was a significant relationship between flexibility and performance; and that when studying in detail the number of mathematical errors made, flexibility was associated with a reduction in such errors. We will now interpret these results according to each of the research questions.

4.1 Pre-service teachers' errors and performance across the sequence

Our analysis of participants' solutions allowed us to categorise their performance level (see Fig. 1) using two types of errors, strategic errors and errors during mathematical work.

In our study, we found that 66% of the errors made were mathematical errors. Among the 306 mathematical errors committed, 204 were errors related to measurement procedures. To solve Fermi problems, it is necessary to make arguments using estimations involving measurements of quantities (e.g., the area of the enclosure or the space occupied by an element). Andrews et al. (2021) pointed out that estimation could appear in four different ways in mathematics teaching and learning. The problems discussed here are in line with measurement estimation, although estimation of numbers is also involved. Indeed, although the aim of the activity is to obtain a quantity estimate, this necessarily requires estimating measurements of the enclosure and, in some cases, of the size of each element. Therefore, measurement and estimation procedures are relevant in the Fermi problem-solving process as is the case in many real-context problems (Hagena, 2015). Prior work on estimation strategies developed by teachers, based on intra-mathematical tasks (Copur-Gencturk, 2022) or verbal problems (Copur-Gencturk & Doleck, 2021b), shows that mathematical errors are due to shortcomings in calculation procedures. This explains why errors made during mathematical work when solving Fermi problems are related to difficulties in measurement or calculation procedures.

If we look at the results at Table 3, which shows the distribution of errors in each problem, we see that problems *People* and *Tiles* have a higher number of mathematical errors. Going deeper into these data, we find that errors related to measurement were quite frequent, especially in the *Tiles* problem, where some solvers estimated the number of tiles that fit in the distance separating the gymnasium and the education building, instead of estimating the number of tiles in the area between both buildings. Ferrando et al. (2021) found that the most frequent strategy in this problem was linearisation (see Table 2). This strategy involves working with distances (rows of tiles), and when solvers used linearisation, they frequently estimated the number of tiles in a row instead of doing an estimation for the whole area. Measurement errors also appeared in *People* and *Grass*, where, trying to apply a base unit strategy, some solvers reasoned using the width of a person or a blade of grass instead of reasoning using the area covered by these elements. For example, in the problem *People*, a solver proposed dividing the area of the porch by the width of a person. In some cases, solvers also referred to volume instead of area. Although these errors did not impede the development of a strategy, they revealed shortcomings in measurement procedures (Segura & Ferrando, 2021). These errors suggest weaknesses in future teachers' ability to visualise and interpret spatial facts and relationships (Lester, 1994).

In the results showed at Table 3, we see that errors related to calculation procedures were considerably less frequent than measurement errors. These errors are mostly inversion errors; this type of error appeared, for example, in some solutions based on Base unit strategy, in which the solver multiplied the total area by the area of the unit element instead of dividing. In other cases, we found reversal errors, for instance, when a solver proposed dividing a car's area by the parking area. Reversal errors also appeared in some solutions based on density strategy, when solvers considered density as a ratio between a unit area and a number of elements. Inversion errors and reversal errors are related to density and base unit strategies, which explains why the number of errors related to mathematical procedures was lower in the *Tiles* problem. Indeed, in this case, the most common solution strategy was linearisation.

Among the solvers who proposed a valid solution strategy for all the problems of the sequence–performance level basic or high, more than one-half (77 out of 134, 57%) made, at least, one mathematical error across the sequence. If we focus on strategic errors leading to incomplete solutions, we find them in 153 out of 896 solutions (17%), which confirms that Fermi problems are accessible (Ärlebäck, 2009). The results on performance throughout the sequence showed that 90 out of 224 prospective teachers (40%) had at least one incomplete solution (low level). These results confirm Chapman's (2015) findings that report problem-solving difficulties for pre-service teachers.

4.2 Pre-service teachers' flexibility across the sequence

The first step in analysing flexibility is to categorise the solution strategies. The results in Section 3.2 show that the most used strategy by prospective teachers throughout the sequence was the base unit. These results confirm the findings presented in Ferrando et al. (2021), which document that, although the context influences the choice of strategy, reasoning from the area occupied by an element and dividing the total area by this value is the most frequent strategy in general. Moreover, base unit strategy is mathematically simpler because it requires operation with quantities of the same type (area of an element and area of the enclosure), whereas the density strategy requires operations involving quantities of a different type (number of elements per unit area and area of the enclosure).

Once we categorised the solutions to each problem in the sequence, we counted the number of valid strategies proposed by each participant and thus deduced their level of flexibility. We found that more than two-thirds of the solvers (155 out of 224) were flexible, as they were able to propose at least two different valid strategies across the sequence. Some previous studies (Chapman, 2015; Van Dooren et al., 2003) have noted that prospective teachers do not have flexibility in problem-solving; we believe that this discrepancy is due to the types of tasks that students are asked to solve. As explained in Ferrando and Segura (2020), the design of Fermi problem sequences based on variation theory (Ko & Marton, 2004) can elicit solvers' flexibility. When looking at the solutions by solvers categorised as non-flexible, we think that that they proposed just one solution strategy, either because they were confident that it worked in all cases (regardless of the context of the problem) or because they did not know any other strategy. In the latter case, as in the last example in Table 2, knowing only one strategy can lead to errors because the solver is not able to adapt the strategy to the context of the problem.

4.3 Relationship between flexibility and performance across the sequence

We found a statistically significant moderately strong relationship between prospective teachers' level of performance and their flexibility level. These results extend, for Fermi problems, those of Elia et al. (2009), Arslan and Yazgan (2015), and Keleş and Yazgan (2021) for non-routine intra-mathematical problem-solving. As shown in Table 4, prospective teachers in our sample who demonstrated no flexibility and those who demonstrated moderate flexibility had similar behaviour in terms of their level of performance across the Fermi problem sequence. However, the behaviour was different for those participants with high flexibility. In particular, the proportion of prospective teachers who failed to complete the sequence was very low in the case of solvers categorised as very flexible (3 out of 66 participants, 5%), while this proportion was higher for the other flexibility levels. In fact, the reason for changing strategy on a single problem may be that the solver did not know how to approach it. In contrast, solvers categorised as very flexible changed several times—indicating a certain fluency—and this could allow them to overcome difficulties, and then complete the Fermi problem sequence. For example, in *Grass*, the disorder and irregularity of the blades of grass, whose number must be estimated, may hinder strategies based on linearisation or base unit, whereas the choice of another strategy, such as density, would avoid them (Ferrando et al., 2021).

These results highlighted the theoretical importance of the three-level categorisation of flexibility, as they allowed us to differentiate performance between those who switched strategy at least twice and those who switched strategy only once. Schoenfeld (1982) defined successful performance in problem-solving as the ability to choose and manage a strategy to obtain a solution. Our results expand this description in the context of Fermi problems: we observed that the level of flexibility related to the ability to complete the sequence. This may be because knowing several strategies and switching flexibly from one to another facilitated the choice of a valid strategy, for a given problem. However, our study is correlational, so it could also be that the ability not to make errors that impede devising a valid strategy influenced the level of flexibility shown by the solver. This possibility seems to us to be a more challenging hypothesis to explain.

4.4 Relationship between flexibility and number of mathematical errors

We found that flexibility was associated with mathematical errors made. When we focused on solvers who completed the Fermi problem sequence (that is, without making strategic errors), we found that prospective teachers categorised as very flexible had significantly fewer mathematical errors than those categorised as non-flexible or moderately flexible. Furthermore, we did not find differences between the number of mathematical errors made by solvers categorised as moderately flexible and non-flexible. This result agrees with the findings from studies that use non-routine intra-mathematical problems (Arslan & Yazgan, 2015; Elia et al., 2009; Keleş & Yazgan, 2021). By differentiating between two types of errors—strategic and mathematical—we were able to refine the findings on the prospective teachers' problem-solving performance and its relation to flexibility in the context of Fermi problems. The interpretation we have used earlier applies here as well: flexibility, which involves strategy management skills (knowing several strategies and switching between them), relates to not making errors that prevent devising a valid strategy. However, when we consider the number of mathematical errors made and their relationship to flexibility, the result is more surprising, as it indicates that high flexibility also relates to the skills—measurement and calculation procedures—needed to correctly implement the strategy.

Further research is needed to determine the possible effects of flexibility on the incidence of mathematical errors, or the possible influence of proficiency in measurement and calculation procedures on the level of flexibility. Newton et al. (2020) highlighted the importance of prior knowledge of concepts and procedures used in problem solving in the development of flexibility. Other authors (e.g., Heinze et al., 2009) related flexibility and adaptive expertise, that is, the ability to effectively apply the most appropriate solution in a given context, which avoids mathematical errors.

Our findings showed that, among the solvers demonstrating low performance level, there was no relationship between flexibility and the number of mathematical errors made. It seems reasonable to assume that if a solver makes strategic errors, even if he or she knows several strategies, he or she does not manage their development. Therefore, it is possible that the solver makes as many mathematical errors as those who know a smaller number of strategies and do not manage them either.

4.5 Limitations

This work is an exploratory study on the relationship between flexibility and performance in solving Fermi problems. This is a correlational study, so its main limitation is that we cannot determine causality between variables. A path analysis (e.g., Schukajlow et al., 2015) could be used to determine whether a solver's level of flexibility influences specific aspects related to performance (e.g., strategic errors, mathematical errors), or whether it is the aspects related to performance that influence solver's level of flexibility. The results of this exploratory study derive from the analysis of the solutions of a sequence of four very specific Fermi problems, all of them contextualised in settings very close to the solvers' reality. This is a limitation of the work, both due to the number of problems and their characteristics. Moreover, in order to analyse the performance and the errors made, we have relied exclusively on the analysis of written productions. A complementary qualitative analysis based on collecting information from observation of the students while they are solving the problems or from subsequent interviews would undoubtedly allow us to better understand and complete the results presented here.

5 Conclusions

Given that studies focused on the use of multiple solving strategies for real-context problems are scarce (Achmetli et al., 2019; Schukajlow et al., 2015), this work contributes a novel approach, using sequences of Fermi problems to study flexibility. Another relevant contribution of this study is extending the notion of inter-task flexibility of Elia et al. (2009), distinguishing between levels of flexibility along a sequence of problems. Furthermore, based on error analysis (Segura & Ferrando, 2021) and drawing on the approach of Copur-Gencturk and Doleck (2021a), we defined performance levels, which allowed us to find that pre-service teachers who demonstrated high flexibility (switching strategy two or more times) performed better: they made fewer errors that impeded developing a valid strategy, and fewer errors in calculation or measurement procedures. These results allowed us to conclude that flexibility is not only a component of problem-solving proficiency (Heinze et al., 2009; Star & Rittle-Johnson, 2008), but it is also directly related to problem-solving performance in terms of errors.

On the other hand, Lu and Kaiser (2022) have recently enriched real-context problem-solving proficiency to include creativity. Flexibility, together with fluency and originality, is one of the criteria for assessing creativity (Levav-Waynberg & Leikin, 2012). A future line of research could deepen and extend this study to investigate the relationship between flexibility, fluency, originality, and performance. This would help to better understand creativity in solving real-context problems, such as Fermi problems. Expanding the study of flexibility to include creativity of prospective teachers when solving real-context problems will contribute to designing training programs for pre-service teachers that reinforce their problem-solving proficiency and ensure the effective introduction of real-context open-ended problems in the classroom.

Author contribution Carlos Segura was the PhD candidate who conducted the research upon which this paper is based. Carlos's PhD was supervised by Irene Ferrando. Carlos Segura prepared the first draft of this document and Irene Ferrando revised, corrected, and added different sections. All authors read and approved the final paper. All authors contributed to the study's conception and design.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work has been done with the financial support of Universitat de València, through the project UV-INV-AE-1557785. Authors are grateful to the support provided by the grants PID2020-117395RB-I00 and PID2021-126707NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe".

Data availability The material presented in this paper was drawn from the data generated in a wider PhD investigation.

Code availability Not applicable

Declarations

Competing interests The authors declare no competing interest.

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

Achmetli, K., Schukajlow, S., & Rakoczy, K. (2019). Multiple solutions to solve real-world problems and students' procedural and conceptual knowledge. *International Journal of Science and Mathematics Education, 17*, 1605–1625. https://doi.org/10.1007/s10763-018-9936-5

Albarracín, L., Ferrando, I., & Gorgorió, N. (2021). The role of context for characterising students' strategies when estimating large numbers of elements on a surface. *International Journal of Science and Mathematics Education, 19*, 1209–1227. https://doi.org/10.1007/s10763-020-10107-4

Albarracín, L., & Gorgorió, N. (2014). Devising a plan to solve Fermi problems involving large numbers. *Educational Studies in Mathematics, 86*(1), 79–96. https://doi.org/10.1007/s10649-013-9528-9

Albarracín, L., & Gorgorió, N. (2019). Using large number estimation problems in primary education classrooms to introduce mathematical modelling. *International Journal of Innovation in Science and Mathematics Education, 27*(2), 45–57. https://doi.org/10.30722/IJISME.27.02.004

Andrews, P., Xenofontos, C., & Sayers, J. (2021). Estimation in the primary mathematics curricula of the United Kingdom: Ambivalent expectations of an essential competence. *International Journal of Mathematical Education in Science and Technology*. https://doi.org/10.1080/0020739X.2020.1868591

Ärlebäck, J. B. & Doerr, H. M. (2015). Moving beyond a single modelling activity. In G. A. Stillman, W. Blum, & M. S. Beimbegut (Eds.), *Mathematical Modelling in Education and Practice* (pp. 293–303). Springer. https://doi.org/10.1007/978-3-319-18272-8_24

Ärlebäck, J. B. (2009). On the use of realistic Fermi problems for introducing mathematical modelling in school. *The Mathematics Enthusiast, 6*(3), 331–364. https://doi.org/10.54870/1551-3440.1157

Arslan, C., & Yazgan, Y. (2015). Common and flexible use of mathematical non routine problem-solving strategies. *American Journal of Educational Research, 3*(12), 1519–1523. https://doi.org/10.12691/education-3-12-6

Ball, D., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59 (5), 389–407. https://doi.org/10.1177/0022487108324554

Berk, D., Taber, S. B., Gorowara, C. C., & Poetzl, C. (2009). Developing prospective elementary teachers' flexibility in the domain of proportional reasoning. *Mathematical Thinking and Learning, 11*(3), 113–135. https://doi.org/10.1080/10986060903022714

Borromeo Ferri, R. (2018). Learning how to teach mathematical modeling in school and teacher education. *Springer*. https://doi.org/10.1007/978-3-319-68072-9

Cevikbas, M., Kaiser, G., & Schukajlow, S. (2022). A systematic literature review of the current discussion on mathematical modelling competencies: State-of-the-art developments in conceptualizing, measuring, and fostering. *Educational Studies in Mathematics, 109*, 205–236. https://doi.org/10.1007/s10649-021-10104-6

Chapman, O. (2015). Mathematics teachers' knowledge for teaching problem solving. *LUMAT: International Journal on Math, Science and Technology Education, 3*(1), 19–36. https://doi.org/10.31129/lumat.v3i1.1049

Copur-Gencturk, Y. (2022). Teachers' knowledge of fraction magnitude. *International Journal of Science and Mathematics Education, 20*, 1021–1036. https://doi.org/10.1007/s10763-021-10173-2

Copur-Gencturk, Y., & Doleck, T. (2021a). Strategic competence for multistep fraction word problems: An overlooked aspect of mathematical knowledge for teaching. *Educational Studies in Mathematics, 107*, 49–70. https://doi.org/10.1007/s10649-021-10028-1

Copur-Gencturk, Y., & Doleck, T. (2021b). Linking teachers' solution strategies to their performance on fraction word problems. *Teaching and Teacher Education, 101*, 103314. https://doi.org/10.1016/j.tate.2021.103314

Denzin, N. K. (2009). *The research act: A theoretical introduction to sociological methods* (3rd ed.). Prentice Hall.

Efthimiou, C. J., & Llewellyn, R. A. (2007). Cinema, Fermi problems and general education. *Physics Education, 42*(3), 253–261. https://doi.org/10.1088/0031-9120/42/3/003

Elia, I., van den Heuvel-Panhuizen, M., & Kolovou, A. (2009). Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. *ZDM the International Journal on Mathematics Education, 41*(5), 605–618. https://doi.org/10.1007/s11858-009-0184-6

Ferrando, I., Segura, C., & Pla-Castells, M. (2021). Analysis of the relationship between context and solution plan in modelling tasks involving estimations. In F. K. S. Leung, G. A. Stillman, G. Kaiser & K.

L. Wong (Eds.), *Mathematical Modelling Education in East and West* (pp. 119–128). Springer. https://doi.org/10.1007/978-3-030-66996-6_10

Ferrando, I., & Segura, C. (2020). Fomento de la flexibilidad matemática a través de una secuencia de tareas de modelización [Fostering mathematical flexibility through a sequence of modelling tasks]. *Avances de Investigación en Educación Matemática, 17*, 84–97. https://doi.org/10.35763/aiem.v0i17.306

Ferrando, I., Segura, C., & Pla-Castells, M. (2020). Relations entre contexte, situation et schéma de résolution dans les problèmes d'estimation [Relationship between context, situation and resolution scheme in estimation problems]. *Canadian Journal of Science, Mathematics and Technology Education, 20*(3), 557–573. https://doi.org/10.1007/s42330-020-00102-w

Govender, R. (2020). Mathematical modelling: A 'growing tree' for creative and flexible thinking in pre-service mathematics teachers. In G. Stillman, G. Kaiser, C. Lampen (Eds.), *Mathematical Modelling Education and Sense-making. International Perspectives on the Teaching and Learning of Mathematical Modelling* (pp 443–453). Springer. https://doi.org/10.1007/978-3-030-37673-4_38

Hagena, M. (2015). Improving mathematical modelling by fostering measurement sense: An intervention study with pre-service mathematics teachers. In G. Stillman, W. Blum & M. Salett Biembengut (Eds.), *Mathematical Modelling in Education Research and Practice. International Perspectives on the Teaching and Learning of Mathematical Modelling*. Springer. https://doi.org/10.1007/978-3-319-18272-8_14

Heinze, A., Star, J. R., & Verschaffel, L. (2009). Flexible and adaptive use of strategies and representations in mathematics education. *ZDM the International Journal on Mathematics Education, 41*, 535–540. https://doi.org/10.1007/s11858-009-0214-4

Kaiser, G., & Sriraman, B. (2006). A global survey of international perspectives on modelling in mathematics education. *ZDM the International Journal on Mathematics Education, 38*(3), 302–310. https://doi.org/10.1007/BF02652813

Keleş, T., & Yazgan, Y. (2021). Gifted eighth, ninth, tenth and eleventh graders' strategic flexibility in non-routine problem solving. *The Journal of Educational Research, 114*(4), 332–345. https://doi.org/10.1080/00220671.2021.1937913

Klock, H., & Siller, H.-S. (2020). A time-based measurement of the intensity of difficulties in the modelling process. In H. Wessels, G. A. Stillman, G. Kaiser, & E. Lampen (Eds.), *International perspectives on the teaching and learning of mathematical modelling* (pp. 163–173). Springer. https://doi.org/10.1007/978-3-030-37673-4_15

Ko, P. Y., & Marton, F. (2004). Variation and the secret of the virtuoso. In F. Marton & A. B. M. Tsui (Eds.), *Classroom discourse and the space of learning* (pp. 43–62). Erlbaum.

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

Lee, M. Y. (2017). Pre-service teachers' flexibility with referent units in solving a fraction division problem. *Educational Studies in Mathematics, 96*(3), 327–348. https://doi.org/10.1007/s10649-017-9771-6

Leikin, R., & Levav-Waynberg, A. (2007). Exploring mathematics teacher knowledge to explain the gap between theory-based recommendations and school practice in the use of connecting tasks. *Educational Studies in Mathematics, 66*(3), 349–371. https://doi.org/10.1007/s10649-006-9071-z

Leikin, R., & Levav-Waynberg, A. (2008). Solution spaces of multiple-solution connecting tasks as a mirror of the development of mathematics teachers' knowledge. *Canadian Journal of Science, Mathematics, and Technology Education, 8*(3), 233–251. https://doi.org/10.1080/14926150802304464

Lester, F. K. (1994). Musings about mathematical problem-solving research: 1970–1994. *Journal for Research in Mathematics Education, 25*(6), 660–675. https://doi.org/10.5951/jresematheduc.25.6.0660

Levav-Waynberg, A., & Leikin, R. (2012). The role of multiple solution tasks in developing knowledge and creativity in geometry. *Journal of Mathematical Behavior, 31*, 73–90. https://doi.org/10.1016/j.jmathb.2011.11.001

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: from theory to practice*. Jossey-Bass. https://doi.org/10.33524/cjar.v14i3.103

Lu, X., & Kaiser, G. (2022). Creativity in students' modelling competencies: Conceptualisation and measurement. *Educational Studies in Mathematics, 109*, 287–311. https://doi.org/10.1007/s10649-021-10055-y

Ministerio de Educación y Formación Profesional (MEFP). (2022). Real Decreto 217/2022, de 29 de marzo, por el que se establece la ordenación y las enseñanzas mínimas de la Educación Secundaria Obligatoria [Royal Decree 217/2022, of 29 March, which establishes the organisation and minimum teaching of Compulsory Secondary Education]. Boletín Oficial del Estado, 76, de 30 de marzo de 2022. https://www.boe.es/eli/es/rd/2022/03/29/217/con

Moreno, A., Marín, M., & Ramírez-Uclés, R. (2021). Errores de profesores de matemáticas en formación inicial al resolver una tarea de modelización [Errors of pre-service mathematics teachers in solving a modelling task]. *PNA, 15*(2), 109–136. https://doi.org/10.30827/pna.v15i2.20746

Newton, K. J., Lange, K., & Booth, J. L. (2020). Mathematical flexibility: Aspects of a continuum and the role of prior knowledge. *The Journal of Experimental Education, 88*(4), 503–515. https://doi.org/10.1080/00220973.2019.1586629

Schoenfeld, A. H. (1982). Measures of problem-solving performance and of problem-solving instruction. *Journal for Research in Mathematics Education, 13*(1), 31–49. https://doi.org/10.5951/jresematheduc.13.1.0031

Schukajlow, S., Kaiser, G. & Stillman, G. (2021) Modeling from a cognitive perspective: theoretical considerations and empirical contributions, *Mathematical Thinking and Learning*, 1–11. https://doi.org/10.1080/10986065.2021.2012631

Schukajlow, S., & Krug, A. (2014). Do multiple solutions matter? Prompting multiple solutions, interest, competence, and autonomy. *Journal for Research in Mathematics Education, 45*(4), 497–533. https://doi.org/10.5951/jresematheduc.45.4.0497

Schukajlow, S., Krug, A., & Rakoczy, K. (2015). Effects of prompting multiple solutions for modelling problems on students' performance. *Educational Studies in Mathematics, 89*(3), 393–417. https://doi.org/10.1007/s10649-015-9608-0

Segura, C., & Ferrando, I. (2021). Classification and analysis of pre-service teachers' errors in solving Fermi problems. *Education Sciences, 11*(8), 451. https://doi.org/10.3390/educsci11080451

Sriraman, B., & Knott, L. (2009). The mathematics of estimation: Possibilities for interdisciplinary pedagogy and social consciousness. *Interchange, 40*(2), 205–223. https://doi.org/10.1007/s10780-009-9090-7

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction, 18*(6), 565–579. https://doi.org/10.1016/j.learninstruc.2007.09.018

Thompson, A. G. (1985). Teachers' conceptions of mathematics and the teaching of problem solving. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives* (pp. 281–294). Erlbaum.

Threlfall, J. (2002). Flexible mental calculation. *Educational Studies in Mathematics, 50*, 29–47. https://doi.org/10.1023/A:1020572803437

Van Dooren, W., Verschaffel, L., & Onghena, P. (2003). Pre-service teachers' preferred strategies for solving arithmetic and algebra word problems. *Journal of Mathematics Teacher Education, 6*(1), 27–52. https://doi.org/10.1023/A:1022109006658

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