Journal of Mathematics Teacher Education (2026) 29:433–458
https://doi.org/10.1007/s10857-025-09685-3

Influence of pre-service primary school teachers' prior knowledge of measurement and measurement estimation in solving modelling problems

C. Segura[1] · **C. Gallart**[1] · **I. Ferrando**[1]

Accepted: 1 February 2025 / Published online: 19 February 2025
© The Author(s) 2025

Abstract
Modelling is an important mathematical competence in many curricula worldwide, in Secondary education but even in Primary Education. Fermi problems (FPs) are suitable activities for introducing mathematical modelling in Primary Education, as they present a real situation, without data, and ask for an estimate of an unattainable quantity. However, implementing such tasks is challenging for primary school teachers, partly due to some gaps in modelling-specific teacher knowledge. Previous studies show a relationship between pre-service teachers' performance and flexibility in solving FPs. Still, the impact of other factors that may affect performance, particularly prior knowledge of measurement and measurement estimation (M&ME), has not been studied. We aim to characterise pre-service teachers' prior M&ME knowledge and to find out what are the relationships between this knowledge, flexibility and performance in solving FPss. To address this, we present the design and validation of a test on M&ME knowledge that pre-service teachers (PSTs) subsequently took. A cluster analysis based on their answers allowed us to characterise three levels of prior knowledge. In addition, PSTs' flexibility and performance in solving a sequence of FPs was analysed, finding that there are significant differences in performance according to their level of prior M&ME knowledge. However, prior knowledge does not influence flexibility. We further find that PSTs' flexibility and prior M&ME knowledge explain an important part of their performance in solving FPs. The results may help to improve PSTs' instruction on modelling problem-solving.

Keywords Fermi problems · Mathematical modelling · Flexibility · Measurement estimation · Pre-service teachers · Performance · Teacher knowledge

✉ I. Ferrando
irene.ferrando@uv.es

C. Segura
carlos.segura@uv.es

C. Gallart
cesar.gallart@uv.es

[1] Departament de Didàctica de les Matemàtiques, Facultat de Magisteri - Universitat de València, Avda. Tarongers, 4, 42022 València, Spain

Introduction

There is a consensus in the academic community of Mathematics Education on the importance of mathematical modelling to develop students' mathematical competence, which in the last decade has been transferred to numerous curricula around the world (Cevikbas et al., 2022). Despite the growing importance of mathematical modelling in educational programmes, its presence in classrooms is still scarce, especially in Primary Education (Borromeo Ferri, 2021). However, Albarracín and Gorgorió (2019) have shown that a type of modelling task, called Fermi problems (FPs), is accessible to primary school students. Indeed, FPs requiring the estimation of a large number of elements in a bounded perimeter or area allow students to develop modelling processes such as developing models to describe the organisation of elements in the space, identify realistic assumptions, estimating measurements of objects and devising strategies for estimating perimeters or areas. As pointed out by Peter-Koop, in FPs "the required modelling process extends beyond the application of a standard algorithm and necessitates multiple modelling cycles with multiple ways of thinking about givens, goals, and solution paths" (2009, p. 142).

To successfully implement these modelling problems in classrooms, teachers must be competent in solving them (Borromeo Ferri, 2018). However, Segura and Ferrando (2021, 2023) show that pre-service primary school teachers (hereafter, PSTs) face difficulties solving FPs, particularly related to the perception of magnitudes (area and perimeter) but also some procedural errors such as using an inverse algorithm or not complete the description of calculation procedures. Therefore, to understand the reasons for these difficulties, it is important to analyse mathematics teachers' specialised knowledge (Ball et al., 2008; Carrillo-Yáñez et al., 2018). This includes content knowledge as a relevant dimension, which comprises knowledge of problem-solving practices (Chapman, 2015) and understanding of the procedures and concepts involved in those practices.

Research (e.g. Heinze et al., 2009) has highlighted that flexibility, that is, the knowledge and use of multiple strategies, is a key skill of problem-solving practices. Segura and Ferrando (2023) found that PSTs should show flexibility in solving FPs, adapting strategies to their contextual characteristics. In addition, to obtain the large number of elements requested by FPs, the strategies require prior knowledge of measurement and measurement estimation (hereafter, M&ME) of lengths and areas (Baturo & Nason, 1996; Gooya et al., 2011; Weiher et al., 2022). Consequently, it is necessary to further research what the PSTs' prior M&ME knowledge actually is and to what extent it has an effect, together with flexibility, on their performance in solving this type of modelling problems.

In this paper, we present the design and validation of a test on knowledge of M&ME of lengths and areas to perform a two-phase study. In the first phase, 87 pre-service teachers took the M&ME test to analyse their prior knowledge. In the second phase, these same participants solved a sequence of FPs. The first objective of this work is to characterise the levels of PSTs' prior M&ME knowledge. In addition, we also analyse their flexibility and performance in solving FPs, which allows us to address the second objective: to study the influence of PSTs' prior M&ME knowledge on their flexibility and performance in solving FPs. Finally, we analyse the effect of flexibility and prior M&ME knowledge on PSTs' performance when solving FPs. It is worthwhile to deepen teachers' M&ME knowledge, as it is a highly relevant topic in primary school curricula (Clements, 2003) and could also be key to solving modelling problems (Hagena, 2015).

Theoretical framework

Teacher knowledge for teaching problem-solving

Shulman (1986) describes teacher's knowledge along seven dimensions, including content knowledge (i.e. of the subject matter) and pedagogical content knowledge (i.e. instruction and learning of the subject matter). Although Shulman stresses the importance of pedagogical content knowledge, other studies demonstrate that content knowledge (CK) serves as a foundational component for the development of pedagogical content knowledge (PCK), highlighting the interdependence between these types of knowledge in effective mathematics teaching (Ball et al., 2001; Blömeke et al., 2022; Depaepe et al., 2013). In this sense, Thompson (1985) had already pointed out that successful learning of problem-solving in Primary Education, requires teachers to be competent problem solvers.

Building on Shulman's work, specific mathematics teacher knowledge frameworks have been developed over the last fifteen years (Chapman, 2013a). For example, Ball and colleagues (Ball et al., 2008; Hill et al., 2008) developed the Mathematical Knowledge for Teaching (MKT), and later Carrillo-Yáñez et al. (2018) developed the Mathematics Teacher's Specialised Knowledge (MTSK) framework. CK is relevant in both frameworks, which characterise it in subdomains: for example, MTSK differentiates between knowledge of topics (KoT) and knowledge of practices in mathematics (KPM).

Problem-solving knowledge brings together knowledge of both topics and mathematical practices (Chapman, 2015), as it requires developing problem-solving strategies but also mobilising prior concepts and procedures (Schoenfeld, 1985). Chapman (2013b) highlights the importance of problem-solving knowledge as part of a specific domain of teaching competence: mathematical task knowledge for teaching.

Fermi problems as modelling activities

The mathematical study of real-world phenomena helps students to make sense of mathematical concepts and procedures (Gravemeijer & Doorman, 1999). Mathematical modelling is an activity that promotes this kind of meaningful learning (Kaiser & Sriraman, 2006). Indeed, modelling problems require the solver to mathematise a real phenomenon or situation, a process in which he or she moves from reality to the mathematical domain (Blum & Niss, 1991; Wijaya et al., 2015). The mathematisation process often involves structuring the real situation, making assumptions, identifying, estimating or measuring relevant aspects or elements to transform them into mathematical objects or variables, and establishing quantifiable relationships between them (Czocher, 2017; Zbiek & Conner, 2006). A mathematical model is generated from these variables and relationships, and when the solver works in the mathematical domain, he/she manipulates this model with mathematical procedures to obtain an answer that must be reinterpreted and validated in the real context (Niss & Blum, 2020). This is why Lesh and Doerr (2003) formulated a definition of a mathematical model that is based on two elements: a conceptual system that describes a real phenomenon and the procedures associated with it for manipulating it and making predictions.

FPs present a real, open-ended situation that requires the solver to provide a quantity that cannot be obtained directly (Ärlebäck, 2009). An example might be to make a reasoned estimate of how many kilograms of waste a city generates in a year. Research has

focused on describing and characterising the solving process when dealing with a FP: it requires the solver to start from prior knowledge about the real world (Henze & Fritzlar, 2010), to decompose the problem into simpler subproblems where informed assumptions (Sriraman & Knott, 2009) can be made to identify and measure or reasonably estimate variables and relationships leading to the solution (Taggart et al., 2007). As pointed out by different research (Albarracín & Gorgorió, 2014; Ärlebäck & Bergsten, 2010; Peter-Koop, 2009), this is a process that requires the development of a mathematical model that is not too complex. That is why FPs have gained importance in recent years as introductory modelling activities. Indeed, some authors (Albarracín & Gorgorió, 2019; Haberzettl et al., 2018; Henze & Fritzlar, 2010) have shown that FPs are suitable for introducing mathematical modelling in Primary Education, especially those that require estimating a large number of elements in a delimited area (e.g. how many people fit on a football field). Despite this, primary school teachers rarely use this type of problem in the classroom (Borromeo Ferri, 2021; Hiebert et al., 2003). FPs present a valuable opportunity for them to expand their knowledge in teaching modelling (Borromeo Ferri, 2018).

Modelling-specific teacher knowledge

We have previously pointed out that within problem-solving knowledge (Chapman, 2013b), knowledge of practices (KPM) and knowledge of topics (KoT) are particularly relevant (Carrillo-Yáñez et al., 2018). Since FPs can be considered as modelling tasks suitable for Primary Education (Albarracín & Gorgorió, 2019; Peter-Koop, 2009), the solving practices and the mathematical contents involved should be analysed from a perspective of modelling-specific primary school teacher knowledge (Borromeo Ferri, 2018). In this perspective, Borromeo Ferri (2018), when describing those skills that are necessary for teaching mathematical modelling, indicates that teachers must be competent in the task dimension, which includes both knowledge of multiple strategies to address modelling problems (which can be considered an aspect of KPM) and knowledge of the cognitive processes, concepts and procedures involved during model development (which is part of the KoT).

Regarding the management of multiple strategies that lead to the development of a mathematical model (KPM), we will focus on flexibility, as it is important for solving FPs, as we will explain in the following subsection. On the other hand, concerning the knowledge of the cognitive processes involved in solving FPs, given that the strategies require identifying and manipulating measures and making estimations (Taggart et al., 2007), in the last subsection, we will concentrate on knowledge of M&ME (KoT).

Flexibility and performance in solving FPs

Ferrando et al. (2020) have analysed and classified the multiple strategies developed by PSTs when solving FPs that consist of estimating a large number of elements on a delimited area (such as those used in the present study). Three strategies were found to develop a model of the situation and obtain the required estimate (Table 1), and we refer to Ferrando et al. (2020) for specific examples of each category.

From Table 1, we infer that the successful development of these strategies requires not only completing the modelling process described in the previous section, but also applying prior M&ME knowledge. In this sense, Schoenfeld (1982a, 1982b) had already indicated that high problem-solving performance depends on two conditions:

Table 1 Ferrando et al. (2020) categories for FP-solving strategies

Linearisation	The solver simplifies the real situation by identifying variables associated with the length to obtain partial estimates of the number of elements per row and the number of rows; and finally, the Cartesian product of the partial estimates is performed. It corresponds to a grid distribution model (Albarracín & Gorgorió, 2014)
Base unit	The solver identifies in the real situation the variables of the total area of the space and area occupied by an element, quantifies them (by measuring or estimating) and relates them through a division of measurements -since the area occupied by an element acts as a benchmark—in the sense established by Joram et al. (2005) in measurement processes
Density	The solver simplifies the real situation by determining a smaller surface (unit area) on which to easily estimate the number of elements, identifying as variables this intensive quantity of "density" (number of elements per unit area) and the total area of the surface, and using proportional reasoning to obtain the required estimate

first, having a "management strategy" (Ibid, p.33) to select appropriate approaches; second, knowing the problem-solving techniques and the mathematical contents involved.

"Management strategy" can be related to the notion of flexibility, which can be defined as the knowledge of multiple strategies and the ability to switch strategies from one problem to another (Elia et al., 2009; Lamb et al., 2023). Several research and educational programmes highlight flexibility as an important learning goal in mathematics education (Hong et al., 2023) not only as a strategic component of problem-solving competence (Heinze et al., 2009; Rittle-Johnson et al., 2012), but also for increasing conceptual and procedural knowledge of mathematics (Rittle-Johnson & Star, 2009). In this line, Segura and Ferrando (2023) defined *flexibility across a sequence of FPs* as the ability to switch strategies from one problem to another. Previous studies (Ferrando et al., 2020; Segura & Ferrando 2023) found that some PSTs have this "management strategy", developing the most appropriate strategy for certain contextual problem characteristics—the size of elements, order or disorder in their distribution, regularity or irregularity in their shapes.

Furthermore, they distinguished between moderate flexibility (change of strategy on only one problem) and high flexibility (change of strategy on more than one problem). Indeed, Segura and Ferrando (2023) found that PSTs' flexibility in solving FPs is related to high problem-solving performance. To this end, the PSTs' *performance in solving FPs* was characterised through a differentiation of the errors committed: On the one hand, there are the errors that prevent the development of a valid strategy to reach the solution, that is, those linked to the modelling process; on the other hand, there are the mathematical errors, committed during the execution of the strategy, linked to prior conceptual and procedural knowledge. Segura and Ferrando (2023) found that highly flexible PSTs made fewer strategic errors, but also fewer mathematical errors (which consisted of calculation errors, proportional reasoning errors and, above all, errors linked to prior knowledge of M&ME of lengths and areas, as can also be seen in Segura and Ferrando, 2021).

While previous studies suggest that knowledge of practice, particularly the ability to choose the most appropriate strategy (flexibility) is related to performance in solving FPs, the role of topic knowledge, particularly M&ME knowledge, needs to be further explored. Indeed, previous studies in other arithmetic or algebra have found that prior

knowledge influences both flexibility and performance (Newton et al., 2020; Rittle-Johnson et al., 2009; Schneider et al., 2011).

Knowledge of measurement and measurement estimation

The results by Albarracín and Gorgorió (2014) suggest that FPs require a deep understanding of mathematical knowledge to build a mathematical model during the development of the solving strategy. As explained above, PSTs showing low performance make numerous errors, most of them related to the M&ME of lengths and areas (Segura & Ferrando, 2021, 2023). Beyond its application to FPs, teachers should have a solid knowledge of this topic, as it is relevant in Primary school curricula (Clements, 2003; Subramaniam, 2014; Towers & Hunter, 2010).

Regarding *measurement*, Baturo and Nason (1996) have studied PSTs' knowledge of area measurement, finding that they have difficulties in applying this knowledge in complex situations. These authors stress that, when solving area measurement tasks, different cognitive processes are involved. The first cognitive process is to know and identify what attribute (magnitude) is being measured and why (e.g. the area enclosed by a plane shape is related to its size and corresponds to a value called quantity). Next, the measurement of a quantity—such as length or area—requires an understanding of the concept unit of measurement—a subdivision of the magnitude—and the process of iterating it over the extent of the quantity without gaps or overlaps (Baturo & Nason, 1996; Dickson et al., 1984). This unit can be taken from the situation to be measured (non-standard) or determined in advance (standard), and the direct measurement process involves counting or quotative division, covering one quantity with another (Joram et al., 2005; Sowder, 1992). However, some PSTs do not show a deep understanding of the concept of area unit, then they tend to use indirect procedures to measure areas, for example, by using the Cartesian product of length and width, or by memorising and applying formulas (Baturo & Nason, 1991; Castillo-Mateo et al., 2012; Dickson et al., 1984). Even using indirect procedures, Castillo-Mateo et al., (2012, 2017) have found that PSTs have errors in terms of magnitude itself (e.g. confusing width or length) or in the application of formulas (confusing area and perimeter, as highlighted by Dickson et al., 1984). Baturo and Nason (1996) also stress that PSTs need to have an understanding of which units of measurement are appropriate by size and shape for measuring an area, and they need to know the relationships between standard units of measurement (e.g. knowing how to convert 1 $m^2 = 10{,}000$ cm^2).

In some measurement situations, direct or indirect comparison of a unit is not possible and must be replaced by internalised referents (benchmarks), i.e. a mental idea of its quantity (Joram, 2003; Joram et al., 1998, 2005). This perceptual ability to approximate different quantities in common objects is called *measurement estimation* (Hogan & Brezinski, 2003; Sowder, 1992). Several authors (Castillo-Mateo et al., 2012, 2017; Gooya et al., 2011; Hildreth, 1983; Hoth et al., 2023; Joram et al., 2005) have characterised measurement estimation strategies based on the measurement skills described above, but applied to benchmarks, without using measurement tools.

Measurement estimation is an important real-life skill (Gooya et al., 2011), but real situations pose several difficulties—the size of the elements, accessibility or visibility of the benchmark—to the estimation (Hoth et al., 2023; Weiher et al., 2022). Since procedural tools cannot be used, measurement estimation is a complex activity involving cognitive processes related to the fundamental principles of measurement (Gooya et al.,

2011; Joram et al., 1998). Indeed, PSTs face challenges in solving modelling problems that require measurement estimates of areas and lengths, such as FPs (Segura & Ferrando, 2021, 2023). However, we need to know more about how this prior M&ME knowledge affects the solution of FPs, as other aspects could also affect performance, such as flexibility in the use of strategies, and we aim to understand more about the relationships between them. To this end, we pose three research questions:

RQ1. What are the levels of pre-service primary school teachers' prior knowledge of measurement and measurement estimation (M&ME) of areas and lengths?
RQ2. Are there differences in pre-service teachers' flexibility and performance in solving Fermi problems according to their level of prior M&ME knowledge?
RQ3. What effect do pre-service teachers' prior M&ME knowledge and flexibility have on their performance in solving Fermi problems?

Methodology

Sample

In Spain, primary school teachers must pass a four-year university degree in which they receive theoretical and practical training that qualifies them to work with pupils aged 6–12 years. The study was carried out during the academic year 2022/2023, with a convenience sample of 87 third-year students of the Degree in Primary School Education at the Faculty of Teacher Training of the University of Valencia (Spain). The sample represents 10% of the students of this course, with an average age of 22.2 years and 63% of women. Moreover, the design of the degree course is regulated by the state and is quite homogeneous in all Spanish universities, so the sample can be considered representative of the students completing the degree. The participants were just starting a mathematics didactics course and had previously taken a 90-h mathematics course that included arithmetic, geometry, statistics, probability and algebra, which guarantees homogeneity in their mathematical background, particularly about basic mathematical concepts.

Data collection

In this subsection, we will describe the instruments and materials used to collect the data, as well as the implementation of the experience.

Design and validation of the test on knowledge of measurement and measurement estimation of areas and lengths (M&ME test)

The authors designed a test focused on the knowledge needed to measure and estimate measures, based on a systematic review of research on the topic, as described in the theoretical framework (e.g. Baturo & Nason, 1996; Hildreth, 1983; Joram et al., 1998, 2005; Weiher et al., 2022). The English version of the test can be found in the Online Appendix. As can be seen in Table 2, we synthesised that research into nine categories, each describing the different cognitive processes involved in the M&ME knowledge. Then, we develop 14 items that specify these categories concerning the ability to measure and estimate areas

Table 2 Content of the M&ME test items and relation to cognitive processes involved in the M&ME knowledge

M&ME cognitive processes	Item	Content
CP1. Perception of magnitude	1	Recognition and understanding of width
CP2. Internalisation (mental idea) of the benchmark	2.1	Approximate idea of the area (length and width) of a benchmark
	2.2	Approximate idea of the numerosity of the referent set
CP3. Concept of unit of measurement	3	Understanding of the area unit
CP4. Appropriate use of the unit of measurement	4	Correct use of the area unit
CP5. Relationships between standard units of measurement	5	Conversion of standard units of area
CP6. Direct measurement	6.1	Understanding measurement as covering one area with another
	6.2	Understanding measurement as quotative division of quantities
CP7. Indirect measurement	7	Area from Cartesian product of lengths
CP8. Terms of the measurement process	8.1	Appropriate use of the necessary terms (lengths, widths) to obtain the number of elements in a rectangular surface
	8.2	Difference between area and perimeter
	8.3	Proportional reasoning to obtain the number of elements in a rectangular area
CP9. Application of formulas	9.1	Understanding the formulae for obtaining the area of a compound shape
	9.2	Understanding the formulae for obtaining the perimeter of a compound shape

and lengths. We decided to design a short test to facilitate its application. However, five of the nine cognitive processes are only covered by a single item, which forces us to be cautious when interpreting the results, as we will explain in the limitations section.

Each test item is formulated as a multiple-choice question. (The correct answer must be chosen from three options.) Incorrect choices reflect obstacles in knowledge about M&ME of areas and lengths related to each content. They have been selected from previous studies on errors in M&ME of areas and lengths (Castillo-Mateo et al., 2012; Segura & Ferrando, 2021). Figure 1 shows an example of two test items.

Once the first version of the M&ME test was designed, researchers with experience in this area of mathematics education were contacted to carry out a content validation, which is a process that aims to ensure that each test item measures the content it is expected to measure (Frank-Stromberg & Olsen, 2004). We constituted a panel of nine experts, who were given a version of the test in which each item was accompanied by a description of the categories of knowledge and content (Table 2) it was intended to measure. Following Lawshe's method, which is widely used to establish and quantify the content validity of each item (Ayre & Scally, 2014), the experts had to choose whether the item asked for what we wanted to detect by assessing its relevance (choosing between Essential, Useful but improvable, and Unnecessary). In addition, in the case of not choosing Essential, they had an open field for suggestions and/or proposals for changes.

To assess the relevance of each item, the CVR (Content Validity Ratio) was calculated after the first round, which measures (with a value between 0 and 1) that a sufficient proportion of experts on the panel have rated the item as essential. With nine experts, the CVR must be equal to or greater than 0.78 (Ayre & Scally, 2014). Six of the 14 items were selected without modifications ($CVR \geq 0.78$), and the other eight had a CVR below 0.78, although no item was considered unnecessary by the experts. The suggestions and changes we collected from the experts were implemented for these items in a second version of the test. After a second round of evaluation by the panel of nine experts, 11 items were left with a $CVR = 1$ and 3 items with $CVR = 0.78$, thus validating the final version of the test.

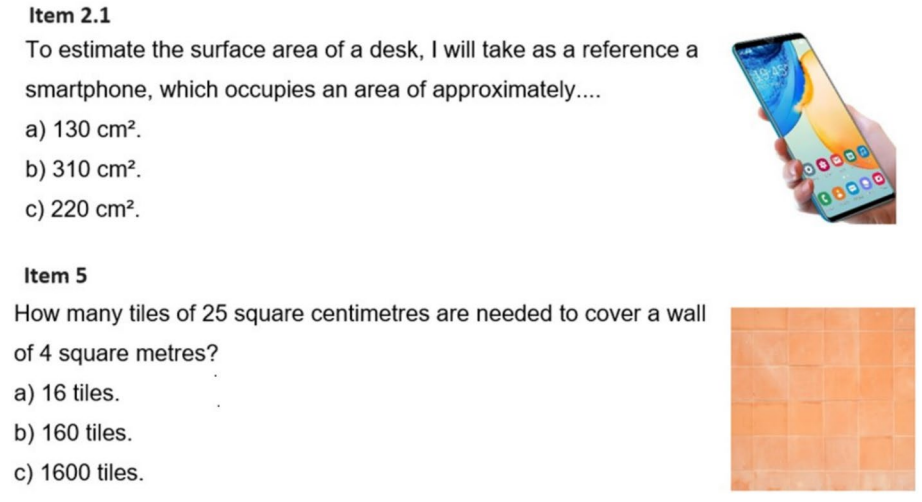


Fig. 1 M&ME test item 2.1 (Internalisation of the benchmark) and item 5 (Relationships between standard units of measurement)

This version was piloted with a small group of 10 PSTs who had to answer the M&ME test and indicate, for each item, whether they had difficulties understanding the statement or the multiple-choice answers. No comprehension problems were reported.

Fermi problems sequence

For the second part of the experience, the authors designed a sequence of 7 FPs requiring the estimate of a large number of elements on a delimited surface (see Online Appendix). All the problems present a written statement with a direct question, accompanied by a picture of the real location of the environment of the Faculty of Teacher Education in which the problem is situated, and some of them have been used in previous studies (Ferrando et al., 2020; Segura & Ferrando 2021, 2023). These works explain that the sequence of FPs was designed by contrasting contextual characteristics such as the size of the elements to be estimated, their regularity in shape or the order/disorder in their distribution, with the aim that the solvers perceive that in some problems it could be more appropriate to use one strategy or another, although all of them are susceptible to be solved with the same strategy (see Table 1). Figure 2 shows two problems in the sequence: in P1, the elements (cars) are large, always rectangular and arranged neatly in rows; in P2, the elements (blades of grass) are small, irregular in shape, and arranged unordered over the surface of the garden.

P1. How many cars can fit in the faculty car park without leaving spaces?



P2. How many blades of grass are there in this garden?



Fig. 2 Two problems (P1, P2) of the sequence of FPs used

Although a model of the problem situation can be built using either strategy, it is clear that we will have a more accurate result for P2 if we use the density strategy—because the size occupied by a blade is irregular and the organisation of the blades is disordered. In P1, the linearisation or unit base strategies are more straightforward—because of the regularity and size of the area occupied by a car—than calculating the density of cars per unit area. Therefore, it is a sequence of problems that can be solved with flexibility in the use of multiple strategies or without flexibility, using a single strategy.

Data collection procedure

The experience was conducted in two consecutive 90-min sessions, within the participants' regular classes. Two of the authors were also the participants' teachers and were responsible for data collection. Participants were notified that they were participating in research and gave their informed consent.

In the first session, participants completed the M&ME test. This took between 30 and 45 min. In the second part of the session, one of the researchers explained what a FP is and also explained in a schematic way (similar to Table 1) the three strategies for developing a mathematical model to obtain the solution. This part of the session is brief because its purpose is not to teach how to solve FPs (assumed accessible), but to remind participants of well-known processes (linearising, iterating unit or calculating density) and to indicate that they will be able to apply one or more of them to FPs in the next session.

In the second session, two days later, we gave the statements of the seven FPs to the participants. Before solving them, one of the researchers gave the prompt that they could choose two ways to solve the problems: (a) using several strategies for the problems in the sequence, or (b) using only one strategy for the whole sequence of problems. Participants spent about 60 min solving the FPs.

Data analysis

The answers to the M&ME test and the FPs solutions collected were analysed and categorised by the researchers.

Categorisation of prior knowledge levels based on a cluster analysis

As explained in Table 2, each of the nine cognitive processes involved in M&ME knowledge is specified in one or more test items. Each of these 14 test items is scored 0 if the answer is incorrect and 1 if the answer is correct. Thus, the score of each participant in each cognitive process can be obtained by averaging the items related to each process. Furthermore, for each participant, we have a 0–14 scale variable reflecting their total knowledge of M&ME of areas and lengths.

To obtain the *levels of PSTs' prior knowledge*, a hierarchical cluster analysis based on Ward's method was applied. This is one of the most widely used methods in practice as it is more discriminating in determining the levels of clustering than other methods based on means, distances or centroids (Kuiper & Fisher, 1975). Looking at Ward's linkage dendrogram generated in SPSS, the researchers opted for 3 clusters (emerging, medium and high knowledge), as three groups of similar size are obtained (37, 26 and 24 participants, respectively).

Categorisation of flexibility in solving FPs

The 609 solutions (87 participants solving 7 FPs) were analysed by two researchers. Following previous work (Ferrando et al., 2020; Segura & Ferrando, 2023), solutions were categorised using the strategies described in Table 2 (Linearisation, Base Unit and Density), and those solutions that did not devise a valid strategy were categorised as Incomplete. These categories have already been validated in previous studies and have high reliability (Ferrando et al., 2020; Segura & Ferrando, 2023).

First, two researchers independently categorised a sample of 126 solutions (26% of the total), with a concordance of $\kappa = 0.88$ on Cohen's kappa. A few cases of disagreement were discussed with the third researcher. Subsequently, the categorisation of the remaining solutions was shared between the two researchers. Once each solution was categorised, the coding process described in Segura and Ferrando (2023) was followed to measure flexibility: a participant is categorised as *non-flexible* (coded as 0) if he/she uses the same strategy in all problems in the sequence; as *moderately flexible* (coded as 1) if he/she switches strategy in only one problem in the sequence; and as *highly flexible* (coded as 2) if he/she switches strategy in two or more problems in the sequence.

Categorisation of performance in solving FPs

As described in Segura and Ferrando (2023, p. 214), before categorising performance in solving FPs, it is necessary to differentiate between strategy errors that prevent the development of a mathematical model (strategic errors) and mathematical errors during the execution of the strategy (mathematical errors). Strategic errors occur during the process of structuring, simplifying and mathematising the real situation posed in the problem, and are due to the failure to formulate some relevant variable or some mathematical relationship between the variables. These are errors when building the mathematical model and can lead to the categorisation as Incomplete. On the other hand, mathematical errors are conceptual or procedural errors associated with the mathematical work within the model, for example, errors in the quantification (estimation or measurement) of the variables or the identification of their relationships.

Two researchers analysed the errors in the 609 solutions. For each participant in each problem, a performance variable with values between 0 and 2 was assigned. Following Segura and Ferrando (2023), performance on a problem was scored 0 if the participant made strategic errors, 1 if the participant made only mathematical errors, and 2 if the participant made no errors. Two researchers independently categorised the performance on 161 problems (26% of the total). Cohen's kappa was calculated to obtain a concordance measure of $\kappa = 0.84$, which gives a high reliability of categorisation. The remaining solutions were split between the two researchers.

Summing the performance on each problem in the sequence, for each participant, yields an overall *FP-solving performance* variable that takes values on a 0–14 scale. To find out the reliability of the performance scale by solving the sequence of FPs, we calculated Cronbach's alpha coefficient, obtaining good reliability ($\alpha = 0.8$).

Statistical analyses performed

From the above analyses, we obtain three scale variables for each PST: their level of prior M&ME knowledge, their level of flexibility and their performance in solving Fermi problems.

To answer the first research question, we first analyse the correlations between the PSTs' scores on the nine cognitive processes measured in the test (Table 2) to ensure a redundancy-free characterisation of M&ME knowledge levels based on their differences. Indeed, a two-way mixed ANOVA is then performed to find whether there are significant differences between the scores in the nine cognitive processes (within-subjects factor, repeated measures) and to find in which of them there are significant differences between the emerging, medium- and high-level groups (between-subjects factor). These results allow us to characterise three profiles of PSTs' knowledge of M&ME of areas and lengths.

To answer the second research question, we describe for each level of prior M&ME knowledge (emerging, medium and high) what is the average flexibility and average performance in solving FPs, and we use a one-way ANOVA to find if the differences are significant.

In cases where there is no normality or homoscedasticity in the data distribution, the sample size is sufficient to opt for parametric tests, which remain preferable because of their robustness (Rasch & Guiard, 2004). In the case of the repeated measures ANOVA test, the data did not meet the assumption of sphericity but the Greenhouse–Geisser correction was used.

To answer the third research question, we first obtain Pearson correlations between prior knowledge, flexibility and performance in solving FPs. Then, to find out to what extent flexibility and prior M&ME knowledge affect performance in solving FPs, we use a multiple linear regression model and the R^2 to estimate its predictive power.

Results

Pre-service teachers' levels of prior M&ME knowledge

The overall mean score (*PrK*) of the 87 PSTs on the M&ME test was $PrK = 8.24$ (out of 14). As explained in the methodology section, to characterise the PSTs' prior knowledge of M&ME of areas and lengths, three clusters were obtained, corresponding to emerging—37 participants with *PrK* of 3 to 7 points, medium—26 participants with *PrK* of 8 to 9 points and high—24 participants with *PrK* of 10–14 points—prior knowledge.

To characterise the three levels of M&ME knowledge based on the differences in the acquisition of the nine cognitive processes (CP1 to CP9, see Table 2), we should first verify that they hardly correlate with each other and do not give rise to latent variables that can explain prior knowledge in a shared way. Table 3 shows the correlation matrix between the PSTs' scores on the cognitive processes involved in M&ME knowledge.

We observe that there are significant correlations only between the internalisation process of the benchmark (CP2) and the understanding of the unit of measurement concept (CP3); between the direct (CP6) and indirect (CP7) measurement process; and between the knowledge of the terms of the measurement process (CP8) and the application of formulas (CP9). However, these correlations are weak; in fact, Bartlett's test of sphericity shows that the correlation matrix in Table 3 is not significantly different from the identity matrix

Table 3 Correlation matrix between PSTs' scores on the nine cognitive processes (CP1 to CP9) involved in M&ME knowledge

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9
CP1	1								
CP2	0.05	1							
CP3	0.05	0.25*	1						
CP4	0.11	0.06	−0.01	1					
CP5	0.11	−0.12	0.17	0.08	1				
CP6	−0.00	0.13	0.18	0.20	−0.05	1			
CP7	0.06	0.10	0.09	0.21	0.04	0.23*	1		
CP8	0.20	0.02	0.14	0.03	0.10	−0.17	0.02	1	
CP9	0.19	−0.11	0.05	0.06	0.16	−0.06	0.05	0.24*	1

* Significant with $p < .05$

Table 4 PSTs' mean scores at each level of prior knowledge on the nine cognitive processes (CP1 to CP9) involved in M&ME knowledge

	N	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9
Emerging	37	0.22	0.57	0.16	0.19	0.08	0.51	0.86	0.62	0.38
Medium	26	0.46	0.81	0.27	0.42	0.27	0.63	1	0.74	0.56
High	24	0.79	0.75	0.58	0.63	0.42	0.71	0.96	0.92	0.85
	87	0.49	0.71	0.34	0.41	0.26	0.62	0.94	0.76	0.60

($p = 0.12$). Therefore, it is not necessary to perform an exploratory factor analysis to group cognitive processes into new latent variables.

Table 4 shows the mean scores of emerging-, medium- and high-level PSTs on the nine cognitive processes involved in M&ME knowledge. This reflects the degree of knowledge at each level for each process, and we can observe appreciable differences both between cognitive processes and between knowledge levels.

We conducted a two-way mixed ANOVA test to find: (i) whether there are significant differences in the degree of knowledge of the nine M&ME cognitive processes; (ii) in which of these nine M&ME cognitive processes there are significant differences in knowledge among the three level groups; (iii) whether the three level groups interact differently with the nine cognitive processes involved in M&ME knowledge.

Thus, we found that there are significant differences between PSTs' scores in the cognitive processes involved in M&ME of areas and lengths, obtaining $F\ (8, 672) = 29.46$ with $p < 0.001$ and a high effect size ($\eta^2 = 0.26$). Bonferroni post hoc tests indicate that:

- PSTs' knowledge of the concept of unit of measurement (CP3, unit of area) and of the relationships between standard units of measurement (CP5, conversion between units of area) is significantly lower than that of the other categories ($p < 0.05$).
- Although higher than the above, PSTs also show significantly lower knowledge ($p < 0.05$) in the perception of magnitude (CP1, recognition and understanding of

width) and in the appropriate use of the unit of measurement (CP4, correct use of the area unit).
- On the other hand, PSTs have a significantly higher knowledge of indirect measurement procedures (CP7, area from the Cartesian product of lengths) than of the other categories ($p < 0.001$).

Furthermore, there are obviously significant differences between the scores of the emerging-, medium- and high-knowledge groups (F (2, 84) = 111.17, $p < 0.001$, $\eta^2 = 0.73$). Post hoc analyses of pairwise comparisons allow us to detect in which cognitive processes involved in M&ME knowledge these differences lie:

High-knowledge PSTs score significantly better than medium ($p < 0.05$) and emerging-knowledge PSTs ($p < 0.001$) in the perception of spatial magnitudes (CP1), understanding of the concept of unit area (CP3), knowledge of terms in the process of measuring areas and lengths (CP8) and application of formulas for area and perimeter (CP9). Moreover, the high-level group scores significantly higher than the emerging-level group also in all other categories ($p < 0.05$), except for knowledge of indirect measurement procedures (CP7), where there is no difference between groups.
In turn, medium-knowledge PSTs score significantly better than emerging-knowledge PSTs ($p < 0.05$) in the perception of spatial magnitudes (CP1), in the internalisation of a benchmark (CP2), in the appropriate use of the area unit (CP4), in the knowledge of terms in the process of measuring areas and lengths (CP8), and in the application of formulas for area and perimeter (CP9).

The 24 PSTs (28% of the total) with high knowledge of M&ME of areas and lengths are characterised by a mean score higher than 0.5 in all cognitive processes involved in this topic, except for the conversion of standard units of area. In contrast, the 37 PSTs (42%) showing emerging knowledge have an average score of insufficient in five of the nine cognitive processes. The 26 PSTs (30%) with medium knowledge have an average score of insufficient in four categories.

Finally, we found that there are no significant differences in the cognitive processes that are easier or more difficult according to the PSTs' level of knowledge (F (16, 672) = 1.62, $p = 0.08$). That is, the three-level groups score higher and lower in the same M&ME categories, as explained above.

Differences in pre-service teachers' flexibility and performance in solving FPs according to their levels of prior knowledge

Among 87 PSTs' 609 solutions to the sequence of 7 FPs, 383 (63%) were classified as the base unit strategy; 92 (15%) as the density strategy; and 32 (5%) as the linearisation strategy. In addition, 102 solutions (17%) were not sufficiently developed to be classified as a strategy and classified as Incomplete. The analysis of the use of multiple strategies per participant allowed us to categorise the participants by their level of flexibility. We found that 25 PSTs (29%) decided to use the same strategy for all problems (non-flexible); 35 PSTs (40%) switched strategy in only one problem (moderately flexible); and 27 PSTs (31%) switched strategy in two or more problems (highly flexible).

The 87 PSTs' average performance in solving FPs was 9.4 out of 14. As explained in the methodology, each problem was assigned 0 to 2 points depending on whether strategic

Table 5 Mean flexibility (standard deviation) and mean performance (standard deviation) in solving FPs according to PSTs' prior M&ME knowledge

Prior knowledge	N	Flexibility	Performance in FPs
Emerging	37	0.95 (0.70)	7.38 (3.42)
Medium	26	0.92 (0.79)	9.62 (3.59)
High	24	1.25 (0.85)	12.29 (2.12)

errors, mathematical errors or no errors were made. We found 131 solutions (21% of the total) in which errors were made that prevented the devising of the strategy. Strategic errors were related to the lack of identification or incorrect identification of variables of the real situation posed by the problem, with the lack of identification of a relationship between them, or with their expression in mathematical language. In 138 solutions (23%), we did not find strategic errors, but we did find errors in the mathematical work; although the variables were well formulated, they were not well quantified or there were procedural errors or errors in the calculation of the mathematical relationships between them. In 340 solutions (56%), no errors were identified.

Table 5 shows the 87 PSTs' average flexibility and average performance in solving FPs according to their level of prior knowledge of M&ME of areas and lengths.

We observe that PSTs with high prior M&ME knowledge are more flexible than the others, although using the one-way ANOVA test, we obtain that this difference is not significant (F (2, 84) = 1.44 with $p = 0.24$). Differences are also evident in the performance in solving FPs. In this case, the one-way ANOVA test confirms that they are significant, with F (2, 84) = 17.53 and $p < 0.001$. Moreover, the effect size in the differences is large ($\eta^2 = 0.30$). With the HSD Tukey post hoc test, we obtain that PSTs with high M&ME knowledge perform significantly better than those with medium ($p = 0.01$) and emerging M&ME knowledge ($p < 0.001$). Furthermore, PSTs showing medium M&ME knowledge also perform significantly better than those showing emerging prior knowledge ($p < 0.05$).

Effect of flexibility and prior knowledge on pre-service teachers' performance in solving FPs

We have seen that PSTs' performance in solving FPs depends on their prior knowledge of M&ME of areas and lengths. However, PSTs' flexibility in the use of multiple strategies does not seem to depend on their prior M&ME knowledge.

To explore the extent to which these variables are related, in Table 6, we show the Spearman correlations between scores on the M&ME test, demonstrated flexibility and performance in solving FPs. Indeed, the relationship between prior knowledge and performance in solving FPs is significant and strong ($\rho = 0.57$). Furthermore, we find that there

Table 6 Spearman correlations between PSTs' prior M&ME knowledge, flexibility and performance in solving FPs

	Prior knowledge	Flexibility	Performance in FPs
Prior knowledge	1		
Flexibility	0.12	1	
Performance in FPs	0.57***	0.37***	1

*** Significant with $p < .001$

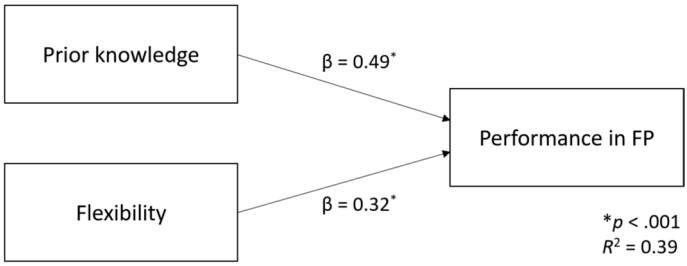


Fig. 3 Regression model of the effect of prior knowledge and flexibility on performance solving FPs

is a significant relationship of medium size ($\rho = 0.37$) between flexibility and performance in solving FPs. However, there is no relationship between prior M&ME knowledge and flexibility.

Based on the above relationships, we hypothesise that prior M&ME knowledge and flexibility are independent predictors of FP-solving performance, and we formulate a multiple linear regression model to obtain the extent to which they explain changes in performance (Fig. 3).

The model is statistically significant with $F\ (2, 84) = 26.91$ and $p < 0.001$. Furthermore, $R^2 = 0.39$ indicates that PSTs' prior M&ME knowledge and their flexibility using multiple strategies explain 39% of the variance in FP-solving performance. Individual factors were further examined and indicated that prior knowledge ($\beta = 0.49$, $t = 5.74$ with $p < 0.001$) and flexibility ($\beta = 0.32$, $t = 3.67$ with $p < 0.001$) are significant predictors that have a positive effect on performance in solving FPs.

Discussion

What are the levels of PSTs' prior knowledge of M&ME of areas and lengths?

It has been pointed out that knowledge of topics is part of the content knowledge that every teacher must master (Ball et al., 2008; Carrillo-Yáñez et al., 2018; Depaepe et al., 2013). The importance of M&ME of areas and lengths in Primary Education (Clements, 2003; Subramaniam, 2014; Towers & Hunter, 2010) makes the design and validation of a test to characterise this PSTs' knowledge a relevant result in itself. The validated M&ME test includes nine cognitive processes involved in the M&ME knowledge: perception of magnitude; internalisation of the benchmark; concept of measurement unit; appropriate use of the measurement unit; relationship between standard units of measurement; direct measurement; indirect measurement; terms of the measurement process; and application of formulas. Analysing the correlations in the PSTs' test scores, we found that some of these cognitive processes are related (understanding the concepts of unit of measurement and benchmark; calculating direct and indirect measurements; using the terms involved in measurement appropriately and applying formulas to obtain it), but cannot be reduced to each other or grouped into new variables. This makes it possible to use the acquisition of cognitive processes to describe PSTs' prior knowledge of M&ME in levels.

Hence, the M&ME test results (see Table 4) show that PSTs demonstrate varying levels of knowledge on these topics, highlighting opportunities for further development (Castillo-Mateo et al., 2012, 2017; Gooya et al., 2011; Joram et al., 1998; Weiher et al., 2022). In particular, we found that PSTs should enhance their knowledge of the cognitive processes related to the concept of unit of area (CP3), as well as of the relationships between different standard area units (CP5). They also need opportunities to strengthen their ability to perceive some spatial magnitudes (CP1, differentiating between width and area), and to use the area unit appropriately (CP4). This stems from the opportunity to deepen their understanding that a unit of measurement is a subdivision of the area that iterates over its extent, as already pointed out by several works (Baturo & Nason, 1996; Dickson et al., 1984; Joram et al., 2005). In contrast, they have hardly any difficulty in understanding the indirect measurement process (CP7, Cartesian product of lengths), which can be explained by previous studies indicating that PSTs are more accustomed to indirect processes for measuring areas (Baturo & Nason, 1996; Castillo-Mateo et al., 2012, 2017).

Our results show that less than 30% of the PSTs have a high M&ME knowledge. The average score of high-knowledge PSTs is satisfactory (mean score above 0.5) in eight of the nine cognitive processes in the M&ME test. Only the average knowledge of the relationship between standard area units (CP5) needs growth opportunities (although it is close to 0.5, while the medium and emerging levels score below 0.3). It is worth noting that those items scoring above 0.75 imply that a large proportion of the high-knowledge PSTs can perceive magnitudes (CP1), give an approximate idea of a benchmark (CP2, measurement estimation for both continuous and discrete quantities, as differentiated by Sowder, 1992), understand indirect measurement (CP7), use all terms of the measurement process appropriately (CP8), and apply formulae to obtain areas and perimeters (CP9). This means that they are proficient in almost all the cognitive processes related to M&ME knowledge. The high-level group stands out from the other two levels, as its degree of knowledge is not only higher in all categories, but this difference is significant in eight of the nine categories (except for CP7, related to indirect measurement procedures, which is similar for all three groups).

Conversely, the PSTs' average score in the emerging-knowledge group indicates improvement needs (mean score below 0.5) in five cognitive processes related to M&ME knowledge. Although the medium-knowledge group demonstrates significantly better scores than the emerging-knowledge group in several M&ME cognitive processes, their average score still has room for improvement in four cognitive processes, only one less than the emerging-knowledge group. Thus, these two groups, comprising more than 70% of the PSTs, highlight significant learning needs in the topic of M&ME of areas and lengths.

Only two of the nine M&ME cognitive processes are successfully addressed by a large number of medium-level PSTs (mean score above 0.75): giving an approximate idea of a benchmark (CP2, measurement estimation) and understanding indirect measurement procedures (CP7). In the case of the emerging-level group, this generalised proficiency is only found in one category: indirect measurement procedures (CP7). So we can state that, among PSTs, only measurement estimation (mastered by medium- and high-level PSTs) and the use of indirect measurement procedures (mastered by all) are sufficiently consolidated cognitive processes related to prior knowledge of M&ME of areas and lengths.

The characterisation by levels presented in this paper is relevant because it broadens and deepens previous studies on knowledge for teaching M&ME of areas and lengths (Baturo & Nason, 1996; Castillo-Mateo et al., 2012, 2017). Furthermore, as we will discuss in the next research question, these levels allow us to address how knowledge of the topics,

together with knowledge of practices (Carrillo-Yáñez et al., 2018), is integrated into problem-solving knowledge (Chapman, 2013b, 2015).

Are there differences in PSTs' flexibility and performance in solving FPs according to their level of prior knowledge?

Knowledge of the mathematical practices involved in problem-solving (Carrillo-Yáñez et al., 2018; Chapman, 2015) requires managing multiple strategies in modelling tasks such as FPs (Borromeo Ferri, 2018; Segura & Ferrando, 2023). In this sense, we found that 70% of PSTs are (moderately or highly) flexible, changing strategy in some problem, a proportion similar to that found in previous works (Segura & Ferrando, 2023).

Since, during strategy implementation, knowledge of the topics is integrated into the knowledge of the mathematical task (Chapman, 2013b), one would expect prior M&ME knowledge to have some influence on flexibility. Indeed, Rittle-Johnson et al. (2009) point out that high prior knowledge benefits the flexible use of multiple strategies. Although in our results in Sect. "Differences in pre-service teachers' flexibility and performance in solving FPs according to their levels of prior knowledge" we find that PSTs with high prior M&ME knowledge demonstrate greater flexibility in solving FPs (see Table 5), the difference is not significant compared to those with medium or emerging knowledge. Thus, in the modelling domain, the influence of prior M&ME knowledge on the flexible use of multiple strategies is less determinant than in other domains such as algebra (Newton et al., 2020). This can be explained by the strong connection with the reality of FPs, as they require prior knowledge about the real situation (Henze & Fritzlar, 2010) and following modelling processes (Albarracín & Gorgorió, 2014; Ärlebäck & Bergsten, 2010; Peter-Koop, 2009). Previous work (Ferrando et al., 2020) confirms that certain contextual characteristics of the real situation—the size and shape of the items, order or disorder in their distribution—do significantly influence the choice of strategies.

Previous studies have shown that PSTs face challenges when solving modelling problems that require estimating areas or lengths (Segura & Ferrando, 2021, 2023). In this study, when analysing performance, we again find these types of difficulties, related to errors in the development of the strategy—such as errors concerning the modelling process, identifying and mathematising variables or relationships between them—and mathematical work—including calculation errors, but also related to M&ME of areas and lengths. Only in half of the PSTs' solutions, there are no errors at all. So we wonder whether some level of prior M&ME knowledge guarantees that PSTs can solve FPs successfully. Our results (see Table 5) show that, indeed, prior knowledge influences performance in solving FPs: PSTs showing high knowledge of M&ME of areas and lengths are significantly more successful in solving FPs than those showing medium and emerging knowledge, and make hardly any errors (12.3 points out of 14). Even the medium-knowledge PSTs also perform significantly better than the emerging-knowledge ones, who only successfully solve half of the FPs (7.4 points out of 14).

This shows that a high proportion of PSTs must enhance their knowledge about M&ME of areas and lengths to successfully solve the FPs. Although these are modelling problems that require extra-mathematical knowledge (Blum & Niss, 1991), the role of prior knowledge in solving them is very important, in line with work in other domains such as equations (Rittle-Johnson et al., 2009; Schneider et al., 2011). In addition, the success of high-knowledge PSTs also shows that if teachers are trained in the knowledge of these topics,

they can overcome difficulties solving FPs (Segura & Ferrando, 2021, 2023), even if they have not received explicit training on modelling. This is because they are accessible problems (Albarracín & Gorgorió, 2019; Ärlebäck, 2009), but only when solvers have a solid prior knowledge of M&ME of lengths and areas.

What effect do PSTs' prior knowledge and flexibility have on their performance in solving FPs?

In our experimental design, PSTs completed a M&ME test, then received schematic instruction on what strategies can be used to solve FPs, and finally had to solve FPs with a prompt to choose whether to use one or several strategies for the sequence. This allows us to complete the results of the third research question by calculating to what extent prior M&ME knowledge and flexibility predict the PSTs' performance in solving FPs. The results of the correlation analysis (Table 6) confirm that there is no relationship between flexibility and prior M&ME knowledge. This reinforces the differentiation in problem-solving between strategy management and strategy implementation that was already pointed out by Schoenfeld (1982a, 1982b) and by Copur-Gencturk and Doleck (2021). We also found that there is a strong correlation between PSTs' prior knowledge and their performance in solving FPs, which complements the findings discussed in the previous research question. Furthermore, there is also a medium correlation between PSTs' flexibility and performance when solving FPs, very similar to that found in a previous study (Segura & Ferrando, 2023), which confirms the robustness of the finding.

Using a multiple linear regression model (see Fig. 3), we find that, indeed, PSTs' prior M&ME knowledge and PSTs' flexibility are significant predictors with a positive effect that explain 39% of their performance in solving FPs. Thus, a very relevant percentage of success can be explained by prior knowledge of M&ME of areas and lengths, and by the knowledge and use of various strategies.

This result has an important impact on the training of primary school teachers, as it highlights the importance of strong knowledge of topics and the management of multiple strategies (as already pointed out by Borromeo Ferri, 2018) in their competence to successfully solve—a necessary condition for their effective teaching—modelling problems suitable for primary school. Consequently, it is necessary to enhance the initial training of future teachers in M&ME and flexibility if we want to introduce specific-modelling teacher knowledge successfully. Above all, deep conceptual and procedural understanding of the area unit (CP3 and CP4) should be improved. To this end, in line with Baturo and Nason (1996), it is necessary to develop practices that help future teachers to correctly interpret the area units, beyond simply calculating area as the product of lengths or as the result of applying a formula. In this sense, it is advisable to use real contexts in which PSTs must determine and use non-standard units of area. Our results also highlight the need to strengthen the instruction of future teachers in the relationships between standard units of measurement (CP5), avoiding algorithmic and decontextualised exercises of unit conversions in which there is no reflection on the procedure. On the other hand, to promote PSTs' flexibility, it is important to incorporate the comparison of strategies across lessons, as pointed out by Rittle-Johnson and Star (2009).

On the other hand, our results let us infer that 60% of the variance in FP-solving performance could be explained by multiple factors, not only cognitive, but also metacognitive (Stillman & Galbraith, 1998) or affective: beliefs and emotions such as interest,

engagement with the task or self-concept (Schukajlow et al., 2017, 2023). Among the cognitive factors, we believe that prior competence in modelling may have an important effect. Indeed, in the problem-solving performance analysis, we found that 21% of the errors detected were related to the modelling process linked to the development of the strategy: structuring and simplifying the real situation, identifying variables and mathematical relationships between them (Czocher, 2017; Zbiek & Conner, 2006).

Anyway, it is important to note that the study of the effect is based on the whole participants' sample. It is worth recalling that the results discussed in the previous section show that PSTs with high prior M&ME knowledge have a very high performance in solving FPs. This leads us to infer that, as Krawitz and Schukajlow (2018) point out, prior knowledge can have an important impact (positive when it is high, negative when it is low) on the cognitive processes involved in modelling.

Limitations

We are conscious of several limitations of our research. First, it would be interesting to replicate the study elsewhere and thus expand the sample size and representativeness, which are limited, although the results are robust.

On the other hand, we should point out that the design of the M&ME test could be improved because, as we have already mentioned in the methodology section, we opted for a short test and several cognitive processes are covered by only one item. This design should make us cautious about interpreting the PSTs' test results in terms of cognitive processes, as they are broader than the specific item they refer to (see Table 2). This also implies another limitation related to the reliability of the test: although we validated the M&ME test using a panel of experts and calculated the relevance of each item with the content validity ratio, we should have included several items per cognitive process in order to measure its reliability with quantitative methods. This in turn would have allowed us to calculate the reliability of the scale "prior knowledge of M&ME". In our study, this construct makes sense from the level-by-level description in response to RQ1, but its consistency needs to be studied quantitatively. Consequently, extending the M&ME test with more items is important to ensure a more systematic, in-depth and accurate assessment of each cognitive process. For example, some studies (Hoth et al., 2023; Meyer & Greefrath, 2023) have further assessed length estimation skills by developing reliable tests. In addition, a more comprehensive M&ME test would allow us to study the effect of each cognitive process on performance in solving FPs, which would allow us to deepen the results of RQ3.

Finally, we must note the limitation of not having measured PSTs' prior modelling competence with a test such as the one developed by Wess et al. (2021). Thus, we would have the possibility to introduce a third predictor in the multiple regression model and calculate what percentage of the performance in solving FPs is explained by these three cognitive factors.

Conclusions

In this paper, we have characterised pre-service primary school teachers' knowledge of measurement and measurement estimation (M&ME) of areas and lengths at three levels, finding that less than one-third of the participants have mastered this knowledge. Besides being an important topic in the Primary Education curriculum, pre-service teachers' prior M&ME knowledge is important because it significantly influences their performance in solving a type of problem that is particularly suitable for introducing modelling in primary school classrooms: Fermi problems (FPs). Consequently, we have found that a high M&ME knowledge guarantees the correct implementation of a strategy for solving FPs. Our results also show that flexibility in the use of multiple strategies also influences pre-service teachers' performance, although this is a strategic management skill that does not seem to be related to prior M&ME knowledge. Finally, we have observed that both factors together explain much of the pre-service teachers' performance in solving FPs.

These findings are relevant because they broaden and deepen the research on teachers' knowledge of important mathematics contents: M&ME of spatial magnitudes, flexibility in the use of multiple problem-solving strategies, and modelling problems. Our study provides empirical evidence that contributes to laying the foundations for consistent and effective primary school teacher education. The deepening of contents related to M&ME, as well as the development of problem-solving flexibility in primary school teacher education will equip pre-service teachers with the necessary knowledge to promote the development of modelling competence among their students, one of the key competences in mathematics curricula around the world (Cevikbas et al., 2022).

A future research direction that opens up with the instruments developed and the results of this work is to study the influence of factors such as flexibility and prior M&ME knowledge on aspects of teachers' competencies such as noticing or the assessment of students' solutions (Blömeke et al., 2022).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10857-025-09685-3.

Acknowledgements The authors are grateful for the support provided by the Grant No. PID2021-126707NB-I00 funded by MCIU/AEI/10.13039/501100011033 and by "ERDF A way of making Europe".

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of 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

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

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

Ärlebäck, J., & Bergsten, C. (2010). On the use of realistic Fermi problems in introducing mathematical modelling in upper secondary mathematics. In R. Lesh, P. Galbraith, C. Haines, & A. Hurford (Eds.), *Modeling students' mathematical modeling competencies* (pp. 597–609). New York: Springer. https://doi.org/10.1007/978-1-4419-0561-1_52

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development, 47*(1), 79–86. https://doi.org/10.1177/0748175613513808

Ball, D. L., Lubienski, S. T., & Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 433–456). Macmillan.

Ball, D. L., Thames, M. H., & 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

Baturo, A., & Nason, R. (1996). Students teachers' subject matter knowledge within the domain of area measurement. *Educational Studies in Mathematics, 31*(3), 235–268. https://doi.org/10.1007/BF00376322

Blömeke, S., Jentsch, A., Ross, N., Kaiser, G., & König, J. (2022). Opening up the black box: Teacher competence, instructional quality, and students' learning progress. *Learning and Instruction, 79*, 101600. https://doi.org/10.1016/j.learninstruc.2022.101600

Blum, W., & Niss, M. (1991). Applied mathematical problem solving, modelling, applications, and links to other subjects—State, trends and issues in mathematics instruction. *Educational Studies in Mathematics, 22*(1), 37–68. https://doi.org/10.1007/BF00302716

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

Borromeo Ferri, R. (2021). Mandatory Mathematical Modelling in School: What Do We Want the Teachers to Know? In F. K. S. Leung, G. Stillman, G. Kaiser & K. L. Wong (Eds.), *Mathematical Modelling Education in East and West* (pp. 103–117). Springer. https://doi.org/10.1007/978-3-030-66996-6_9

Carrillo-Yañez, J., Climent, N., Montes, M., Contreras, L. C., Flores-Medrano, E., Escudero-Ávila, D., Vasco, D., Rojas, N., Aguilar-González, Á., Ribeiro, M., & Muñoz-Catalán, M. C. (2018). The mathematics teacher's specialised knowledge (MTSK) model. *Research in Mathematics Education, 20*(3), 236–253.

Castillo-Mateo, J. J., Segovia, I., Castro, E., & y Molina, M. (2012). Categorización de errores en la estimación de cantidades de longitud y superfície. In D. Arnau, J. L. Lupiáñez, & A. Maz (Eds.), *Investigaciones en Pensamiento Numérico y Algebraico e Historia de la Matemática y Educación Matemática* (pp. 63–74). SEIEM.

Castillo-Mateo, J., Segovia, I., & Molina, M. (2017). Estudio comparativo de la estimación de cantidades continuas que hacen los estudiantes de secundaria y futuros maestros. *PNA, 12*(1), 45–62. https://doi.org/10.30827/pna.v12i1.6538

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

Chapman, O. (2013a). Investigating teachers' knowledge for teaching mathematics. *Journal of Mathematics Teacher Education, 16*, 237–243. https://doi.org/10.1007/s10857-013-9247-2

Chapman, O. (2013b). Mathematical-task knowledge for teaching. *Journal of Mathematics Teacher Education, 16*, 1–6. https://doi.org/10.1007/s10857-013-9234-7

Clements, D. (Ed.). (2003). *Learning and teaching measurement*. National Council of Teachers of Mathematics.

Copur-Gencturk, Y., & Doleck, T. (2021). Strategic competence for multistep fraction word problems: An overlooked aspect of mathematical knowledge for teaching. *Educational Studies in Mathematics, 107*, 49–70.

Czocher, J. A. (2017). How can emphasizing mathematical modeling principles benefit students in a traditionally taught differential equations course? *The Journal of Mathematical Behavior, 45*, 78–94. https://doi.org/10.1016/j.jmathb.2016.10.006

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and Teacher Education, 34*, 12–25. https://doi.org/10.1016/j.tate.2013.03.001

Dickson, L., Brown, M., & Gibson, O. (1984). *Children learning mathematics: A teacher's guide to recent research*. Cassell Educational Ltd.

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 Mathematics Education, 41*(5), 605–618. https://doi.org/10.1007/s11858-009-0184-6

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

Frank-Stromberg, M., & Olsen, S. J. (2004). *Instruments for clinical health-care research*. Jones & Bartlett.

Gooya, Z., Khosroshahi, L. G., & Teppo, A. R. (2011). Iranian students' measurement estimation performance involving linear and area attributes of real-world objects. *ZDM Mathematics Education, 43*, 709–722. https://doi.org/10.1007/s11858-011-0338-1

Gravemeijer, K., & Doorman, M. (1999). Context problems in realistic mathematics education: A calculus course as an example. *Educational Studies in Mathematics, 39*(1–3), 111–129. https://doi.org/10.1023/A:1003749919816

Hagena, M. (2015). Improving mathematical modelling by fostering measurement sense: An intervention study with pre-service mathematics teachers. In G. Stillman, W. Blum, y B. Maria S (Eds.), *Mathematical Modelling in Education Research and Practice. International Perspectives on the Teaching and Learning of Mathematical Modelling*. (pp. 185–194). Springer. https://doi.org/10.1007/978-3-319-18272-8_14

Haberzettl, N., Klett, S., & Schukajlow, S. (2018). Mathematik rund um die Schule—Modellieren mit Fermi-Aufgaben. In K. Eilerts, & K. Skutella (Eds.), *Neue Materialien für einen realitätsbezogenen Mathematikunterricht 5. Ein ISTRON-Band für die Grundschule* (pp. 31–41). Springer Spectrum.

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

Henze, J. y Fritzlar, T. (2010). Primary school children's model building processes by the example of Fermi questions. In A. Ambrus, & E. Vásárhelyi (Eds.), *Problem Solving in Mathematics Education. Proceedings of the 11th ProMath conference* (pp. 60–75). Eötvös Loránd University.

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., Chui, A. M-Y., Wearne, D., Smith, M., Kersting, N., Manaster, A., Tseng, E., Etterbeek, W., Manaster, C., Gonzales, P., & Stigler, J. W. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*. National Center for Education Statistics.

Hildreth, D. J. (1983). The use of strategies in estimating measurements. *Arithmetic Teacher, 30*(5), 50–54.

Hill, H., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education, 39*(4), 372–400. https://doi.org/10.5951/jresematheduc.39.4.0372

Hogan, T. P., & Brezinski, K. L. (2003). Quantitative estimation: One, two, or three abilities? *Mathematical Thinking and Learning, 5*(4), 259–280. https://doi.org/10.1207/S15327833MTL0504_02

Hong, W., Star, J. R., Liu, R. D., Jiang, R., & Fu, X. (2023). A systematic review of mathematical flexibility: Concepts, measurements, and related research. *Educational Psychology Review, 35*(4), 104. https://doi.org/10.1007/s10648-023-09825-2

Hoth, J., Heinze, A., Huang, H. M. E., Weiher, D. F., Niedermeyer, I., & Ruwisch, S. (2023). Elementary school students' length estimation skills—Analyzing a multidimensional construct in a cross-country study. *International Journal of Science and Mathematics Education, 21*(6), 1841–1864. https://doi.org/10.1007/s10763-022-10323-0

Joram, E. (2003). Benchmarks as tools for developing measurement sense. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement* (pp. 57–67). National Council of Teachers of Mathematics.

Joram, E., Gabriele, A. J., Bertheau, M., Gelman, R., & Subrahmanyam, K. (2005). Children's use of the reference point strategy for measurement estimation. *Journal for Research in Mathematics Education, 36*(1), 4–23.

Joram, E., Subrahmanyam, K., & Gelman, R. (1998). Measurement estimation: Learning to map the route from number to quantity and back. *Review of Educational Research, 68*(4), 413–449.

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

Krawitz, J., & Schukajlow, S. (2018). Activation and monitoring of prior mathematical knowledge in modelling processes. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.). *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 243–250). PME.

Kuiper, F. K., & Fisher, L. (1975). A Monte Carlo comparision of six clustering procedures. *Biometrics, 31*, 777–783.

Lamb, L., Bishop, J., Whitacre, I., & Philipp, R. (2023). Flexibility across and flexibility within: The domain of integer addition and subtraction. *The Journal of Mathematical Behavior, 70*, 101031. https://doi.org/10.1016/j.jmathb.2023.10103

Lesh, R. A., & Doerr, H. M. (Eds.) (2003). *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Lawrence Erlbaum Associates.

Meyer, M., & Greefrath, G. (2023). Students' measurement estimation skills and their relation to the process of solving Fermi problems. In *Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)* (No. 29). ERME.

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

Niss, M., & Blum, W. (2020). *The learning and teaching of mathematical modelling,* Routledge.

Peter-Koop, A. (2009). Teaching and Understanding Mathematical Modelling through Fermi-Problems. In B. Clarke, B. Grevholm & R. Millman (Eds.), *Tasks in primary mathematics teacher education* (pp. 131–146). Springer. https://doi.org/10.1007/978-0-387-09669-8_10

Rasch, D., & Guiard, V. (2004). The robustness of parametric statistical methods. *Psychology Science, 46*, 175–208.

Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology, 101*(3), 529–544.

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2009). The importance of prior knowledge when comparing examples: Influences on conceptual and procedural knowledge of equation solving. *Journal of Educational Psychology, 101*(4), 836. https://doi.org/10.1037/a0016026

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2012). Developing procedural flexibility: Are novices prepared to learn from comparing procedures? *British Journal of Educational Psychology, 82*(Pt 3), 436–455. https://doi.org/10.1111/j.2044-8279.2011.02037.x

Schneider, M., Rittle-Johnson, B., & Star, J. R. (2011). Relations among conceptual knowledge, procedural knowledge, and procedural flexibility in two samples differing in prior knowledge. *Developmental Psychology, 47*(6), 1525. https://doi.org/10.1037/a0024997

Schoenfeld, A. H. (1982a). Measures of problem-solving performance and of problem-solving instruction. *Journal for Research in Mathematics Education, 13*(1), 31–49.

Schoenfeld, A. H. (1982b). 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

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

Schukajlow, S., Rakoczy, K., & Pekrun, R. (2017). Emotions and motivation in mathematics education: Theoretical considerations and empirical contributions. *ZDM Mathematics Education, 49*(1), 307–322. https://doi.org/10.1007/s11858-017-0864-6

Schukajlow, S., Rakoczy, K., & Pekrun, R. (2023). Emotions and motivation in mathematics education: Where we are today and where we need to go. *ZDM Mathematics Education, 55*(2), 249–267. https://doi.org/10.1007/s11858-022-01463-2

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

Segura, C., & Ferrando, I. (2023). Pre-service teachers' flexibility and performance in solving Fermi problems. *Educational Studies in Mathematics*, *113*(2), 207–227. https://doi.org/10.1007/s10649-023-10220-5

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. https://doi.org/10.2307/1175860

Sowder, J. (1992). Estimation and Number sense. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning. A project of the National Council of Teachers of Mathematics* (pp. 371–389). Macmillan.

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

Stillman, G., & Galbraith, P. L. (1998). Applying mathematics with real world connections: Metacognitive characteristics of secondary students. *Educational Studies in Mathematics, 36*(2), 157–194. https://doi.org/10.1023/A:1003246329257

Subramaniam, K. (2014). Prospective secondary mathematics teachers' pedagogical knowledge for teaching the estimation of length measurements. *Journal of Mathematics Teacher Education, 17*, 177–198. https://doi.org/10.1007/s10857-013-9255-2

Taggart, G. L., Adams, P. E., Eltze, E., Heinrichs, J., Hohman, J., & Hickman, K. (2007). Fermi Questions. *Mathematics Teaching in the Middle School, 13*(3), 164–167. https://doi.org/10.5951/MTMS.13.3.0164

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). Routledge.

Towers, J., & Hunter, K. (2010). An ecological reading of mathematical language in Grade 3 classroom: A case of learning and teaching measurement estimation. *The Journal of Mathematical Behavior, 29*, 25–40. https://doi.org/10.1016/j.jmathb.2009.12.001

Weiher, D. F., Ruwisch, S., Huang, H.-M.E., Hoth, J., & Heinze, A. (2022). Modelling the complexity of measurement estimation situations: A theoretical framework for the estimation of lengths. *Mathematica Didactica*. https://doi.org/10.18716/ojs/md/2022.1593

Wess, R., Klock, H., Siller, HS., Greefrath, G. (2021). Test Instrument. In: *Measuring Professional Competence for the Teaching of Mathematical Modelling. International Perspectives on the Teaching and Learning of Mathematical Modelling*. Springer, Cham. https://doi.org/10.1007/978-3-030-78071-5_3

Wijaya, A., van den Heuvel-Panhuizen, M., & Doorman, M. (2015). Opportunity-to-learn context-based tasks provided by mathematics textbooks. *Educational Studies in Mathematics, 89*(1), 41–65. https://doi.org/10.1007/s10649-015-9595-1

Zbiek, R. M., & Conner, A. (2006). Beyond motivation: Exploring mathematical modeling as a context for deepening students' understandings of curricular mathematics. *Educational Studies in Mathematics, 63*, 89–112. https://doi.org/10.1007/s10649-005-9002-4

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