European Journal of Science and Mathematics Education, 2026, **14**(1), 1-27
ISSN: 2301-251X (Online)

OPEN ACCESS

Research Article

Design and validation of the P-MATHEX questionnaire: Assessing preservice teachers' perceptions of pedagogical knowledge in mathematics and science

María Santágueda-Villanueva [1]
0000-0002-5472-7972

Maria Teresa Sanz [2*]
0000-0002-7146-8087

Emilia López-Iñesta [2]
0000-0002-1325-2501

Lidón Monferrer [1]
0000-0001-8117-6814

Gil Lorenzo-Valentín [1]
0000-0002-2812-5740

[1] Departamento de Educación y Didácticas Especificas, Universitat Jaume I, Castellón de la Plana, SPAIN
[2] Departamento Didáctica de la Matemática, Universidad de Valencia, Valencia, SPAIN
[*] Corresponding author: m.teresa.sanz@uv.es

Citation: Santágueda-Villanueva, M., Sanz, M. T., López-Iñesta, E., Monferrer, L., & Lorenzo-Valentín, G. (2026). Design and validation of the P-MATHEX questionnaire: Assessing preservice teachers' perceptions of pedagogical knowledge in mathematics and science. *European Journal of Science and Mathematics Education, 14*(1), 1-27. https://doi.org/10.30935/scimath/17522

ARTICLE INFO

Received: 3 May 2025

Accepted: 16 Nov 2025

ABSTRACT

This study presents the design and validation of the P-MATHEX questionnaire, an instrument aimed at assessing preservice primary teachers' perceptions regarding pedagogical knowledge in mathematics and experimental science education. Initially, a questionnaire comprising 50 Likert-scale items and 10 demographic questions was developed and subjected to a cognitive pre-test with 50 preservice teachers from a Spanish university. The instrument was then validated using a larger training sample of 264 participants. Psychometric validation involved exploratory factor analysis (EFA) and confirmatory factor analysis, item discrimination indices, and internal consistency reliability assessed by McDonald's omega coefficient. EFA results identified five distinct factors: teacher training improvement, methodological shortcomings, applied teaching methods, trial-and-error approach, and student problem-solving difficulties. The revised questionnaire was further validated with an additional sample of 166 preservice teachers. Results confirmed excellent internal consistency (global ω = 0.95) and robust factorial structure (CFI = 0.986; TLI = 0.985; RMSEA = 0.061). The P-MATHEX questionnaire effectively highlighted preservice teachers' perceived strengths in practical teaching methods and identified areas requiring further training, particularly inclusive education practices and integration of information and communication technologies skills. This validated instrument provides educators and policymakers a reliable and actionable tool to evaluate and enhance teacher education programmes in mathematics and experimental sciences, ultimately aiming to improve pedagogical practices in primary education classrooms.

Keywords: preservice teachers, mathematics education, science education, perceptions of pedagogical knowledge, questionnaire validation

INTRODUCTION

The development of mathematical competence in preservice teachers is a key factor for ensuring high-quality instruction in primary education classrooms. Over the years, it has become clear that many future teachers arrive at university with significant gaps in their mathematical knowledge, which may affect their confidence and ability to teach mathematics effectively. As highlighted by recent studies, many of these students have not taken mathematics courses in the final years of secondary education, further aggravating this situation (Nortes-Martínez-Artero & Nortes-Checa, 2018).

Since the implementation of the European higher education area (EHEA), university training has focused on the development of competences that prepare future teachers to face the challenges of their profession. In this sense, possessing a solid mastery of mathematics is essential not only for effective teaching but also for fostering critical and autonomous decision-making. However, many university students who are preparing to become teachers arrive with preconceived and limited ideas about how to teach mathematics, often focusing on isolated procedures and far removed from the problem-solving required in everyday life.

Recent research has demonstrated that teachers' mathematical knowledge is directly related to student achievement. For example, Hill et al. (2005) found that teachers' mathematical knowledge for teaching significantly influences student learning gains in the early years of primary education. In addition, studies such as Kleickmann et al. (2013) highlight the importance of structural differences in teacher education and how these affect the development of content knowledge and pedagogical content knowledge (PCK).

However, while numerous studies have explored preservice teachers' mathematical knowledge and classroom experiences, few have developed valid and reliable instruments that capture their perceptions of pedagogical competence across different subject areas. In particular, there is a clear need for tools that jointly assess perceptions of didactic knowledge in mathematics and experimental science, as these disciplines share common methodological challenges but are often studied in isolation. Recent contributions have addressed this gap by designing and validating instruments that explore teachers' beliefs (Angel-Cuervo et al., 2024), PCK (Vargas et al., 2024), and teaching competencies in STEAM education (Kim & Kim, 2016). Furthermore, Henríquez-Rivas and Vergara-Gómez (2025) have proposed a questionnaire to characterise the geometric work of mathematics teachers, offering a theoretical framework grounded in the concept of mathematical workspaces.

To address this gap, the present study introduces and validates the P-MATHEX questionnaire, a new instrument specifically designed to evaluate preservice teachers' perceptions of their pedagogical knowledge in mathematics and science. The main objectives of this study are:

- To construct and refine a theoretically grounded questionnaire based on prior research and expert review.
- To examine its psychometric properties through exploratory factor analysis (EFA) and confirmatory factor analysis (CFA).
- To assess the internal consistency, item discrimination, and dimensionality of the instrument.
- To validate the revised structure (P-MATHEX 2.0) and evaluate its measurement precision using item response theory (IRT).

By integrating these methodological approaches, this research provides a robust, multidimensional measure that captures preservice teachers' self-perceptions of didactic competence across content domains. The P-MATHEX represents an innovative contribution to the field, offering teacher educators and policymakers a validated tool to diagnose training needs and enhance the quality of teacher education programmes.

THEORETICAL FRAMEWORK

Mathematical Competence in Preservice Teachers

Many institutions and research bodies have highlighted the significant gaps in basic mathematical understanding among primary education preservice teachers. For example, the Spanish Society for Research in Mathematics Education (2014) reported widespread difficulties in fundamental mathematical concepts at the primary level. A key contributing factor is the heterogeneous mathematical background of students entering primary teacher education degrees, with many not having studied mathematics during their last years of secondary education (Nortes-Martínez-Artero & Nortes-Checa, 2018).

Since the Bologna declaration (European Ministers of Education, 1999), university teaching has increasingly emphasised the development of competences. Teacher educators must promote a set of professional competences enabling future teachers to respond to diverse classroom situations. A strong foundation in mathematical knowledge is crucial, as it facilitates critical and autonomous decision-making and supports the ongoing professional development of mathematics teachers.

Research demonstrates that teachers' capacity to identify key elements in teaching and learning situations and to analysed them from a professional perspective is essential for high-quality instruction (Godino et al., 2007; Jacobs et al., 2010; Mason, 2016). Regardless of their previous experience, all preservice teachers must complete mathematics and mathematics education courses during their university studies and are ultimately responsible for teaching mathematics in schools.

However, many university students still have a rigid and partial image of mathematical knowledge and teaching, largely acquired from their experiences as pupils. Their ideas tend to focus on procedural skills disconnected from the application of mathematics to real-world problems, resulting in frequent misunderstandings. Therefore, innovative methodological proposals are necessary to challenge these limited views and prepare future teachers more coherently.

Numerous studies have identified gaps in the mathematical and PCK of primary education preservice teachers, both in Spain and internationally. For example, Hill et al. (2005) found that teachers' mathematical knowledge for teaching is strongly linked to student achievement in the early grades. Kleickmann et al. (2013) emphasise the importance of both content knowledge and PCK, noting how variations in teacher education structure can affect these competencies.

There are also documented difficulties in conceptual areas such as geometry, where preservice teachers often depend on prototypical images and struggle to define necessary and sufficient conditions (Codes et al., 2019; Montes Navarro et al., 2022; Ulusoy, 2021). Moreover, recent systematic reviews have revealed that primary school teachers possess limited knowledge for teaching statistics and probability, often due to insufficient training in these subjects, which in turn leads to insecurity in the classroom (Franco & Alsina, 2022). To address these gaps, scholars recommend reinforcing teacher education programmes with practical projects, technological resources, and context-based learning experiences (Franco & Alsina, 2022).

In summary, strengthening mathematical competence in preservice teachers requires not only an emphasis on content knowledge but also a reflective, research-based approach to didactics and methodology–one that prepares future teachers to foster meaningful mathematics learning in diverse classroom contexts.

The Role of the Practicum in Teacher Education

Numerous studies highlight the importance of the practicum in university teacher education degrees (Bretones, 2013; Mendoza et al., 2020). This type of analysis has also been extended to other educational levels, such as secondary and higher education (González et al., 2020; León-Urquijo et al., 2018; Méndez & Magaña, 2019). However, most of these studies focus on aspects different from those addressed in this work, such as student competences and the roles of different agents involved in the process (Poveda et al., 2021). The practicum is generally valued as a formative stage and is increasingly demanded to play a prominent role in teacher education curricula (Martínez-Izaguirre et al., 2019).

Other studies investigate the learning process of students at the university and in preschool and primary classrooms, highlighting the substantial gap between both institutions (Cortés-González et al., 2020). Furthermore, the impact of practicum experiences on educational agents and institutions involved has been explored (Colén & Castro, 2017; Gairín-Sallán et al., 2019), as well as ways to use these experiences as mechanisms for change and learning (Díez-Fernández et al., 2018).

A considerable number of publications focus on the construction of professional teacher identity during these practices (González et al., 2019), and on the improvement of professional competence (Fierro & Majós, 2016; Sanmamed & Abeledo, 2011), aiming to mitigate the challenges this period may also generate (Larrañaga, 2012).

Recent research confirms once again the relevance of the practicum in future teachers' training. For instance, Verger-Gelabert et al. (2023) observed not only a strong similarity between the practicum guidelines of the analysed universities, but also that this period usually amounts to 43 ECTS credits, most commonly starting from the fifth semester, i.e., the beginning of the third academic year. The comparatively low practical component versus theoretical courses, and its placement at the end of the degree, further accentuate the traditional divide between theory and practice, i.e., between academic knowledge imparted at the university and professional knowledge used by teachers in schools. This gap is one of the main challenges in initial teacher education. Overcoming this dichotomy undoubtedly requires fostering and advancing collaboration between schools and universities (Onrubia et al., 2020).

As highlighted by Verger-Gelabert et al. (2023), it is essential to train practicum tutors–both in schools and at the university–within a constructive accompaniment model in which reflective practice becomes a key component to bridge the gap between theory and practice through inquiry-based discourse. This approach encourages bidirectional learning (Trumbull & Fluet, 2008). Björck and Johansson (2018) also emphasise the need for more horizontal relationships that overcome some of the collaborative limitations imposed by institutional boundaries.

Innovative practicum programmes, such as those recently implemented at the Universitat Jaume I (UJI) in Castellón, Spain, exemplify these new trends. The redesign of the primary and preschool teacher education degrees at UJI included an alternating practicum distributed throughout the second, third, and fourth years, allowing students to engage in weekly reflective seminars at the university while undertaking placements in schools of different types–rural, highly diverse, and mainstream (Gil & Martí, 2024).

Continuous feedback and guidance are provided through these seminars, especially for students who have expressed difficulties in mathematical content and pedagogy, often observed when collaborating with their mentors at the placement school. Recent research shows that both in-service teachers and student teachers often do not perceive mathematics as inherently difficult to teach or learn. The most commonly proposed strategies to address classroom difficulties and perceived training gaps are "presenting problem situations with increasing difficulty" and "re-explaining content" (Santágueda-Villanueva & Lorenzo-Valentín, 2024).

In light of these findings, this work aims to analyse how preservice primary teachers develop their didactic perceptions and competences through both coursework and practicum experiences, and how the structure and quality of practicum placements can shape their professional growth.

Perceptions and Attitudes Towards Mathematics and Science Teaching

Preservice teachers' perceptions and attitudes towards mathematics and science teaching play a decisive role in shaping their future classroom practice and their students' learning experiences. Research indicates that these attitudes are often influenced by previous educational experiences, leading to the persistence of rigid and partial conceptions of mathematics and its teaching. Many preservice teachers report low self-confidence in their mathematical abilities, which can negatively impact their willingness to engage in innovative practices or to address students' difficulties (García et al., 2014).

Attitudes toward mathematics are multidimensional and include affective, cognitive, and behavioural components (Di Martino & Zan, 2010). The literature suggests that positive attitudes are associated with greater willingness to use active methodologies, while negative attitudes can reinforce the reliance on traditional and procedural approaches (Philipp, 2007). Moreover, teachers' beliefs about the nature of mathematics–whether they view it as a set of fixed procedures or as a creative, problem-solving discipline–influence their instructional decisions (Wilkie, 2016).

International studies highlight the importance of developing both content knowledge and PCK (Kleickmann et al., 2013; Shulman, 1986), as these dimensions are closely connected to teachers' self-efficacy and professional identity (Fives & Buehl, 2012). For science education, similar trends are observed: preservice teachers who lack confidence in their understanding of scientific concepts are less likely to adopt inquiry-based strategies (Franco & Alsina, 2022; Kind & Kind, 2007).

Recent research emphasises that supporting preservice teachers in reflecting critically on their attitudes and conceptions is essential for professional growth (Wilkie, 2016). Programmes that provide opportunities for collaborative reflection, classroom observation, and hands-on activities have proven effective in fostering more positive and open attitudes towards mathematics and science teaching (Philipp, 2007; Wilkie, 2016).

Ultimately, recognising and addressing these perceptions is a necessary step toward enhancing the quality of mathematics and science education, as preservice teachers' attitudes can significantly shape their future students' learning opportunities and attitudes toward these disciplines.

A central component underpinning preservice teachers' perceptions and instructional decisions is their PCK–that is, the blend of content and pedagogical expertise needed to teach specific subjects effectively (Shulman, 1986). Recent studies have emphasised that the development of PCK is crucial for preservice teachers to design, implement, and adapt instructional strategies that foster meaningful learning in mathematics and science classrooms (Depaepe et al., 2013; Kleickmann et al., 2013). Inadequate opportunities to build PCK during initial teacher education can contribute to insecurities and reinforce traditional, less effective approaches (Depaepe et al., 2013). Consequently, evaluating and supporting preservice teachers' perceptions of their own didactic knowledge is fundamental for ensuring their readiness to engage with diverse learners and contemporary curricular demands. As such, instruments that assess these perceptions provide valuable insights for both teacher educators and policy-makers seeking to improve the quality of mathematics and science education at the primary level.

MATERIALS AND METHODS

Participants

This study followed a cross-sectional, quantitative, instrumental research design aimed at developing and validating a new questionnaire for assessing preservice teachers' perceptions of their didactic knowledge in mathematics and science. The validation process included two independent samples for EFA and CFA.

The initial sample used for instrument development and exploratory analyses comprised 264 preservice teachers (215 women, 49 men) enrolled in primary or early childhood education degrees at a Spanish university. All participants were undertaking mandatory practicum courses at the time of the study: practicum I (57.41%) was conducted in the second year (with no prior didactics coursework), practicum II in the third year (15.56%, after an introductory didactics course), and practicum III in the fourth year (with complete didactic training).

For the confirmatory phase, a separate cohort of 166 preservice teachers was recruited (80 in early childhood education, 86 in primary education), ensuring independent validation. The total sample, combining both phases, consisted of 430 participants. Sampling followed psychometric guidelines recommending 5-10 respondents per item for factor analysis, well exceeding the recommended minimum for the 50-item scale (Gorsuch, 1983; Tabachnick & Fidell, 2007).

Sociodemographic characteristics and practicum distribution were analysed for potential bias. Kruskal-Wallis tests found significant but negligible effect sizes for year, practicum type, and gender ($\varepsilon^2 < 0.09$), so no stratification was applied in further analyses. The gender imbalance observed in both samples, with a clear predominance of female students, reflects the typical demographic distribution of early childhood and primary education degrees in Spain and across many European countries. According to recent Eurydice data, women represent the vast majority of students enrolled in initial teacher education programmes (European Commission/EACEA/Eurydice, 2021). Therefore, the gender composition of the present sample is consistent with the population characteristics of these degrees rather than a selection bias.

Instrument

The P-MATHEX questionnaire consists of 50 five-point Likert items, structured in two thematic blocks (mathematics: I1-I31; science: I32-I50), plus 10 sociodemographic items. Item content covers practical experience, didactic strategies, perceived student difficulties, and areas for improvement in teacher training, based on prior literature and expert review (see **Appendix A** for the full questionnaire).

Items were grouped by both subject area and didactic category (e.g., observed methods, methods applied by trainees, problem-solving strategies, perceived student difficulties, and required training improvements), as justified by educational theory and curriculum standards (Driver et al., 1994; European Commission, 2018; Kilpatrick et al., 2001;).

Instrument construction followed established guidelines for test development, including item generation from the literature, expert review, cognitive pretesting (n = 50), and iterative revision. The final instrument was administered electronically during regular university sessions. All procedures were approved by the University's ethics committee, and informed consent was obtained from all participants.

Data Analysis

All data analyses were conducted using R (v. 4.4.1), following a rigorous and sequential psychometric validation protocol. Analyses included descriptive statistics, item response distributions, item discrimination indices (Dk), reliability coefficients, dimensionality exploration, and model-based validation.

Descriptive statistics (means [Ms], standard deviations [SDs], and frequency distributions) were calculated for all questionnaire items and sociodemographic variables. Subgroup comparisons by sociodemographic factors were performed using ANOVA for normally distributed variables, or the Kruskal-Wallis test for non-normal distributions. Statistical significance was set at $p < .05$.

To evaluate item quality, classical item analysis was conducted. Item Dk were calculated using Kelley's (1939) formula, comparing the upper and lower 27% of the total score distribution. Items with Dk > 0.30 were considered good discriminators, those between 0.10-0.30 as moderate, and Dk < 0.10 as low (Ebel & Frisbie, 1991; Hair et al., 2018). Item-total correlations were also computed to assess internal consistency.

Internal consistency was estimated using both Cronbach's alpha (α) and McDonald's omega (ω) coefficients. Alpha is the most widely used indicator in social sciences, but ω is increasingly recommended due to its less restrictive assumptions about tau-equivalence (Cho, 2016; Dunn et al., 2014). Reliability was interpreted according to Nunnally and Bernstein's (1994) benchmarks (≥ 0.70 acceptable; ≥ 0.80 good; ≥ 0.90 excellent).

The dimensionality of the P-MATHEX was explored using EFA with the minimum residuals (MinRes) extraction method, which is robust in the presence of non-normal data distributions (Forero et al., 2009). Factor rotation was performed using the Quartimax orthogonal method, as inter-factor correlations were weak ($r < 0.2$). The number of factors to remain was determined by parallel analysis, which compares observed eigenvalues to those expected by chance (Horn, 1965). Sampling adequacy was checked via the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity.

The factorial structure suggested by EFA was tested using CFA, implemented with robust weighted least squares (WLSMV) estimators appropriate for ordinal Likert-type data (Brown, 2015). Model fit was assessed via multiple indices:

- Comparative fit index (CFI) and Tucker-Lewis index (TLI): Good fit was defined as CFI/TLI ≥ 0.90; 0.80-0.89 as moderate (Hu & Bentler, 1999).
- Root mean square error of approximation (RMSEA): ≤ 0.05 excellent, 0.05-0.08 acceptable, 0.08-0.10 marginal, > 0.10 poor (MacCallum et al., 1996).
- Standardised root mean square residual (SRMR): ≤ 0.08 desirable, 0.08-0.10 marginally acceptable (Hu & Bentler, 1999). Model selection was guided by both empirical fit and theoretical interpretability.

To evaluate the test's measurement precision across the latent trait continuum, item and test information functions (TIFs) were estimated using IRT, specifically the graded response model for polytomous items (Samejima, 1969). TIF curves were computed for each subscale, indicating the reliability and sensitivity of each factor along the ability spectrum.

All statistical decisions and interpretations were based on current psychometric guidelines (American Educational Research Association et al., 2014; Brown, 2015; Hair et al., 2018), ensuring methodological robustness and replicability.

RESULTS

Descriptive and Preliminary Analyses

A careful conceptual review of the P-MATHEX questionnaire revealed that its 50 items, each rated on a five-point Likert scale, can be meaningfully classified into two broad categories based on their content and the directionality of their scores:

(1) favourable attitudes and effective teaching practices and

(2) perceived barriers, difficulties, and areas for improvement in didactic training.

On the one hand, fifteen items are associated with positive aspects of didactic competence, motivation, and classroom practice. Specifically, items such as I19 (enjoyment in teaching mathematics), I50 (enjoyment in teaching science), and items reflecting effective didactic practices (I4, I5, I6, I8, I9, I10, I20, I34, I35, I36, I37, I39, I40, and I41) indicate favourable attitudes towards the teaching and learning of mathematics and science, the use of active methodologies, and a high degree of professional satisfaction.

On the other hand, the remaining items capture perceived difficulties in teaching, barriers to effective instruction, and shortcomings in didactic training. This group includes items such as I1 and I2 (perceived difficulty in teaching mathematics), I32 and I33 (perceived difficulty in teaching science), items related to the implementation of less effective teaching strategies (I3, I7, I34, and I38), and items reflecting challenges observed in students (I21-I31). Furthermore, items I11-I18 and I42-I49 point to areas for improvement in didactic training for mathematics and science, respectively.

This thematic separation is not arbitrary; it is grounded both in conceptual and empirical criteria. According to Messick's (1995) theory of construct validity, a measurement instrument must be capable of distinguishing between clear and theoretically justified dimensions of the phenomenon under study. The classification described here ensures that the items grouped together reflect specific and distinguishable constructs, such as positive attitudes towards teaching and perceived barriers or difficulties in the teaching and learning process.

From a psychometric perspective, Nunnally (1978) emphasised that questionnaires should contain both positively and negatively worded items to capture the full range of attitudes and to minimise response bias. The present categorisation strictly follows this recommendation, balancing positive and negative items. Moreover, the tripartite model of attitudes (Rosenberg & Hovland, 1960)–which distinguishes cognitive, affective, and behavioural components–supports this item structure. Favourable items in P-MATHEX reflect positive emotions and behaviours associated with teaching, while unfavourable items capture beliefs and experiences that may impede effective practice.

From the educational research literature, this dual structure is also consistent with Shulman's (1987) framework on PCK, which posits that effective teacher preparation involves not only subject-matter competence but also the ability to identify and overcome obstacles in classroom practice. Favourable items in the P-MATHEX, therefore, highlight strengths and successful practices, whereas negative items signal potential gaps and areas needing targeted professional development. Previous validation studies of similar instruments (e.g., Osborne & Dillon, 2008) have also distinguished between positive and negative item clusters, further justifying the approach used here.

The conceptual review and initial descriptive analysis thus provide a robust foundation for the psychometric evaluation and inferential analyses that follow. By clearly distinguishing between items reflecting favourable attitudes and those highlighting barriers, the P-MATHEX questionnaire ensures both comprehensive coverage of the relevant constructs and interpretability of the results. This structure also enables the design of more effective pedagogical interventions, tailored to the specific strengths and needs identified in the sample.

Moreover, the classification adopted in this study adheres to established guidelines for instrument construction and validation, enhancing the instrument's construct validity and ensuring that subsequent reliability and factor analyses are meaningful and theoretically grounded.

P-MATHEX Training Results

The initial training sample consisted of 264 preservice teachers (81.4% women), with the majority enrolled in the primary education programme and most in their first practicum. Binary questions S10 and S11 showed a predominance of affirmative and negative responses, respectively, providing a stable baseline for inferential comparisons.

Inferential analysis

An inferential analysis was carried out to explore whether sociodemographic variables such as year, gender, degree, practicum stage, and other academic background factors influenced the responses to the P-MATHEX questionnaire. The total training sample consisted of 264 future teachers, primarily women, mainly enrolled in the primary education degree programme, and with most respondents participating in practicum I. Two binary variables relating to prior academic experiences also displayed homogeneous patterns, with the majority of participants selecting the same response.

Given that preliminary tests indicated a clear violation of normality in the data distribution, the Kruskal-Wallis test was selected as a robust non-parametric alternative to one-way ANOVA.

The results of these analyses indicated that there were no systematic or substantial differences in the responses to the P-MATHEX items across the main sociodemographic groups considered. Although isolated statistically significant differences were observed for a few specific items, these effects were not consistent or sufficiently robust to justify constructing separate models or factor structures by subgroup. Rather, the overall pattern of responses suggested a high degree of homogeneity in the perceived didactic competencies and challenges among the future teachers in the sample, regardless of their year, gender, degree pathway, or practicum phase.

This lack of group effects supports the methodological decision to treat the training sample as a unified whole for the purposes of subsequent reliability estimation and factor analysis. Such an approach not only maximises statistical power but also aligns with the principle of parsimony in psychometric validation, especially when differential item functioning is not supported by empirical evidence. Accordingly, prior to proceeding to reliability analysis and factor extraction, negatively worded items were reverse-scored, following established recommendations to ensure proper interpretability of the instrument's scales and avoid artifacts in the measurement of internal consistency (Kline, 2015).

Reliability and validity

For the overall questionnaire, the estimated $\omega = 0.861$, with an M of 2.80 and an SD of 0.336. The inter-item correlation matrix showed that all items had maximum correlations below 0.484, suggesting that none should be eliminated due to redundancy or low discriminatory power.

The items were thematically organised, with mathematics (I1-I31) and science (I32-I50) subscales analysed separately. For mathematics, $\omega = 0.814$ ($M = 2.92$, $SD = 0.385$); for science, $\omega = 0.761$ ($M = 2.67$, $SD = 0.386$). Both values indicate adequate reliability, with the mathematics block showing higher internal consistency. These findings support the conceptual structure of the instrument and justify further factorial analysis.

From a construct validity perspective, the conceptual grouping of items into "favourable attitudes and effective practices" versus "perceived barriers and difficulties" was empirically supported by the structure of the inter-item correlation matrix. This distinction is consistent with Messick's (1995) framework for construct validity and aligns with the theoretical expectations drawn from the literature. The inclusion of both positively and negatively valanced items, as recommended by Nunnally (1978), allows for a comprehensive measurement of the latent constructs and reduces potential response bias. Furthermore, the instrument structure echoes the multidimensional attitude model of Rosenberg and Hovland (1960), encompassing cognitive, affective, and behavioural dimensions, as well as the pedagogical knowledge components highlighted by Shulman (1987).

Figure 1 displays the heatmap of the item correlation matrix for the P-MATHEX questionnaire, providing a visual summary of the internal relationships between items. The heatmap reveals several distinct clusters of significant positive correlations, suggesting the existence of latent factors that structure the responses. Specifically, blocks of strong correlation are evident among items associated with favourable attitudes (e.g.,

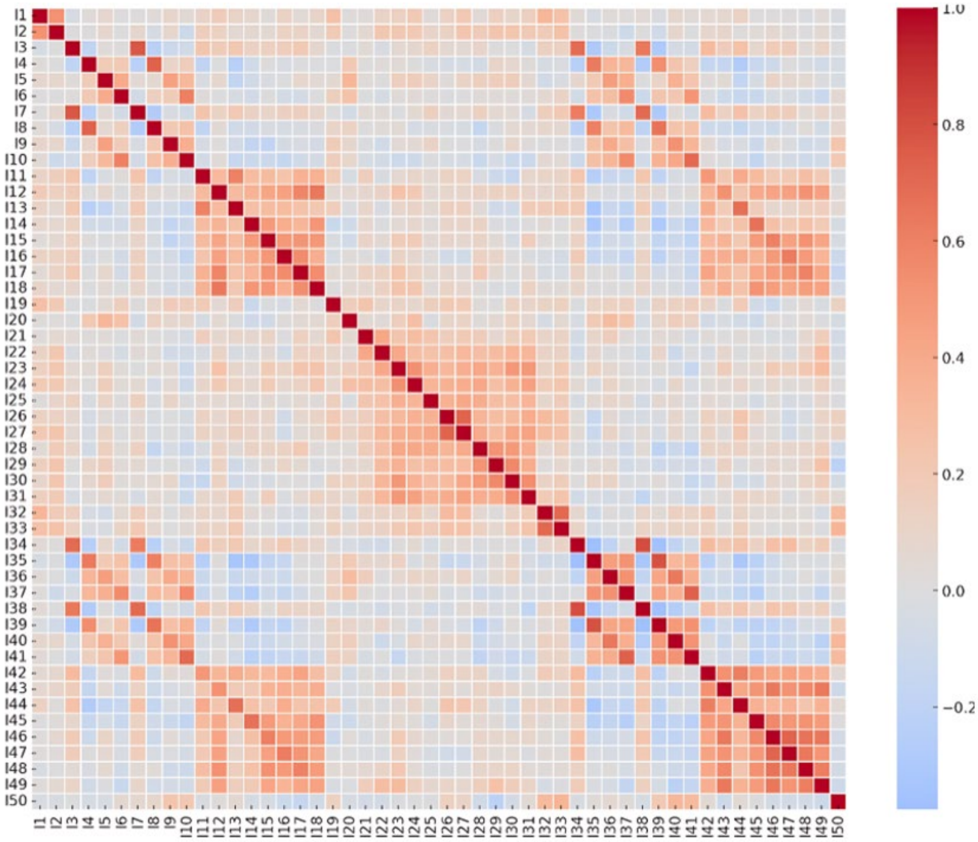

Figure 1. Heatmap of the item correlation matrix for the P-MATHEX questionnaire (Figure created by the authors)

I19, I50, I4-I6, I8-I10, I20, I34-I37, and I39-I41) and among those reflecting difficulties or didactic barriers (e.g., I1, I2, I32, I33, I3, I7, I21-I31, I11-I18, I42-I49, I34, and I38). These clusters correspond to the previously established theoretical categories and support the interpretation that the instrument effectively differentiates between key constructs.

A more detailed inspection of the **Figure 1** shows, for example, that items I1-I2 and I32-I33 form two clear blocks related to perceived difficulty in mathematics and science, respectively, while items I4-I6 and I35-I37 group together as "observed classroom methods." Similar clustering is seen for the "methods applied by practicum students" (I8-I10 and I39-I41) and for "improvements in teacher training" (I11-I18 and I42-I49). The "student difficulties" block (I21-I31) also forms a distinct cluster. Conversely, items such as I19, I20, and I50 display lower or null correlations with other items, suggesting conceptual independence or the potential for item elimination in future revisions.

The heterogeneous structure observed in the correlation matrix, with well-defined clusters and some isolated items, signals the likely presence of multiple underlying factors, thus supporting the need for EFA to further investigate the factorial architecture of the questionnaire. The separation of mathematics and science items into similar blocks also suggests that, despite theoretical recommendations for area-specific differentiation, the empirical structure may justify a joint factor model.

Discrimination indices

A crucial step in evaluating the psychometric properties of the P-MATHEX questionnaire is the analysis of item Dk, which measure the ability of each item to distinguish between respondents with high and low levels of the latent trait–here, perceived didactic competence.

The analysis revealed that 12% of the P-MATHEX items achieved good Dk (> 0.30), indicating that these items are highly effective at differentiating among participants with varying levels of perceived competence. The majority, 64%, showed moderate indices (0.10 ≤ Dk ≤ 0.30), while the remaining 24% had low discrimination values (Dk < 0.10), suggesting a need for further revision of some items.

Table 1. Summary of EFA results for the P-MATHEX questionnaire

	KMO	Bartlett's test χ^2 (df), p	Factors	Total variance	TLI	BIC	RMSEA	90% CI RMSEA
Global	0.708	4,552 (1,225), p < 0.001	6	47.9%	0.668	-2,754	0.0783	(0.0730, 0.0848)
Mathematics (I1-I31)	0.721	2,604 (465), p < 0.001	4	37.0%	0.678	-853	0.0828	(0.0755, 0.0909)
Science (I32-I50)	0.736	1,713(171), p < 0.001	3	46.4%	0.717	-137	0.1240	(0.1090, 0.1400)

Items I1 (Dk = 0.011), I2 (Dk = 0.036), I4 (Dk = 0.031), I8 (Dk = 0.027), and I14 (Dk = 0.038) exhibited particularly low discrimination and weak item-total correlations, which, as discussed by Tabachnick and Fidell (2019), often correspond to items with limited capacity to capture individual differences and may warrant rewording or exclusion in future iterations.

Disaggregated analysis by domain provided additional insight. Among mathematics-related items (I1-I31), 35.5% demonstrated high discrimination, 54.8% were moderate, and only 9.7% were low, underscoring that the mathematics subscale of the instrument is particularly effective at capturing the spectrum of perceived competence among respondents. The items I4, I8, and I10 stood out as those with lower Dk and correlations, meriting careful review.

Within the science-related items (I32-I50), 26.3% had high discrimination, 68.4% moderate, and just 5.3% low (notably, item I38). This distribution reflects an overall satisfactory functioning of the science subscale, although, as in mathematics, a few items with weak indices may benefit from future refinement.

Exploratory factor analysis

To further examine the underlying structure of the P-MATHEX questionnaire, an EFA was conducted. **Table 1** summarises the global and domain-specific indices, confirming the suitability of the data for factor analysis and providing an overview of sampling adequacy, model fit, and variance explained.

The global EFA yielded six principal factors with eigenvalues greater than 1, together accounting for 47.9% of the total variance. These factors are reflected:

(1) improvement in teacher training (I11-I18 and I42-I49),

(2) student difficulties in mathematics (I2 and I21-I31),

(3) applied methodologies in the classroom (I5, I6, I9, I10, I36, I37, I40, and I41),

(4) trial-and-error approaches (I4, I8, I35, and I39),

(5) deficiencies in teaching methods (I3, I7, I34, and I38), and

(6) motivation for mathematics despite perceived difficulty (I1 and I19).

A seventh factor associated with perceived difficulty and motivation in science (I32, I33, and I50) emerged (eigenvalue 0.9291) but did not meet retention criteria based on parallel analysis. **Figure 1** further illustrates the presence of coherent factor blocks, with clusters of items showing strong internal correlations consistent with the theoretical structure of the instrument.

To test the stability of the factor structure across content areas, separate EFAs were conducted for mathematics (I1-I31) and science (I32-I50). In mathematics, four factors were identified–student difficulties, training improvement, applied methodologies, and methodological deficiencies–explaining 37% of the variance. In science, three analogous factors emerged, explaining 46.4% of the variance. Items related to perceived difficulty and motivation (I1, I2, I19, I32, I33, and I50) showed weaker loadings and cross-loadings in both domains, suggesting partial conceptual independence and highlighting candidates for refinement in future iterations of the instrument.

These results empirically substantiate the multidimensionality of the P-MATHEX questionnaire, confirming that it robustly captures distinct facets of didactic competence and perceived barriers among preservice teachers. The emergence of consistent factors across domains, together with the convergence of blocks in the correlation matrix, strengthens the theoretical coherence of the instrument. At the same time, the identification of items with low loadings or problematic cross-loadings provides concrete guidance for further refinement.

In summary, the EFA demonstrates that the P-MATHEX structure is largely congruent with its conceptual foundation, supporting the validity of subsequent CFA and practical application in teacher education contexts.

Table 2. Statistics of CFA of each model

	Theoretical model	χ^2	g.l.	p-value	CFI	TLI	RMSEA	SRMR	90% CI RMSEA
Together	M1	801	169	< 0.001	0.704	0.668	0.1520	0.1050	(0.142, 0.163)
	M2	1527	458	< 0.001	0.688	0.662	0.1220	0.0950	(0.116, 0.129)
	M3	1825	650	< 0.001	0.677	0.651	0.1090	0.0910	(0.103, 0.115)
	M4	2,533	1,106	< 0.001	0.647	0.624	0.0960	0.0940	(0.091, 0101)
Mathematics	M1	173	34	< 0.001	0.860	0.814	0.1250	0.0694	(0.107, 0.143)
	M2	339	98	< 0.001	0.839	0.803	0.0967	0.0692	0.086, 0.108)
	M3	408	138	< 0.001	0.834	0.795	0.0655	0.0861	(0.077, 0.096)
	M4	930	395	< 0.001	0.794	0.767	0.0694	0.0732	(0.067, 0.079)
Experimental science	M1	168	34	< 0.001	0.856	0.809	0.1280	0.0780	(0.109, 0.148)
	M2	361	98	< 0.001	0.826	0.786	0.1280	0.0810	(0.114, 0.142)
	M3	403	138	< 0.001	0.837	0.798	0.1090	0.0760	(0.097, 0.121)
Bifactorial	M1	2,164	609	< 0.001	0.566	0.526	0.1290	0.1160	(0.123, 0.135)
	M2	1,736	597	< 0.001	0.682	0.645	0.1120	0.1090	(0.106, 0.118)
	M3	1,592	612	< 0.001	0.731	0.691	0.1020	0.0720	(0.096, 0.108)
	M4	2,324	1,067	< 0.001	0.689	0.657	0.0920	0.0830	(0.087, 0.097)

Confirmatory factor analysis

A series of CFA were conducted to test the factorial structure of the P-MATHEX instrument, using robust estimation procedures and established psychometric criteria. All models were evaluated according to best practices for structural equation modelling, with particular attention to fit indices beyond the traditional chi-square statistics, which is known to be overly sensitive in moderate and large samples (Kline, 2015).

The CFA models were informed by both conceptual considerations and prior EFA results. Two factors were especially robust across analyses: teacher training improvement (I11-I18 and I42-I49) and deficiencies in teaching methods (I3, I7, I34, and I38). The factor applied methodologies in the classroom (I4-I6, I8-I10, I35-I37, and I39-I41) sometimes included items related to the trial-and-error approach (I4, I8, I35, I39), especially when analysed by subject area. The factor student difficulties in problem-solving (I21-I31) emerged as specific to mathematics.

Some items–namely those related to perceived difficulty (I1, I2, I32, and I33) and teaching motivation (I19 and I50)–consistently showed weak or cross-factor loadings, in line with psychometric recommendations to consider their removal if further scale refinement is pursued. Item I20, which assesses problem-solving in real-life contexts, demonstrated weak associations with its theoretical factor, suggesting potential conceptual independence or the need for exclusion.

Four theoretically-driven models were tested via CFA:

- M1. Teacher training improvement, deficiencies in teaching methods
- M2. M1 + applied methodologies in the classroom, trial-and-error
- M3. M2 + perceived difficulty, teaching motivation
- M4. M3 + student difficulties in problem-solving

Table 2 presents the full set of fit statistics for each model across the full scale and by subject area. The results indicated that Model M3 provided the best compromise between model complexity and fit, yielding the most favourable fit indices in both global and domain-specific analyses. While not all indices reached the optimal cutoffs, they were within the ranges commonly accepted for complex multidimensional educational measures. Notably, model fit improved when mathematics and science domains were analysed separately, a result supported by bifactorial CFA models.

In the bifactorial analysis, model M3 was the most parsimonious solution (CFI = 0.731, TLI = 0.691, RMSEA = 0.102, SRMR = 0.072), with all indices in the moderately acceptable range. Model M4, which included the student difficulties in problem-solving factors, showed slightly lower CFI and TLI but better RMSEA (0.092) and SRMR (0.083), also within acceptable limits. The main distinction is theoretical: model M4's added complexity is justified in mathematics-specific validation, as the additional factor provides nuanced measurement of this critical dimension (Reise et al., 2010).

P-MATHEX 2.0

Based on the evidence obtained from the EFA and CFA, as well as the internal consistency and Dk, the final version of the questionnaire–P-MATHEX 2.0–was developed with an improved conceptual and psychometric structure.

Structure and thematic blocks

The revised questionnaire is organised into two clearly differentiated thematic blocks: mathematics and experimental sciences. Each block comprises subscales reflecting the main factors identified through factor analysis. This restructuring enhances both respondent comprehension and the theoretical coherence of the measurement instrument.

Mathematics block:

Improvement in teacher training (items I11-I18)

Shortcomings in didactic methods (I3 and I7)

Didactic methodologies applied in the classroom (I5, I6, I9, and I10)

Trial-and-error methodology (I4 and I8)

Difficulties in problem-solving (I21-I31)

Experimental sciences block:

Improvement in teacher training (I42-I49)

Shortcomings in didactic methods (I34 and I38)

Didactic methodologies applied in the classroom (I36, I37, I40, and I41)

Trial-and-error methodology (I35 and I39)

A brief introductory section explains the purpose of the questionnaire, the response scale (5-point Likert), and provides sample items. The questionnaire concludes with a socio-demographic section to capture variables such as year, gender, degree, practicum type, and binary items relevant to the educational context.

In the P-MATHEX 2.0, items are grouped according to the subfactors revealed by the best-fitting confirmatory factor model (model M4 for mathematics, model M3 for experimental sciences). The item sequence within each block follows a logical progression–from items reflecting prior teaching experience, through those related to classroom methodologies, to items capturing difficulties observed in students. This logical order is intended to reduce respondent fatigue and optimise data quality.

Given the psychometric evidence, some items with persistently low factor loadings or Dk–such as item I20 (problem-solving in real-life contexts)–are recommended for exclusion in future applications to streamline the instrument and ensure conceptual purity of the subscales.

Internal consistency

The accuracy of the P-MATHEX 2.0 scores was evaluated using total ω and α for both the global scale and each subscale. The overall ω value was 0.95, indicating excellent reliability and confirming that the set of items stably measures a common construct of didactic competence in mathematics and science.

At the subscale level, results were equally outstanding: teacher training improvement (ω = 0.97; α = 0.94), shortcomings in didactic methods (ω = 0.98; α = 0.92), applied methodologies (ω = 0.96; α = 0.88), trial-and-error methodology (ω = 0.97; α = 0.84), and difficulties in problem-solving (ω = 0.95; α = 0.92). All values comfortably exceed the recommended threshold of 0.70 for applied studies (Nunnally & Bernstein, 1994), thus ensuring the stability of the scores obtained with the sample of 166 participants.

Confirmatory factor model fit

The correlated five-factor M4 model was evaluated through CFA using the WLSMV estimator, which is suitable for ordinal data derived from Likert-type scales. The model demonstrated excellent overall fit (CFI = 0.986, TLI = 0.985), both surpassing the established threshold of 0.95 for good fit (Hu & Bentler, 1999). The RMSEA was 0.061, with a 90% confidence interval ranging from 0.055 to 0.067, well within the acceptable range (0.05-0.08). The SRMR was 0.087, which is slightly above the conventional cut-off of 0.08 but still considered acceptable in the context of instruments with a large number of items (Kline, 2015).

The chi-square statistic was significant ($\chi^2 = 1,372.10$; df = 850; $p < 0.001$), a result frequently observed in large samples. Consequently, interpretation of model fit primarily relied on the robust indices mentioned above.

Standardised factor loadings

Standardised factor loadings for the five-factor M4 model ranged from 0.53 to 0.98, with an average loading of approximately 0.79. All items made substantial contributions to their respective factors, exceeding the commonly recommended cut-off of 0.40 for practical significance (Brown, 2015). Particularly notable were the very high loadings observed for items assessing methodological shortcomings (I34 and I38 ≥ 0.98) and those reflecting improvements in teacher training (I42-I48 ≥ 0.82), confirming the strong conceptual coherence of these factors. No items exhibited problematic cross-loadings or high uniqueness that would warrant their removal, thereby supporting the retention of the full item set and the robust structure of the P-MATHEX 2.0 instrument.

Inter-factor correlations

The standardised inter-factor correlations observed in the five-factor model ranged from low to moderate ($|r| = 0.18$-0.63), providing evidence that the factors are conceptually distinct yet related. The highest correlation was found between “applied methodologies” and “trial-and-error” ($r = 0.63$), suggesting that participants who observe trial-and-error strategies in the classroom are also more likely to report the presence of active methodologies, potentially as a compensatory response. There was a moderate negative correlation between “teacher training” and “trial-and-error” ($r = -0.34$), indicating that a better perception of the training received is associated with less frequent use of trial-and-error methods, consistent with professional development theories.

Additionally, a slight negative correlation was found between “teacher training” and “applied methodologies” ($r = -0.18$), which may suggest that stronger training is perceived as somewhat opposed to the frequent need for methodological change. All remaining correlations were below $|0.20|$, supporting the relative independence of the factors, especially between “problem-solving difficulties” and the other dimensions.

Collectively, this correlation pattern confirms the discriminant validity of the instrument and supports the use of both subscale and global scores in future research and practice.

Test information curves by subscale

IRT analysis using the graded response model was conducted to examine the measurement precision of each P-MATHEX 2.0 subscale across the latent trait continuum (θ). The test information curves (TIC) (**Figure 2**) showed differentiated sensitivity profiles, indicating that each subscale provides optimal information at distinct competence levels.

A summary of peak information values, θ positions, and effective measurement ranges is presented in **Table 3**.

In summary, the information curves demonstrate that:

- “Teacher training” and “shortcomings in didactic methods” subscales are most precise for medium-high and medium levels, respectively.
- “Applied methodologies” and “trial-and-error” are more informative at below-average levels.
- “Problem-solving difficulties” discriminates best at slightly above-average levels but maintains coverage across the entire spectrum.

These patterns indicate that P-MATHEX 2.0 offers high and differentiated measurement precision across low, medium, and high levels of perceived didactic competence. This supports the robustness of the instrument for both research and diagnostic applications in teacher education.

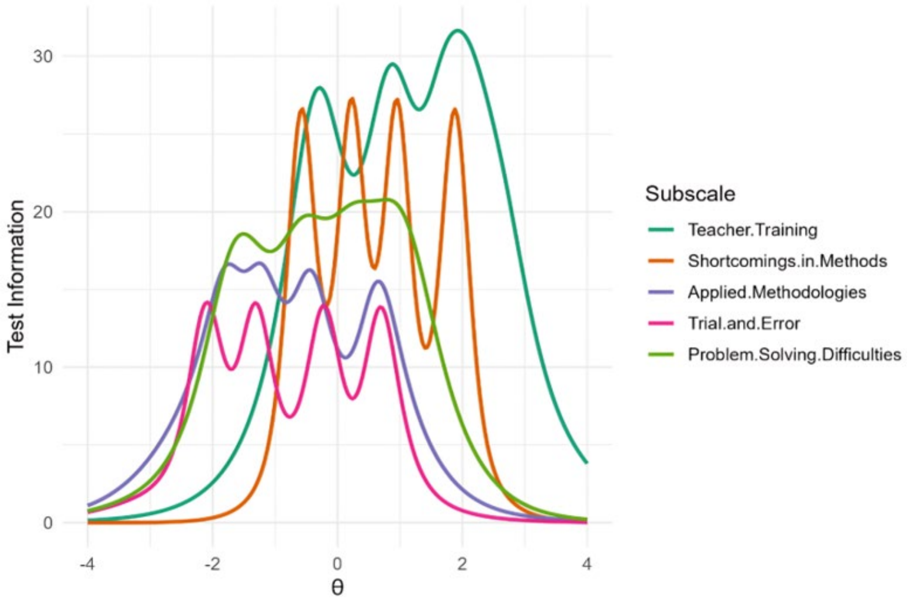

Figure 2. TIC for each P-MATHEX 2.0 subscale in the validation sample (n = 166) (Figure created by the authors)

Table 3. Descriptive statistics for P-MATHEX 2.0 factors by practicum stage

Subscale	Peak information	θ at peak	50% information range
Teacher training	31.634	+1.9	-0.0 to +2.9
Shortcomings in didactic methods	37.119	+0.2	-0.8 to -2.1
Applied methodologies	16.636	-1.3	-2.4 to +1.1
Trial-and-error method	14.169	-2.1	-2.4 to +1.0
Problem-solving difficulties	20.778	+0.8	-2.1 to +1.6

Analysis of Perceptions Regarding Didactic Knowledge in Science and Mathematics Among Preservice Teachers

The final stage of analysis combined the data from both cohorts–the training sample (n = 264) and the validation sample (n = 166)–yielding a total of 430 future teachers. This approach enabled a comprehensive assessment of perceived didactic competence across the full sample and allowed for robust comparisons by degree programme and practicum stage.

Descriptive statistics for the five P-MATHEX 2.0 factors revealed distinct patterns in the perceptions of preservice teachers (see **Figure 3**). The "teacher training improvement" factor exhibited the lowest average scores (M ≈ 1.87, SD = 0.68), suggesting that participants perceive notable deficits in their didactic training.

In contrast, "applied methodologies" (M ≈ 3.72, SD = 0.81) and "trial-and-error" (M ≈ 3.55, SD = 1.06) showed the highest values, indicating a strong confidence in the use of practical pedagogical strategies in the classroom. The "problem-solving difficulties" factor occupied an intermediate position (M ≈ 3.12, SD = 0.75), while "shortcomings in didactic methods" remained at a moderately low level (M ≈ 2.40, SD = 1.09). Overall, the profile depicted in the radar chart (**Figure 3**) reflects a clear imbalance: preservice teachers generally feel more confident in their use of classroom methodologies and trial-and-error strategies yet recognise substantial room for improvement in their formal didactic training.

Further analysis by degree programme (**Table 4**) indicated that preservice teachers specialising in early childhood education systematically reported higher M scores than those in primary education across almost all factors and both content domains. For instance, the M global score in science was 2.603 (SD = 0.525) for early childhood education and 2.539 (SD = 0.487) for primary education; in mathematics, the Ms were 2.926 (SD = 0.491) and 2.764 (SD = 0.383), respectively. Notably, early childhood education students showed higher values in "applied methodologies" and "trial-and-error," suggesting greater confidence in implementing practical teaching strategies, while both groups consistently identified "teacher training improvement" as the lowest-rated factor.

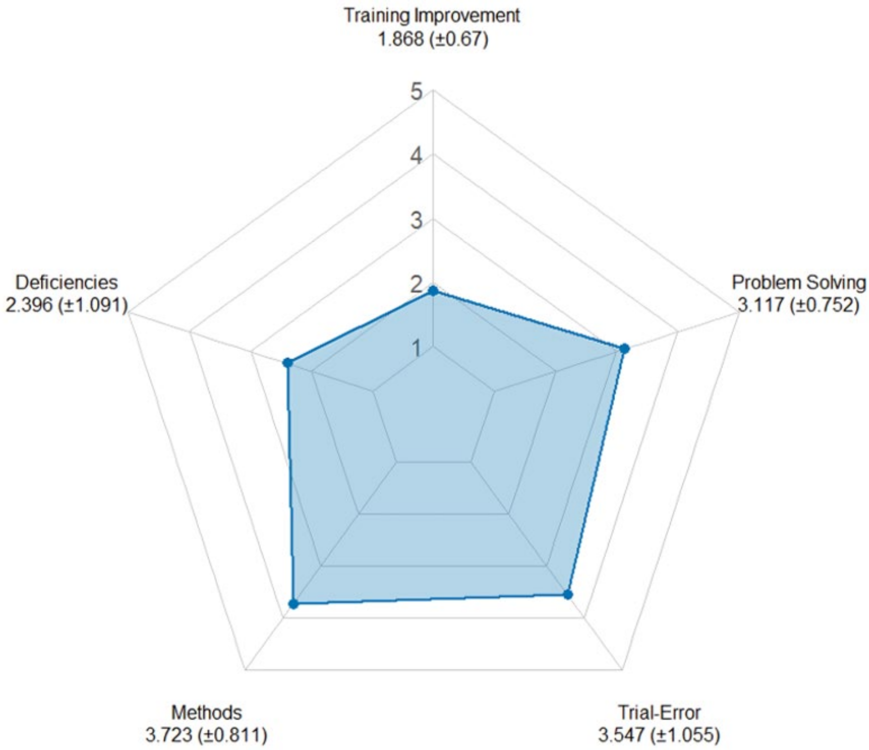

Figure 3. Radar chart illustrating the means and standard deviations of the five P-MATHEX 2.0 factors in the total sample (n = 430) (Figure created by the authors)

Table 4. Descriptive statistics for P-MATHEX 2.0 factors by degree programme

Degree programme	Domain	Factor	M (SD)
Primary education: 2.695 (±0.369)	Experimental sciences: 2.539 (±0.487)	Shortcomings	2.472 (±1.078)
		Trial-and-error	3.318 (±1.109)
		Training	1.794 (±0.636)
		Methodologies	3.699 (±0.805)
	Mathematics: 2.764 (±0.383)	Shortcomings	2.500 (±1.093)
		Problem-solving difficulties	2.990 (±0.627)
		Trial-and-error	3.477 (±1.049)
		Training	1.881 (±0.657)
		Methodologies	3.718 (±0.786)
Early childhood education: 2.826 (±0.426)	Experimental sciences: 2.603 (±0.525)	Shortcomings	2.179 (±1.088)
		Trial-and-error	3.796 (±0.986)
		Training	1.913 (±0.745)
		Methodologies	3.769 (±0.889)
	Mathematics: 2.926 (±0.491)	Shortcomings	2.221 (±1.074)
		Problem-solving difficulties	3.407 (±0.919)
		Trial-and-error	3.897 (±0.884)
		Training	1.934 (±0.686)
		Methodologies	3.737 (±0.814)

Analysis by practicum (**Table 5**) stage demonstrated remarkable stability in perceived didactic competence during the first two placements, with a slight decline in mathematics during the final placement (practicum III), particularly in the "problem-solving difficulties" dimension (M = 2.871, SD = 0.848, compared to 3.164 and 3.208 in practicum I and practicum II, respectively). This may reflect a heightened self-critical perspective as students accumulate classroom experience. In science, practicum III showed a slight increase in the overall score (M = 2.616, SD = 0.394), particularly in the "trial-and-error" dimension, relative to earlier stages.

Collectively, these findings highlight both the strengths and developmental needs perceived by preservice teachers in mathematics and science education. While confidence in classroom practice and active methodologies is high, there is a persistent self-reported gap in foundational didactic knowledge–underscoring the importance of targeted interventions in teacher training curricula.

Table 5. Descriptive statistics for P-MATHEX 2.0 factors by practicum stage

Degree programme	Domain	Factor	M (SD)
Practicum I: 2.744 (±0.372)	Experimental sciences: 2.543 (±0.53)	Shortcomings	2.348 (±1.097)
		Trial-and-error	3.397 (±1.156)
		Training	1.801 (±0.694)
		Methodologies	3.746 (±0.833)
	Mathematics: 2.827 (±0.39)	Shortcomings	2.393 (±1.104)
		Problem-solving difficulties	3.164 (±0.661)
		Trial-and-error	3.602 (±1.087)
		Training	1.858 (±0.672)
		Methodologies	3.742 (±0.772)
Practicum II: 2.777 (±0.434)	Experimental sciences: 2.552 (±0.524)	Shortcomings	2.432 (±1.066)
		Trial-and-error	3.436 (±1.088)
		Training	1.854 (±0.674)
		Methodologies	3.678 (±0.875)
	Mathematics: 2.871 (±0.464)	Shortcomings	2.508 (±1.037)
		Problem-solving difficulties	3.208 (±0.791)
		Trial-and-error	3.576 (±0.960)
		Training	1.939 (±0.677)
		Methodologies	3.798 (±0.811)
Practicum III: 2.670 (±0.375)	Experimental sciences: 2.616 (±0.394)	Shortcomings	2.399 (±1.108)
		Trial-and-error	3.649 (±0.944)
		Training	1.880 (±0.640)
		Methodologies	3.716 (±0.790)
	Mathematics: 2.716 (±0.439)	Shortcomings	2.335 (±1.152)
		Problem-solving difficulties	2.871 (±0.848)
		Trial-and-error	3.690 (±0.916)
		Training	1.949 (±0.641)
		Methodologies	3.578 (±0.800)

DISCUSSION AND CONCLUSION

This study sought to design and validate the P-MATHEX questionnaire, a comprehensive instrument for assessing preservice teachers' perceptions of their pedagogical knowledge in mathematics and science. Through a rigorous process encompassing item development, pilot testing, and sequential psychometric validation, the research provides valuable insights into both the instrument's structure and the state of preservice teacher education in these domains. The results offer important implications for teacher education programmes and future research on teacher professional development.

Validation and Psychometric Properties

The results indicate that the P-MATHEX instrument exhibits excellent psychometric properties, as evidenced by both EFA and CFA. The emergence of five robust factors–teacher training improvement, methodological shortcomings, applied teaching methods, trial-and-error approach, and student problem-solving difficulties–demonstrates the instrument's capacity to capture a multidimensional understanding of didactic knowledge as perceived by preservice teachers. Internal consistency was outstanding at both the global and subscale levels (with ω values exceeding 0.95), surpassing established benchmarks for reliability in educational measurement (Dunn et al., 2014; Nunnally & Bernstein, 1994).

The construct validity of the instrument is further reinforced by the congruence between the theoretically derived domains and the empirically observed factor structure. The inclusion of both positive and negative item value allowed for a nuanced measurement of attitudes, in line with established recommendations (Messick, 1995; Nunnally, 1978; Rosenberg & Hovland, 1960). Moreover, the use of modern statistical approaches, including IRT-based information functions, provided additional evidence that the subscales offer precise and sensitive measurement across the full spectrum of perceived didactic competence.

Notably, the CFA results confirmed the stability and replicability of the five-factor structure, with global model fit indices (CFI, TLI, and RMSEA) exceeding or approaching rigorous standards for multidimensional educational instruments (Brown, 2015; Hu & Bentler, 1999; MacCallum et al., 1996). This suggests that the instrument is suitable for both research and diagnostic use in teacher education contexts.

Beyond its strong psychometric performance, the P-MATHEX represents an original contribution to the field of teacher education research. To our knowledge, no previous validated instrument has jointly examined preservice teachers' perceptions of pedagogical knowledge in both mathematics and experimental science. By integrating domain-specific and cross-disciplinary dimensions within a single validated framework, this study addresses a critical gap identified in the literature and provides an innovative methodological tool for comparative analyses across content areas.

Perceptions and Self-Reported Competence: Strengths and Gaps

The large-scale application of P-MATHEX across two cohorts (n = 430) revealed several salient patterns in preservice teachers' self-assessed didactic knowledge. Across the sample, participants reported high confidence in their ability to implement applied methodologies and trial-and-error approaches, echoing recent trends in teacher education that encourage the adoption of active, student-centered pedagogies (Fives & Buehl, 2012; Philipp, 2007; Wilkie, 2016).

Conversely, the lowest scores were consistently found in the "teacher training improvement" factor, indicating that preservice teachers recognise significant gaps in their foundational pedagogical preparation, particularly in areas such as inclusive education and information and communication technologies (ICT) integration. This result is aligned with prior research documenting preservice teachers' concerns about the adequacy of their university preparation for real classroom challenges (Depaepe et al., 2013; Franco & Alsina, 2022; García et al., 2014). The identification of these specific weaknesses highlights the need for curricular innovation in teacher education, with a focus on bridging the persistent gap between theoretical coursework and the practical realities of teaching (Onrubia et al., 2020; Verger-Gelabert et al., 2023).

Interestingly, early childhood education preservice teachers systematically reported higher perceived competence than those in primary education, especially in the use of applied methodologies and trial-and-error strategies. This could reflect differences in curricular emphasis, prior training, or classroom experience, and suggests that programme-level variations deserve further attention in the design of teacher education pathways.

The observed pattern of a slight decline in mathematics-related perceived competence during practicum III may signal an increased self-critical perspective as students accumulate real-world teaching experience. Such trends have been noted in prior studies, where greater classroom exposure leads to a more nuanced–and sometimes more critical–self-assessment of professional strengths and weaknesses (Hill et al., 2005; Kleickmann et al., 2013).

Relevance for Teacher Education and Policy

Although the present study draws on participants from a single Spanish university, this context is not institutionally idiosyncratic. Initial teacher education programmes in Spain follow the structure mandated by the EHEA, including four-year concurrent degrees, competency-based curricula, and practicum components aligned with the Bologna process (European Commission/EACEA/Eurydice, 2021; European Ministers of Education, 1999). Comparative analyses indicate that these structural features are broadly shared across many European systems, which follow common principles for teacher competences and qualifications (European Commission, 2018). Furthermore, national studies show a high degree of uniformity in practicum organisation across Spanish universities–typically around 40-45 ECTS credits and concentrated from the third academic year onward–suggesting that the institutional model examined here is representative of the broader Spanish context rather than an isolated case (Verger-Gelabert et al., 2023). Although cross-national validation remains necessary, these structural similarities support the claim that the sample can offer insights that are potentially transferable to other European initial teacher education settings.

The P-MATHEX 2.0 instrument provides teacher educators and policymakers with a robust, empirically grounded tool for diagnosing and monitoring the evolution of preservice teachers' didactic perceptions and competencies. Unlike traditional course assessments or informal feedback, P-MATHEX 2.0 offers a comprehensive, multidimensional profile that captures both strengths (e.g., engagement with active methodologies) and areas needing targeted intervention (e.g., inclusive practices, ICT skills, and theoretical grounding).

Based on these results, several concrete competency-development actions can be implemented in initial teacher education programmes. First, the low scores obtained in the teacher training improvement factor suggest the need to design targeted workshops focusing on inclusive education, ICT integration, and curriculum interpretation, areas frequently identified as challenging for preservice teachers (Depaepe et al., 2013; Franco & Alsina, 2022). Second, the strong performance in applied methodologies but the simultaneous reliance on trial-and-error strategies indicate that practicum seminars could incorporate structured micro-teaching sessions where students practice modelling explicit instructional strategies, scaffolding problem-solving, and using formative assessment techniques. Third, mentor teachers in schools could use P-MATHEX profiles to personalise their feedback, prioritising observations and coaching in areas where preservice teachers report lower confidence. Finally, integrating reflective journals aligned with the P-MATHEX dimensions may help preservice teachers connect university coursework with classroom decision-making more effectively, strengthening the theory-practice link highlighted in recent research (Onrubia et al., 2020; Trumbull & Fluet, 2008).

In addition to implications for teacher education programmes, the results also allow for concrete classroom recommendations for in-service teachers. The prevalence of trial-and-error approaches and repeated explanations suggests the need to incorporate more explicit modelling strategies, such as think-aloud demonstrations during problem-solving or the use of worked-example sequences to support struggling learners (Kalyuga, 2014). Furthermore, the difficulties frequently observed in students–particularly in identifying relevant information, recognising problem structures, and verifying solutions–indicate that teachers may benefit from integrating structured problem-solving heuristics (e.g., Polya-based questioning prompts) and metacognitive routines that foster monitoring and checking (Schoenfeld, 2017). The low confidence reported in inclusive and ICT-related practices also suggests the usefulness of implementing accessible, low-barrier tools such as multimodal representations, manipulatives, and digital simulations to support diverse learners (Ainsworth, 2006). Together, these evidence-informed strategies provide actionable guidance for primary teachers seeking to respond to the didactic challenges highlighted by the P-MATHEX results.

The findings strongly support the need for programmatic responses at the institutional and policy levels. Given that perceived training gaps persist despite curricular reforms, teacher education programmes should redouble efforts to integrate reflective practice, authentic classroom experiences, and explicit instruction in critical areas such as inclusion and technology (Franco & Alsina, 2022; Kleickmann et al., 2013). The alternating practicum model at UJI, which couples school placements with ongoing university-based reflection, may serve as a model for other institutions aiming to strengthen the theory-practice link (Gil & Martí, 2024).

Moreover, the results suggest that effective teacher education cannot be one-size-fits-all. The variability by degree programme (early childhood vs. primary education) and the evolving self-perceptions across practicum stages highlight the need for differentiated, developmental approaches in both programme design and assessment. Instruments like P-MATHEX 2.0 enable the identification of subgroup-specific needs and can inform the allocation of resources and supports for particular populations of preservice teachers.

Theoretical Implications

The structure of the P-MATHEX 2.0 aligns closely with key theoretical frameworks in educational psychology and teacher learning, particularly the concepts of PCK and multidimensional attitudes (Shulman, 1986; Rosenberg & Hovland, 1960). The presence of distinct yet interrelated factors mirrors the complexity of teacher expertise, which is not reducible to content knowledge alone but also encompasses affective and behavioural dimensions such as attitudes, self-efficacy, and adaptive methodological practice (Depaepe et al., 2013; Fives & Buehl, 2012).

The five-factor structure empirically supports Shulman's (1986, 1987) proposition that effective teaching depends on the integration of content knowledge, pedagogical reasoning, and reflective practice. Likewise, the differentiation between factors such as "applied methodologies," "trial-and-error," and "teacher training improvement" reflects the cognitive, affective, and behavioural components described in the classic attitudinal model of Rosenberg and Hovland (1960). These findings therefore provide measurable evidence of how such dimensions coexist and interact in preservice teachers' perceptions.

Furthermore, the coherent alignment between theoretically expected and empirically derived factors strengthens the construct validity of the instrument in Messick's (1995) sense–integrating cognitive, affective, and contextual evidence to capture the meaning of didactic competence. The P-MATHEX thus offers a concrete operationalisation of these long-established theoretical models, translating them into a robust measurement framework applicable across mathematics and science education.

From a methodological standpoint, the successful application of modern psychometric approaches, including EFA and CFA and IRT, underscores the value of using advanced statistical validation to refine complex constructs in teacher education research. The stability of the factorial structure across both mathematics and science domains suggests a degree of generalisability while also revealing subject-specific nuances that merit further investigation in future studies.

Overall, the P-MATHEX bridges a long-standing gap between theoretical conceptions of didactic competence and their empirical measurement. By operationalising multidimensional frameworks such as PCK and attitudinal theory within a single validated instrument, it contributes both theoretically and methodologically to the advancement of research in teacher education.

Author contributions: MSV: conceptualization, project administration, supervision, writing – original draft; **MTS:** formal analysis, methodology, validation, writing – original draft, writing – review & editing; **ELI:** formal analysis, methodology, validation, writing – original draft; **LM:** conceptualization, project administration, writing – review & editing; **GLV:** conceptualization, project administration, writing – review & editing. All authors approved the final version of the article.

Funding: This article was supported by the Conselleria d'Educació, Cultura, Universitats i Ocupació (Generalitat Valenciana) under Grant CIAICO/2022/154.

Ethics declaration: This study was supported by Universitat Jaume I, under registration number CEIDH/31/022. Participants were fully informed that they were taking part in a research study, in accordance with ethical principles governing scientific practice. The objectives of the study were clearly communicated prior to their participation. In compliance with national legislation and institutional requirements, ethical review and approval were not required for this study involving human participants, as data collection was limited to anonymous questionnaire responses. No interviews or recordings were conducted, and researchers did not have direct contact with participants, interacting only with regular teaching staff. Similarly, written informed consent was not required under applicable legal and institutional guidelines, given that participation was entirely anonymous and voluntary.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analysed during this study are available from the authors on request.

REFERENCES

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183-198. https://doi.org/10.1016/j.learninstruc.2006.03.001

American Educational Research Association. (2014). Standards for educational and psychological testing. *American Educational Research Association*. https://www.testingstandards.net/open-access-files.html

Angel-Cuervo, Z.-M., Briceño-Martínez, J.-J., & Bernal-Ballén, A. (2024). Validation of a questionnaire to evaluate mathematics teachers' beliefs about mathematics, teaching, and learning. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(5), Article em2435. https://doi.org/10.29333/ejmste/14447

Björck, V., & Johansson, K. (2019). Problematising the theory–Practice terminology: A discourse analysis of students' statements on work-integrated learning. *Journal of Further and Higher Education, 43*(10), 1363-1375. https://doi.org/10.1080/0309877X.2018.1483016

Bretones, A. (2013). El prácticum de magisterio en educación primaria: Una mirada retrospective [The teaching practicum in primary education: A retrospective look]. *Revista Complutense de Educación, 24*(2), 443-471. https://doi.org/10.5209/rev_RCED.2013.v24.n2.42088

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational Research Methods, 19*(4), 651-682. https://doi.org/10.1177/1094428116656239

Codes, M., Climent, & N., Oliveros, I. (2019). Prospective primary teachers' knowledge about the mathematical practice of defining. In U. T. Jankvist, M. Van den Heuvel-Panhuizen, & M. Veldhuis (Eds.), *Proceedings of the 11th Congress of the European Society for Research in Mathematics Education* (pp. 3871-3878). ERME.

Colén, M. T., & Castro, L. (2017). El desarrollo de la relación teoría y práctica en el grado de maestro en educación primaria [The development of the relationship between theory and practice in the primary education teacher degree]. *Profesorado. Revista de Currículum y Formación del Profesorado, 21*(1), 59-79. https://doi.org/10.30827/profesorado.v21i1.10352

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

Di Martino, P., & Zan, R. (2010). 'Me and maths': Towards a definition of attitude grounded on students' narratives. *Journal of Mathematics Teacher Education, 13*(1), 27-48. https://doi.org/10.1007/s10857-009-9134-z

Díez-Fernández, Á., & Domínguez-Fernández, R. (2018). University tutors as facilitators of student-teachers' reflective learning during their teaching practice. *Estudios Pedagógicos, 44*(2), 311-328. https://doi.org/10.4067/S0718-07052018000200311

Driver, R., Asoko, H., Leach, J., Mortimer, E., & Scott, P. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*(7), 5-12. https://doi.org/10.3102/0013189X023007005

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412. https://doi.org/10.1111/bjop.12046

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.

European Commission. (2018). Key competences for lifelong learning. *European Commission*. https://op.europa.eu/en/publication-detail/-/publication/297a33c8-a1f3-11e9-9d01-01aa75ed71a1/language-en

European Commission/EACEA/Eurydice. (2021). Teachers in Europe: Careers, development and well-being. *Eurydice*. https://eurydice.eacea.ec.europa.eu/sites/default/files/teachers_in_europe_2020_chapter_1.pdf

European Ministers of Education. (1999). The Bologna declaration of 19 June 1999: Joint declaration of the European Ministers of Education. *EHEA*. https://ehea.info/Upload/document/ministerial_declarations/1999_Bologna_Declaration_English_553028.pdf

Fierro, B. T., & Majós, T. M. (2016). Aprendizaje como participación durante el prácticum en un aula de educación infantil [Learning as participation during the practicum in an early childhood education classroom]. *Estudios Pedagógicos, 42*(2), 343-363. https://doi.org/10.4067/S0718-07052016000200020

Fives, H., & Buehl, M. M. (2012). Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook: Vol. 2. Individual differences and cultural and contextual factors* (pp. 471-499). American Psychological Association. https://doi.org/10.1037/13274-019

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling: A Multidisciplinary Journal, 6, 16*(4), 625-641. https://doi.org/10.1080/10705510903203573

Franco, J., & Alsina, Ángel. (2022). El conocimiento del profesorado de Educación Primaria para enseñar estadística y probabilidad: Una revisión sistemática [Primary school teachers' knowledge of statistics and probability: A systematic review]. *Aula Abierta, 51*(1), 7-16. https://doi.org/10.17811/rifie.51.1.2022.7-16

Gairín-Sallán, J., Díaz-Vicario, A., Del Arco Bravo, I., & Alarcia, O. F. (2019). Effect and impact of curricular practices of the early childhood education and primary education bachelor's degrees: Opinion of students, tutors and coordinators. *Educación XX1, 22*(2), 17-43. https://doi.org/10.5944/educxx1.21311

García, A., Buforn, A., & Torregrosa, G. (2014). Un módulo de enseñanza centrado en desarrollar el razonamiento configural: Características desde una perspectiva cognitiva. In M. T Tortosa, J. D. Álvarez, & N. Pellín (Eds.), *Proceedings of the 12th Conference on Research Networks in University Teaching*. Universidad de Alicante.

Gil, J., & Martí, J. (Eds.) (2024). *Las prácticas curriculares en alternancia: propuestas para el ámbito universitario*. Graó.

Godino, J. D., Batanero, C., & Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM Mathematics Education, 39*, 127-135. https://doi.org/10.1007/s11858-006-0004-1

González, E., Muñoz, M., Cruz, A., & Olivares, M. A. (2019). Construction of the teaching professional identity in early childhood education in Córdoba (Spain). *Revista de Ciencias Sociales, 25*(3), 30-41. https://doi.org/10.31876/rcs.v25i3.27354

González, F., Bisquert, M., Haba-Osca, J., & Osca-Lluch, J. (2020). Secondary teacher training in Spain: A study through the official masters in secondary education in public universities. *Revista Interuniversitaria de Formación del Profesorado, 34*(2), 205-224. https://doi.org/10.47553/rifop.v34i2.78055

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203781098

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis* (8th ed.). Cengage Learning.

Henríquez-Rivas, C., & Vergara-Gómez, A. (2025). Design and validation of a questionnaire to explore the geometric work of mathematics teachers. *European Journal of Science and Mathematics Education, 13*(2), 103-118. https://doi.org/10.30935/scimath/16161

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371-406. https://doi.org/10.3102/00028312042002371

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185. https://doi.org/10.1007/BF02289447

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Jacobs, V. R., Lamb, L. C., & Philipp, R. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2), 169-202. https://doi.org/10.5951/jresematheduc.41.2.0169

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 576-597). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.028

Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology, 30*(1), 17-24. https://doi.org/10.1037/h0057123

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. National Academy Press.

Kim, B.-H., & Kim, J. (2016). Development and validation of evaluation indicators for teaching competency in STEAM education in Korea. *Eurasia Journal of Mathematics, Science and Technology Education, 12*(7), 1909-1924. https://doi.org/10.12973/eurasia.2016.1537a

Kind, V., & Kind, P. (2007). Creativity in science education: Perspectives and challenges for developing school science. *Studies in Science Education, 43*(1), 1-37. https://doi.org/10.1080/03057260708560225

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education, 64*(1), 90-106. https://doi.org/10.1177/0022487112460398

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Larrañaga, A. E. (2012). Problemática de los alumnos de magisterio durante el prácticum: Estudio en la facultad de educación y humanidades de Melilla [Problems faced by student teachers during the practicum: A study at the faculty of education and humanities of Melilla]. *Publicaciones de la Facultad de Educación y Humanidades del Campus de Melilla, 42*, 107-120.

León-Urquijo, A. P., García-Jiménez, O. L., & Rendón-Lara, D. B. (2018). Investigaciones sobre prácticas pedagógicas de maestros en formación de instituciones de educación superior. *Espacios, 39*(53).

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130-149. https://doi.org/10.1037/1082-989X.1.2.130

Martínez-Izaguirre, M., Barrenetxea, L., & Díaz-Iso, A. (2019). Aprendizaje entre iguales y desarrollo profesional de los estudiantes en prácticas del grado de educación primaria a través de los foros online [Peer learning and professional development of students in primary education degree placements through online forums]. *Aloma, 37*(1), 45-53. https://doi.org/10.51698/aloma.2019.37.1.45-53

Mason, J. (2016). Perception, interpretation and decision making: Understanding gaps between competence and performance–A commentary. *ZDM Mathematics Education, 48*, 219-226. https://doi.org/10.1007/s11858-016-0764-1

Méndez, V. G., & Magaña, E. C. (2019). Los tutores de prácticas en los centros: Comparativa autonómica del prácticum del máster de educación secundaria [Practice supervisors in schools: A regional comparison of the practicum for the master's degree in secondary education]. *Revista Española de Educación Comparada, 34*(1), 163-181. https://doi.org/10.5944/reec.34.2019.23524

Mendoza, M., Drouilly, N., & Covarrubias, C. G. (2020). Dimensiones formativas del prácticum de pedagogía en educación general básica [Formative dimensions of the practicum in basic general education]. *Estudios Pedagógicos, 46*(2), 139-157. https://doi.org/10.4067/S0718-07052020000200139

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Montes Navarro, M. Ángel, Climent, N., & Contreras, L. C. (2022). Construyendo conocimiento especializado en geometría: Un experimento de enseñanza en formación inicial de maestros [Building specialized knowledge in geometry: A teaching experiment in initial teacher training]. *Aula Abierta, 51*(1), 27-36. https://doi.org/10.17811/rifie.51.1.2022.27-36

Nortes-Martínez-Artero, R., & Nortes-Checa, A. (2018). Conocimientos matemáticos de futuros maestros en resolución de problemas de 6.º de primaria [Mathematical knowledge of future teachers in problem solving for 6th grade of primary school]. *Educatio Siglo XXI, 36*(3), 201-230. https://doi.org/10.6018/j/349971

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Onrubia, J., Mauri, T., Colomina, R., & Ginesta, A. (2020). Promover la colaboración entre maestros tutores/as y tutores/as de universidad en "espacios compartidos" de evaluación formativa en el prácticum de maestro de educación primaria. In C. Lindín, M. B. Esteban, J. C. Bergmann, N. Castells, & P. Rivera-Vargas (Eds.), *Proceedings of the 1st International Conference on Research in Education. Education 2019: Challenges, Trends and Commitments* (pp. 192-198). Institut de Recerca en Educació, Universitat de Barcelona.

Osborne, J., & Dillon, J. (2008). Science education in Europe: Critical reflections. *The Nuffield Foundation*. https://www.nuffieldfoundation.org/about/publications/science-education-in-europe-critical-reflections

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 257-315). Information Age Publishing.

Poveda, B., Barceló, M. L., Rodríguez, I., & López-Gómez, E. (2021). Perceptions and beliefs of education students about learning at university and during the practicum: A qualitative study. *Revista Complutense de Educación, 32*(1), 41-53. https://doi.org/10.5209/rced.67953

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544-559. https://doi.org/10.1080/00223891.2010.496477

Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In M. J. Rosenberg, & C. I. Hovland (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1-14). Yale University Press.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, 34*(S1), 1-97. https://doi.org/10.1007/BF03372160

Sanmamed, M. G., & Abeledo, E. J. F. (2011). Practicum in the learning of teaching professionals. *Revista de Educación, 354*(3), 47-70.

Santágueda-Villanueva, M., & Lorenzo-Valentín, G. (2024). Autosuficiencia matemática en docentes en formación inicial y en ejercicio: Estudio de caso [Mathematical self-sufficiency in pre-service and in-service teachers: A case study.]. *Encuentros. Revista De Ciencias Humanas, Teoría Social Y Pensamiento Crítico*, (20 (enero-abril)), 193–205. https://encuentros.unermb.web.ve/index.php/encuentros/article/view/647

Schoenfeld, A. H. (2017). Learning to think mathematically: Problem-solving, metacognition, and sense making in mathematics. *Journal of Education, 196*(2), 1-38. https://doi.org/10.1177/002205741619600202

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4-14. https://doi.org/10.3102/0013189X015002004

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

Spanish Society for Research in Mathematics Education. (2014). Editorial. Boletín de la SEIEM, 37, 2 [Editorial. SEIEM bulletin, 37, 2]. *Spanish Society for Research in Mathematics Education*. https://www.seiem.es/wp-content/uploads/2025/04/boletin37.pdf

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Allyn & Bacon.

Trumbull, D. J., & Fluet, K. (2008). What can be learned from writing about early field experiences? *Teaching and Teacher Education, 24*(6), 1672-1685. https://doi.org/10.1016/j.tate.2007.12.001

Ulusoy, F. (2021). Prospective early childhood and elementary school mathematics teachers' concept images and concept definitions of triangles. *International Journal of Science and Mathematics Education, 19*, 1057-1078. https://doi.org/10.1007/s10763-020-10105-6

Vargas, S. D., Bernal-Ballén, A., Briceño-Martínez, J. J., & Ariza-Bareño, Y. (2024). Design and validation of an instrument to determine the relationship between pedagogical content knowledge and practical work in science instruction. *Eurasia Journal of Mathematics, Science and Technology Education, 20*(1), Article em2382. https://doi.org/10.29333/ejmste/13962

Verger-Gelabert, S., Riera-Negre, L., Rosselló-Ramón, M. R., & Mut-Amengual, B. (2023). Análisis del prácticum de educación infantil y primaria en las universidades Españolas [Analysis of the practicum in early childhood and primary education in Spanish universities]. *Aula Abierta, 52*(1), 81-90. https://doi.org/10.17811/rifie.52.1.2023.81-90

Wilkie, K. J. (2016). Learning to teach upper primary school algebra: Changes to teachers' mathematical knowledge for teaching functional thinking. *Mathematics Education Research Journal, 28*(2), 245-275. https://doi.org/10.1007/s13394-015-0151-1

APPENDIX A: P-MATHEX QUESTIONNAIRE 2024-2025

Information: Confidentiality (English)

Basic information on data protection according to the general data protection regulation (EU) 2016/679: The data you provide will be processed by the mathematics and experimental sciences didactics research group at UJI (DIMATHEX). The purpose of this processing is solely for research, meaning that the data will not be used for purposes other than scientific research, nor will they be shared with third parties. Furthermore, DIMATHEX commits to using the data in a strictly confidential manner. Therefore, we guarantee the implementation of the necessary technical and organisational measures to ensure data security and prevent their alteration, loss, processing, or unauthorised access, considering the state of technology, the nature of the data, and the risks to which they may be exposed, whether due to human actions or physical or natural environments.

Data controller: UJI. DIMATHEX group.

Purpose of processing: Research on the teaching and learning of mathematics and experimental sciences in early childhood/primary education classrooms.

I have been informed and I accept that the DIMATHEX research group at UJI will process my personal data in accordance with the general data protection regulation (EU) 2016/679.

☐ Yes

☐ No

Information

This questionnaire is intended for student teachers in the early childhood and primary education degrees at UJI in Castelló.

It is a Likert-type questionnaire divided into two independent thematic sections (mathematics and experimental sciences), aiming to understand your practicum experience.

If you have any questions about any aspect of the questionnaire, or would like more information about the study, you can contact us via the following email: santague@uji.es

Thank you very much for your collaboration.

Mathematics Block

After my practicum, in my training in mathematics didactics, I would like to improve in the following aspects.

Rating scale: 1–strongly disagree, 2–disagree, 3–neither agree nor disagree, 4–agree, & 5–strongly agree

- Knowledge and understanding of the contents.
- Pedagogical and methodological competencies for teaching mathematics.
- Knowledge of the mathematics curriculum.
- ICT skills applied to mathematics teaching.
- Knowledge of inclusive education and its implementation.
- Cross-curricular and interdisciplinary teaching.
- Approaches for developing competencies in everyday life situations.
- Use of manipulative materials.

When students have difficulty understanding a specific mathematics content, what the teacher in my classroom does is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Explain it again in the same way but more slowly.

When students have had difficulty understanding a specific mathematics content, what I have done during my practicum is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Explain it again in the same way but more slowly.

When students have difficulty understanding a specific mathematics content, what the teacher in my classroom does is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Use a different methodology to revisit that content.
- Present problem situations with progressively increasing difficulty.

When students have had difficulty understanding a specific mathematics content, what I have done during my practicum is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Use a different methodology to revisit that content.
- Present problem situations with progressively increasing difficulty.

When students have difficulty understanding a specific mathematics content, what the teacher in my classroom does is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Encourage trial and error.

When students have had difficulty understanding a specific mathematics content, what I have done during my practicum is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Encourage trial and error.

Among the following difficulties in problem solving, which have you observed in the students in your practicum classroom and how frequently?

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- They do not understand the problem statement because it is not in their native language.
- They have poor reading comprehension even though the problem is in their native language.
- Although they know how to perform the necessary operations, they do not identify them with the problem.
- They do not recognise similar problems.
- They have difficulties with operations: they do not handle decimals, forget zeros in some operations ...
- They do not identify the data in the problem.
- They do not identify the unknowns in the problem.
- They do not change their point of view, even if they get an impossible result.
- They do not check their solutions.
- They do not reflect on what the problem is asking.
- They cannot explain the procedure they followed, whether correct or incorrect.

Experimental Sciences Block

After my practicum, in my training in experimental sciences didactics, I would like to improve in the following aspects:

Rating scale: 1–strongly disagree, 2–disagree, 3–neither agree nor disagree, 4–agree, & 5–strongly agree

- Knowledge and understanding of the contents.
- Pedagogical and methodological competencies for teaching experimental sciences.
- Knowledge of the experimental sciences curriculum.
- ICT skills applied to the teaching of experimental sciences.
- Knowledge of inclusive education and its implementation.
- Cross-curricular and interdisciplinary teaching.
- Approaches for developing competencies in everyday life situations.
- Activities in the science laboratory.

When students have difficulty understanding a specific topic in experimental sciences, what the teacher in my classroom does is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Explain it again in the same way but more slowly.

When students have had difficulty understanding a specific topic in experimental sciences, what I have done during my teaching practice is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Explain it again in the same way but more slowly.

When students have difficulty understanding a specific topic in experimental sciences, what the teacher in my classroom does is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Use a different methodology to revisit that topic.
- Present problem-solving situations with progressively increasing difficulty.

When students have had difficulty understanding a specific topic in experimental sciences, what I have done during my teaching practice is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Use a different methodology to revisit that topic.
- Present problem-solving situations with progressively increasing difficulty.

When students have difficulty understanding a specific topic in experimental sciences, what the teacher in my classroom does is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Encourage trial and error.

When students have had difficulty understanding a specific topic in experimental sciences, what I have done during my teaching practice is:

Frequency scale: 1–never, 2–rarely, 3–sometimes, 4–often, & 5–always

- Encourage trial and error.

Socio-Demographic Data

Gender

- Male
- Female
- Other

Age

University degree you are studying

- Bachelor’s degree in early childhood education
- Bachelor’s degree in primary education

Course you are currently taking

- MI1847–Practicum I
- MI1852–Practicum II
- MI1856–Practicum III
- MP1847–Practicum I
- MP1852–Practicum II
- MP1856–Practicum III (primary education)
- MP1857–Practicum III (music specialisation)
- MP1858–Practicum III (physical education specialisation)
- MP1859–Practicum III (English specialisation)

In which educational center are you doing your practicum?

In which town/city?

What type of school is it?

- Public
- Semi-private (concertado)

Mark the grade level in which you are doing your practicum (if you are in a multilevel classroom, mark all the grade levels present):

- Early childhood 2 years
- Early childhood 3 years
- Early childhood 4 years
- Early childhood 5 years
- Primary 1st grade
- Primary 2nd grade
- Primary 3rd grade

- Primary 4th grade
- Primary 5th grade
- Primary 6th grade

End of the questionnaire

Thank you very much for your collaboration.