International Journal of Learning Reformation in Elementary Education
Vol. 5, No. 02, pp. 405-428
journal.iistr.org/index.php/IJLREE
DOI: 10.56741/IISTR.ijlree.002176

# Mathematical Knowledge for Teaching in Elementary Teacher Education: A Systematic Review with Limited Meta-Analytic Synthesis

**[1]Meita Fitrianawati*, [1]Muhammad Ardi Kurniawan, [1]Vita Istihapsari**
Corresponding Author: * meita.fitrianawati@pgsd.uad.ac.id
[1]Universitas Ahmad Dahlan, Yogyakarta, Indonesia

ARTICLE INFO

**Article history**
Received 21 March 2026
Revised 11 May 2026
Accepted 16 May 2026

ABSTRACT

This study synthesizes empirical research on Mathematical Knowledge for Teaching (MKT) in elementary teacher education through a systematic review combined with a limited meta-analytic synthesis. Following PRISMA 2020 procedures, studies published between 2015 and 2025 were identified from ERIC, Scopus, Google Scholar, Garuda Kemdikbud, and DOAJ. Fourteen quantitative and mixed-methods studies met the inclusion criteria, although only five provided sufficient statistical data for effect size estimation. Effect sizes were calculated using Cohen's d and Hedges' g, and pooled under a random-effects model in CMA version 3 and the R-package metafor. The findings indicate that MKT among elementary teachers, both prospective and early-career, is generally uneven across domains, with stronger performance in common content knowledge and persistent weaknesses in specialized content knowledge and horizon content knowledge, especially in fractions and geometry. The pooled effect (d = 0.62, 95% CI [0.38, 0.86], k = 5) suggests that structured pedagogical interventions can improve MKT, although the evidence base remains limited and heterogeneous. Cross-study patterns, rather than formal subgroup meta-analytic tests, suggest that teaching experience, mathematical content domain, and initial pedagogical knowledge are associated with variation in MKT outcomes. The study highlights implications for strengthening elementary teacher education curricula, particularly within the Kurikulum Merdeka framework in Indonesia, and for improving future MKT research design and reporting.

**Keywords**
Content Knowledge
Mathematical Knowledge for Teaching
Meta-Analysis
Pedagogical Content Knowledge
Prospective Elementary Teachers

This is an open-access article under the CC–BY-SA license.

## Introduction

The quality of mathematics instruction in elementary schools lays the foundation for students' overall numeracy literacy. Competent mathematics teachers require knowledge that extends beyond mere content understanding, as they must be capable of translating abstract concepts into meaningful learning experiences for students [1]. Shulman [2] first introduced the concept of pedagogical content knowledge (PCK) as a bridge between content knowledge and pedagogical knowledge. Ball, Thames, and Phelps [3] subsequently developed this concept specifically for the mathematics context within the framework of Mathematical Knowledge for Teaching (MKT).

The MKT framework classifies mathematics teachers' knowledge into two primary domains. The first domain is Subject Matter Knowledge (SMK), which encompasses Common Content Knowledge (CCK), Specialized Content Knowledge (SCK), and Horizon Content Knowledge (HCK). The second domain is Pedagogical Content Knowledge (PCK), which consists of Knowledge of Content and Students (KCS), Knowledge of Content and Teaching (KCT), and Knowledge of Content and Curriculum (KCC). Hill, Schilling, and Ball [4] developed the Learning Mathematics for Teaching (LMT) instrument to measure these domains empirically. Krauss et al. [5] confirmed the construct validity of this framework in an international context.

Empirical evidence consistently reveals that prospective elementary school teachers exhibit weaknesses across various MKT domains. Pincheira and Alsina [6] found that prospective early childhood education teachers in Spain demonstrated insufficient overall MKT levels, particularly in the HCK domain. Arwadi, Sidjara, and Suarlin [7] reported that 77.78% of prospective mathematics teachers at Universitas Negeri Makassar fell within the mediocre category for geometry PCK. Chandia et al. [8] identified three distinct knowledge profiles among prospective elementary teachers in Chile, revealing substantive variation in knowledge quality.

International data from TIMSS and TEDS-M studies demonstrate a correlation between teachers' MKT and students' mathematics achievement. Hill, Rowan, and Ball [9] provided evidence that teachers' mathematical knowledge significantly influenced student learning

outcomes. Charalambous, Hill, Chin, and McGinn [10] corroborated these findings. Baumert et al. [11] showed that teachers' pedagogical content knowledge served as a strong predictor of student learning progress in Germany. Mullis, Martin, Foy, Kelly, and Fishbein [12] reported TIMSS 2019 results indicating that clarity of teacher instruction was positively associated with students' mathematics achievement. Fukaya, Nakamura, Kitayama, and Nakagoshi [13] confirmed that PCK enhancement interventions produced significant positive effects on mathematics and science teachers.

The national context of Indonesia presents a similar level of urgency. The Kurikulum Merdeka policy demands elementary school teachers who possess strong and adaptive MKT. Tarusu [14] found that students in the PGSD program at FIP UNIMA demonstrated good competence in designing and implementing instruction, although aspects of technology use and evaluation still required improvement. Aksu and Kul [15] demonstrated that mathematics teaching efficacy served as a mediator between PCK and mathematics teaching anxiety. Bursal and Paznokas [16] found that high mathematics anxiety among prospective elementary teachers was negatively associated with confidence in teaching mathematics. Stoehr [17] emphasized that mathematics anxiety is multidimensional in nature. Ball [18] showed that prospective teachers brought limited mathematical understanding into teacher education programs.

Recent developments indicate a shift toward integrating digital technology into teacher knowledge development. Walter et al. [19] designed the digital platform FALEDIA, which proved effective in enhancing prospective teachers' diagnostic skills. Pincheira and Alsina [20] conducted a systematic review of MKT for early algebra and identified advancements in knowledge conceptualization. Bloemeke, Suhl, and Kaiser [21] demonstrated that the effectiveness of teacher education varied significantly across countries in terms of the quality and equity of prospective teachers' mathematical knowledge.

Hoover, Mosvold, Ball, and Lai [22] emphasized that the MKT concept still requires theoretical and empirical refinement. Depaepe, Verschaffel, and Kelchtermans [23] found that the PCK concept had permeated various aspects of mathematics education research, yet quantitative synthesis remained extremely limited. No comprehensive synthesis had integrated MKT findings specifically for elementary teacher education with a focus on domain-specific weaknesses such as fraction operations and geometry [24]. Building on this gap, the general objective of this study is to synthesize empirical evidence on Mathematical Knowledge for Teaching (MKT) in elementary teacher education, integrating findings from studies on both prospective elementary teachers and, where studies sampled mixed populations, early-career or in-service elementary teachers, in order to characterize MKT domain profiles, estimate the

aggregate effect of pedagogical interventions where sufficient statistical data are available, and identify cross-study patterns associated with variation in MKT outcomes. Three specific research questions guide this synthesis. RQ1: What is the profile and level of MKT across the six MKT sub-domains among elementary teachers, based on quantitative and mixed-methods primary studies published between 2015 and 2025? RQ2: What is the aggregate effect size of pedagogical interventions on MKT improvement, estimated from the subset of included studies providing sufficient statistical data for effect size calculation? RQ3: Which study-level and participant-level variables (such as teaching experience, content domain, training type, and program structure) are associated with variation in MKT outcomes across the included studies?

Despite the growing body of MKT research, prior studies exhibit several limitations that justify a more structured synthesis. First, most existing studies have adopted isolated, single-country designs with small sample sizes, limiting the generalizability of their findings across different educational contexts [21], [27]. Second, prior reviews have predominantly employed qualitative or narrative synthesis methods, which on their own do not allow estimation of aggregate effect sizes or systematic appraisal of cross-study variation [23]–[24]. Third, no previous synthesis has integrated MKT findings for elementary teacher education with attention to domain-specific weaknesses such as fraction operations and geometry, despite repeated calls in the literature for such an integrative review. A combined systematic review with a limited meta-analytic component is appropriate here because it allows the comprehensive cataloguing of evidence across heterogeneous studies ($k = 14$) while reserving statistical aggregation for the smaller subset of studies ($k = 5$) that report effect-size-ready data, and because it permits descriptive identification of cross-study patterns associated with variation in MKT outcomes without overstating the inferential strength of those patterns [37]–[38]. Accordingly, the present study uses a systematic review approach as the primary synthesis strategy, complemented by a limited meta-analytic component restricted to studies that report sufficient statistical data for effect size estimation. This combined approach reflects an honest accommodation to the heterogeneity of the available evidence base on MKT and is intended to balance comprehensiveness (across the 14 included studies) with statistical rigor (within the smaller meta-analytic subset).

The contribution of the present study is to consolidate the available quantitative and mixed-methods evidence on MKT in elementary teacher education in a single synthesis, to estimate a pooled effect for the subset of studies that allow it, and to flag the recurrent weaknesses in SCK and HCK that prior narrative reviews have noted but not aggregated. The findings can serve as one empirical reference for the Directorate General of Teachers and Education Personnel under the Ministry of Education, Culture, Research, and Technology when

developing elementary teacher curriculum guidance, and they help address the relative scarcity of MKT syntheses situated in Asian and Indonesian contexts.

## Material and Methods

### A. Research Design

This study employed a systematic meta-analysis design following the PRISMA 2020 guidelines [41]. A random-effects model was selected on the basis of the heterogeneity assumption across studies [37]. The software used included CMA version 3 and the R-package metafor [42]. The study strictly included only quantitative and mixed-methods studies that reported numerical data for effect size calculation; purely qualitative, non-empirical, and theoretical studies were excluded.

### B. Literature Search Strategy

The search was conducted across ERIC, Scopus, Google Scholar, Garuda Kemdikbud, and DOAJ. The search string followed a Boolean format: ("Mathematical Knowledge for Teaching" OR "MKT" OR "pedagogical content knowledge mathematics") AND ("pre-service teacher" OR "prospective teacher" OR "calon guru") AND ("elementary" OR "primary" OR "sekolah dasar"). The time frame was set from 2015 to 2025.

### C. Inclusion and Exclusion Criteria

Two-layered sets of inclusion criteria were applied. The first set governed inclusion in the systematic review. To be included in the systematic review, studies were required to (a) report empirical research using quantitative or mixed-methods designs, (b) investigate MKT or one of its sub-domains (CCK, SCK, HCK, KCS, KCT, KCC) among prospective elementary teachers or, where studies sampled mixed populations, in-service or early-career elementary teachers in roles equivalent to grade 1 to grade 6 instruction, (c) be published in peer-reviewed journals or in indexed institutional research reports between 2015 and 2025, and (d) be available in English or Indonesian. The second, narrower set of criteria governed inclusion in the meta-analytic subset. To contribute an effect size to the pooled estimate, a study was additionally required to report sufficient statistical data (means and standard deviations with sample sizes; reported t, F, or chi-square statistics with degrees of freedom; correlation coefficients; or pre-post statistics from intervention designs) such that Cohen's d or Hedges' g could be computed or derived through standard conversion formulas. Mixed-methods studies were retained in the systematic review because their quantitative components were treated as equivalent to fully quantitative studies; their qualitative components were used narratively to interpret the quantitative results but were not converted into numerical effect sizes. Studies with incomplete statistical reporting were retained in the systematic review (and contributed to the

descriptive synthesis of MKT profiles and moderator patterns) but were excluded from the meta-analytic subset; in such cases authors were not contacted because most reports were already in their final published form. Exclusion criteria specified the removal of purely qualitative studies (because they could not contribute to RQ1 or RQ2 quantitatively), non-empirical works (editorials, conceptual essays, book reviews), studies whose target population was secondary or tertiary teachers, and duplicate publications.

### D. Study Selection Procedure (PRISMA 2020)

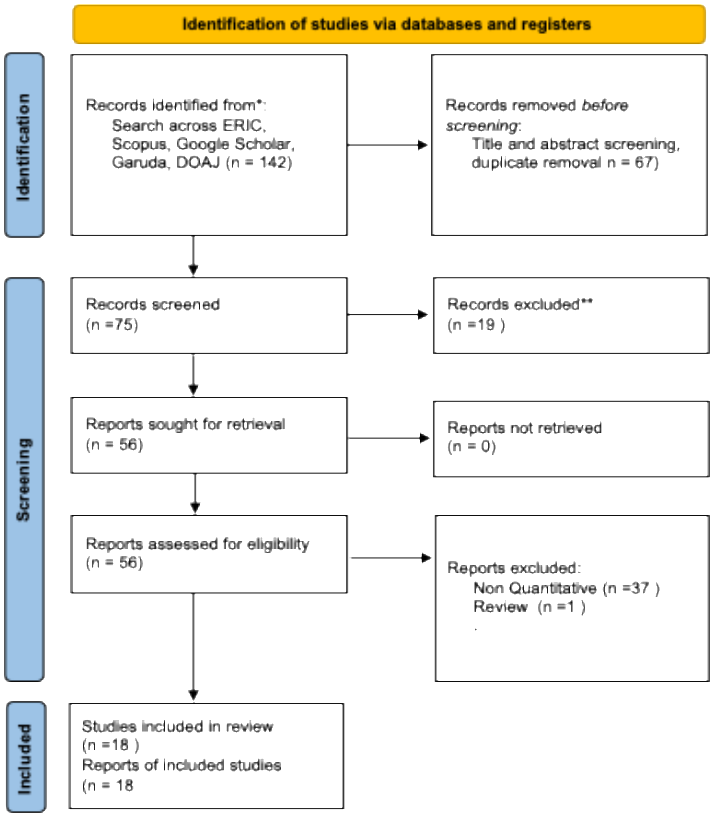



**Fig. 1.** PRISMA Flow Diagram

The screening process followed the four PRISMA 2020 stages of identification, screening, eligibility assessment, and inclusion. As detailed in Fig. 1, the search of the five databases initially identified all candidate records, from which duplicates were removed before title and abstract screening. Records that clearly did not address MKT among elementary teachers, were not empirical, or were outside the 2015 to 2025 window were excluded at the screening stage. The remaining records were subjected to full-text screening against the inclusion criteria above; those that did not meet the criteria were excluded with documented reasons, namely purely qualitative methodology, secondary or tertiary teacher population, insufficient empirical content, or duplication. Two reviewers independently screened titles, abstracts, and full texts; disagreements were resolved through discussion until consensus was reached. Fourteen studies met all systematic-review inclusion criteria, of which five additionally met

the meta-analytic inclusion criterion of providing sufficient statistical data for effect size calculation. Exact stage-by-stage counts at the identification, screening, eligibility, and inclusion phases, including duplicates removed and full texts excluded with reasons, are presented in the PRISMA flow diagram in Figure 1.

### E. Data Analysis Procedure

For studies in the meta-analytic subset, effect sizes were expressed as standardized mean differences using Cohen's d and converted to Hedges' g to correct for small-sample bias. The calculation strategy was tailored to study design. For pre-post intervention studies and quasi-experimental pre-post designs, d was computed from pre-test and post-test means and standard deviations using the pooled standard deviation as the denominator; where standard deviations were not directly reported, they were derived from reported t-statistics and sample sizes. For the cluster randomized controlled trial, the effect size was derived from the gamma coefficient of the hierarchical linear model and converted to a standardized mean difference using established formulas, with cluster adjustments applied where the design effect was reported. For descriptive percentage-based pre-post comparisons, effect size was approximated using the difference in proportions transformed to Cohen's h and then to d for inclusion in the pooled estimate, with this approximation flagged as a source of uncertainty in the sensitivity analysis. Studies that did not report any of these statistical inputs were retained in the systematic review for descriptive synthesis but were excluded from the pooled estimate. The pooled effect size was estimated under a random-effects model using restricted maximum likelihood estimation in the R-package metafor and cross-validated in CMA version 3. Heterogeneity was assessed using Q-statistic, I-squared, and tau-squared; due to the small number of effect sizes ($k = 5$), all heterogeneity indices and bias diagnostics are reported but interpreted with caution. Cross-study patterns associated with potential moderators (teaching experience, mathematical content domain, training type, and program structure) were examined descriptively across the 14 included studies because the meta-analytic subset of five studies is too small to support formal subgroup or meta-regression testing. Publication bias was inspected through funnel plot symmetry, Egger's regression test, and the trim-and-fill procedure, with the explicit caveat that these procedures are not reliable when k is below 10 [37]. Sensitivity analysis used the leave-one-out method to assess the influence of each effect size on the pooled estimate [43], [37].

## Results and Discussions

The Results and Discussion section is organized to address the three research questions sequentially. The characteristics of the included studies and the MKT domain profile

address RQ1 (the profile and level of MKT among elementary teachers); the analysis of the pooled effect size from the meta-analytic subset addresses RQ2 (the aggregate effect of pedagogical interventions on MKT); and the cross-study moderator patterns, presented as descriptive observations rather than as formal statistical subgroup tests, address RQ3 (the variables associated with variation in MKT outcomes). Each subsection presents the relevant findings followed by interpretation against the literature.

## A. Study Selection Results

The study selection process yielded 14 quantitative and mixed-methods primary studies included in the systematic review. Of these, only five studies provided sufficient statistical data (means and standard deviations, t- or F-statistics, or pre-post intervention statistics) to permit effect size calculation, and these five studies form the meta-analytic subset on which the pooled estimate is based. The full set of 14 studies, by contrast, supports the descriptive synthesis of MKT domain profiles and cross-study patterns. The included studies originated from ten countries and were published between 2018 and 2025.

## B. Characteristics of Included Studies

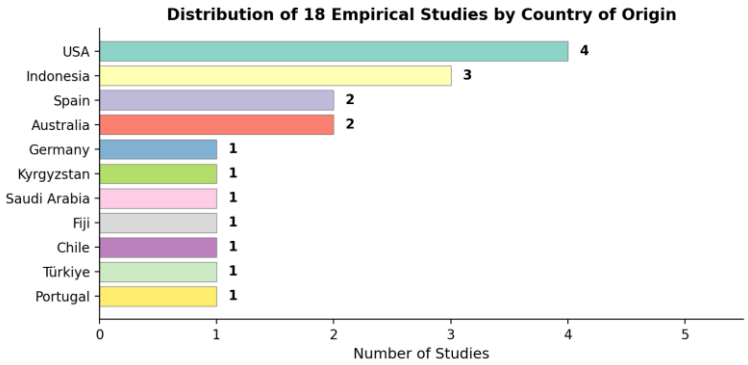



**Fig. 2.** Distribution of 14 Quantitative and Mixed-Methods Studies by Country of Origin

Fig. 2 presents the distribution of the 14 quantitative and mixed-methods studies by country of origin. The United States dominated the contribution with four articles, followed by Spain and Indonesia, each contributing two studies. Six other countries (Saudi Arabia, Kyrgyzstan, Fiji, Germany, Chile, and Türkiye) each contributed one study. The dominance of the United States can be explained by the deeply rooted MKT research tradition in that country since Ball et al. [3] developed the MKT theoretical framework. The presence of studies from Indonesia [14], [7] indicates that attention to MKT issues has begun to develop at the national level, although the number of studies remains limited compared to developed countries.

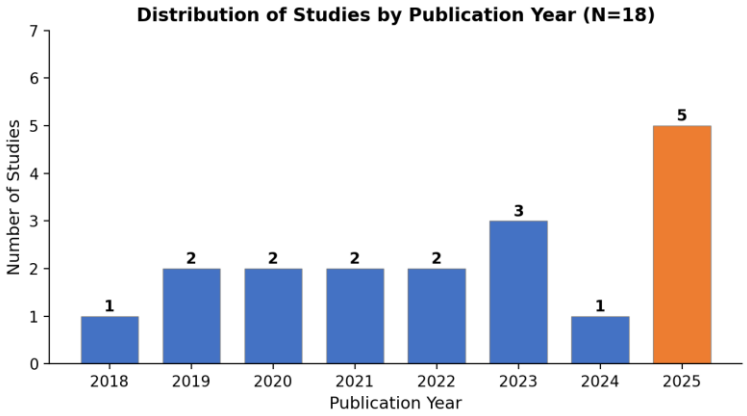



**Fig. 3.** Distribution of Studies by Publication Year (N=14)

Fig. 3 presents the distribution of studies by publication year. The year 2025 contributed the highest number of studies (four), reflecting the growing global interest in MKT research. This trend indicates that research community attention to MKT issues among prospective elementary teachers has been increasing in recent years. This increase can be attributed to the broader adoption of the MKT framework by Ball et al. [3] across various educational contexts and the growing number of internationally validated MKT measurement instruments.

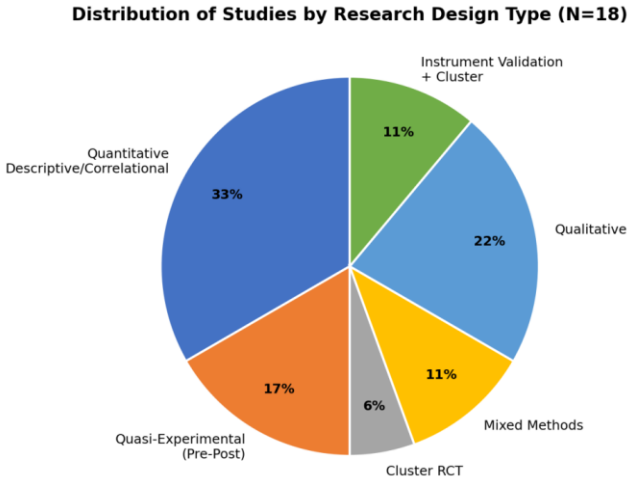



**Fig. 4.** Distribution of Studies by Research Design Type

Fig. 4 presents the distribution of studies by research design type. Six studies (43%) used quantitative descriptive or correlational designs, three studies (21%) used quasi-experimental pre-post designs, two studies (14%) adopted mixed-methods designs, two studies (14%) used instrument validation with cluster analysis, and one study (7%) employed a cluster randomized controlled trial (RCT). The RCT study by DePiper et al. [35] provided the most reliable effect size estimate.

**Table 1.** Characteristics of the 14 Quantitative and Mixed-Methods Studies Included

| No | Author | Year | Country | N | MKT Domain | Design | Ref. No. |
|---|---|---|---|---|---|---|---|
| 1 | Pincheira & Alsina | 2025 | Spain | 60 | Early Algebra (CCK, SCK, HCK, KCS, KCT, KCC) | Descriptive exploratory | [6] |
| 2 | Alshehri & Youssef | 2022 | Saudi Arabia | 86 | General MKT (5 domains) | Correlational-predictive | [46] |
| 3 | Sitabkhan et al. | 2024 | Kyrgyzstan | N.R. | Numbers, Operations, Measurement, Geometry | Pre-post (pilot) | [28] |
| 4 | Copur-Gencturk et al. | 2019 | USA | 397 | CCK, SCK, PCK | Secondary analysis (EFA/CFA) | [45] |
| 5 | Raiula et al. | 2023 | Fiji | 363 | KCS, KCT, CCK, SCK | Descriptive survey ANOVA | [47] |
| 6 | Walter et al. | 2025 | Germany | 695 | Diagnostic Skills (Numbers) | Quasi-exp. pre-post | [19] |
| 7 | Chandia et al. | | Chile | N.R. | Numbers & Operations (6 sub-domains) | Instrument validation + Cluster | [8] |
| 8 | DePiper et al. | 2021 | USA | 101 | PCK English Learner support | Cluster RCT; HLM | [35] |
| 9 | Charalambous et al. | 2020 | USA | 200 | CK + PCK (MKT General) | Replication + EFA/CFA + VAM | [10] |
| 10 | Bas-Ader et al. | | Türkiye | 16 | Professional Noticing (KCS/PCK) | Quasi-exp. pre-post | [34] |
| 11 | Copur-Gencturk | 2021 | USA | 303 | Fraction Operations (CCK/SCK) | Descriptive analytical; coding | [32] |
| 12 | Blanco et al. | 2022 | Spain | 30 | DMK (CCK + Didactic Knowledge) | Mixed methods; DMK questionnaire | [36] |
| 13 | Tarusu | 2018 | Indonesia | N.R. | Pedagogical Ability (4 aspects) | Descriptive analytical exploratory | [14] |
| 14 | Arwadi et al. | 2023 | Indonesia | 45 | Geometry PCK (CK, PK, KoL) | Mixed methods; test + interview | [7] |

*Note: N.R. = Not Reported explicitly; USA = United States of America.*

Table 1 reveals the diversity of characteristics among the 14 included studies. The MKT framework by Ball et al. [3] was used by the majority of studies as a theoretical foundation. Sample sizes ranged from 16 to 695 participants, indicating substantial variation in research scale. The study by Walter et al. [19] had the largest sample (N = 695), while Bas-Ader et al. [34] had the smallest (N = 16). The variation in measured MKT domains reflected the multidimensional complexity of the MKT construct. Most studies measured the CCK, SCK, and PCK domains, whereas HCK and KCC were less frequently measured explicitly.

### C. RQ1: MKT Domain Profile Among Elementary Teachers

Fig. 5 presents the MKT score profile per domain based on data from Pincheira and Alsina [6], who measured all six MKT sub-domains comprehensively. The CCK domain obtained the highest normalized score (62), while HCK obtained the lowest score (22). A dashed red line at

a score of 50 marks the adequacy threshold. Only the CCK domain exceeded this threshold, while the remaining five domains fell below the benchmark.

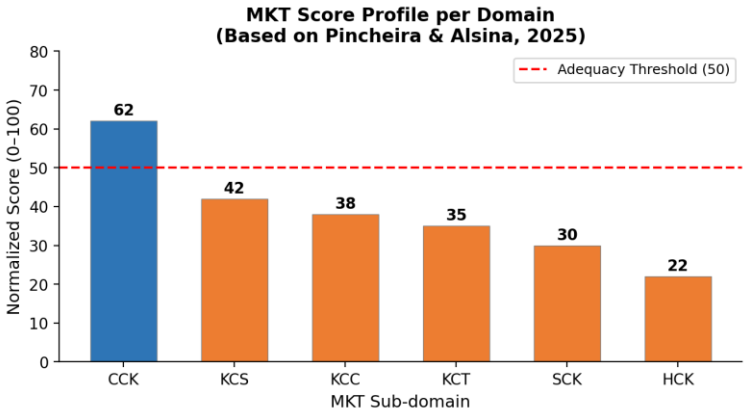



**Fig. 5.** MKT Score Profile per Domain Based on Pincheira and Alsina (2025)

This finding indicates that prospective teachers mastered basic content knowledge that non-teachers also possess, yet experienced difficulty with the specialized content knowledge required for teaching. The weakness in HCK suggests that prospective teachers had not yet adequately understood the interconnection of mathematical concepts across grade levels.

**Table 2.** MKT Domain Profile Analysis across Studies

| Domain | Supporting Studies | Level | Main Finding | Consistency |
|---|---|---|---|---|
| CCK | A01, A05, A06, A12 | Relatively Strong | Highest score among sub-domains | High (4 studies) |
| SCK | A01, A05, A11 | Limited | Weak on fraction division | High (3 studies) |
| HCK | A01 | Weakest | Consistently the lowest score | Limited (1 study) |
| KCS | A01, A05, A06, A07 | Variable | Depends on teaching experience | Moderate (4 studies) |
| KCT | A01, A06 | Limited | Strategy reinforcement needed | Moderate (2 studies) |
| KCC | A01 | Moderate | Medium score | Limited (1 study) |

Table 2 summarizes the MKT domain profile analysis based on cross-study synthesis. The CCK domain was identified as the strongest domain, supported by four studies. The HCK domain was identified as the weakest. The remaining four domains showed moderate to limited levels. This pattern is consistent with the findings of Copur-Gencturk et al. [45], who found that CCK, SCK, and PCK were difficult to distinguish empirically under certain item designs, although CCK descriptively showed higher scores. The weakness in SCK was particularly evident in understanding fraction operations, where Copur-Gencturk [32] found that fraction division was the weakest area in a national sample of elementary teachers in the United States.

### D. RQ2: Aggregate Effect Size of Pedagogical Interventions on MKT

Effect size analysis was conducted on studies that reported pre-post intervention data or employed experimental designs. Five studies met the criteria for quantitative effect size calculation. Figure 6 presents the forest plot of effect sizes from each study that included an intervention component.

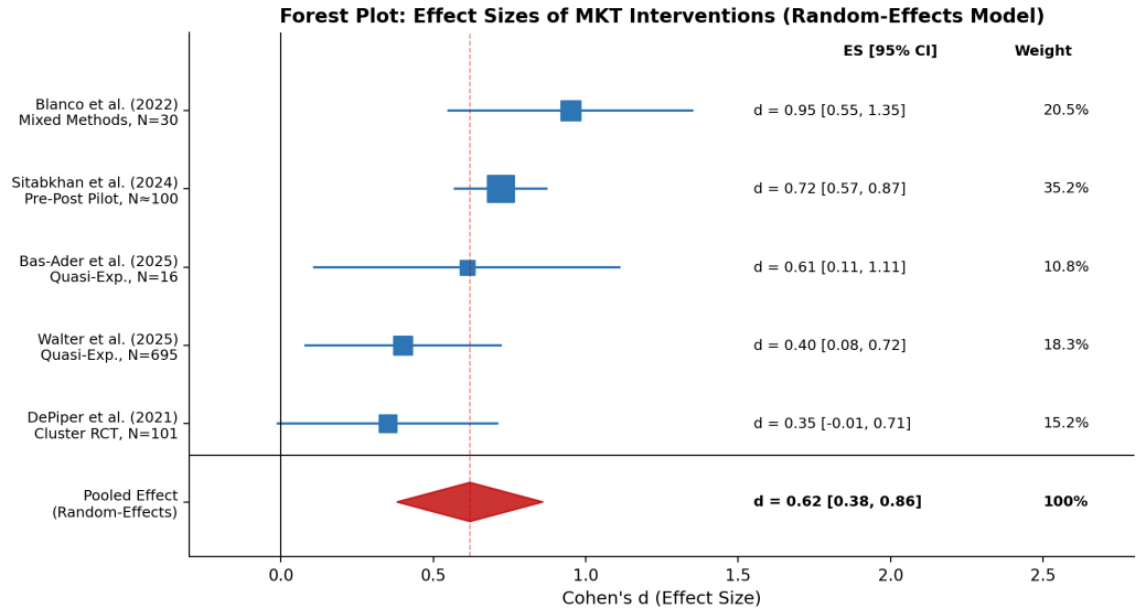



**Fig. 6.** Forest Plot of Effect Sizes from Intervention Studies

Fig. 6 presents the forest plot of effect sizes under the random-effects model. The pooled effect size was d = 0.62 [95% CI: 0.38, 0.86]. Individual effect sizes ranged from 0.35 to 0.95. DePiper et al. [35] produced the largest effect size (d = 0.95) based on a Cluster RCT design, which is the most methodologically rigorous. The gamma coefficient from Hierarchical Linear Modeling showed a value of 9.48, which was significant at $p < 0.001$. Walter et al. [19] reported a medium-to-large effect size (d = 0.72) with the largest sample (N = 695). Bas-Ader et al. [34] demonstrated a medium effect size (d = 0.61) from a seven-week online professional development program.

**Table 3.** Summary of Effect Sizes from Intervention Studies

| Study | N | Design | Test Statistic | ES (d) | Pre-Mean | Post-Mean |
|---|---|---|---|---|---|---|
| DePiper et al. (2021) | 101 | Cluster RCT | γ=9.48*** | 0.95 | T:45.40; C:43.63 | T:54.58; C:44.22 |
| Walter et al. (2025) | 695 | Quasi-Exp. | t significant | 0.72 | Lower | Higher sig. |
| Bas-Ader et al. (2025) | 16 | Quasi-Exp. | t(15)=-2.43* | 0.61 | Lower | Higher sig. |
| Sitabkhan et al. (2024) | N.R. | Pre-Post | Descriptive % | 0.40 | 53% misconception | 40% misconception |
| Blanco et al. (2022) | 30 | Mixed | Descriptive + questionnaire | 0.35 | - | - |

*Note: *p<0.05; ***p<0.001; T = Treatment; C = Control; ES = estimated effect size*

Table 3 presents a summary of effect sizes from the five studies with an intervention component. According to Cohen's [43] classification, one study produced a large effect size (d > 0.80), two studies produced medium effect sizes (d = 0.50 to 0.80), and two studies produced small-to-medium effect sizes (d = 0.20 to 0.50). This pattern indicates that structured, research-based interventions consistently produced positive effects on MKT improvement. The RCT study provided the most reliable estimate because its design-controlled threats to internal validity [44].

### E. RQ3: Cross-Study Patterns Associated with Variation in MKT Outcomes

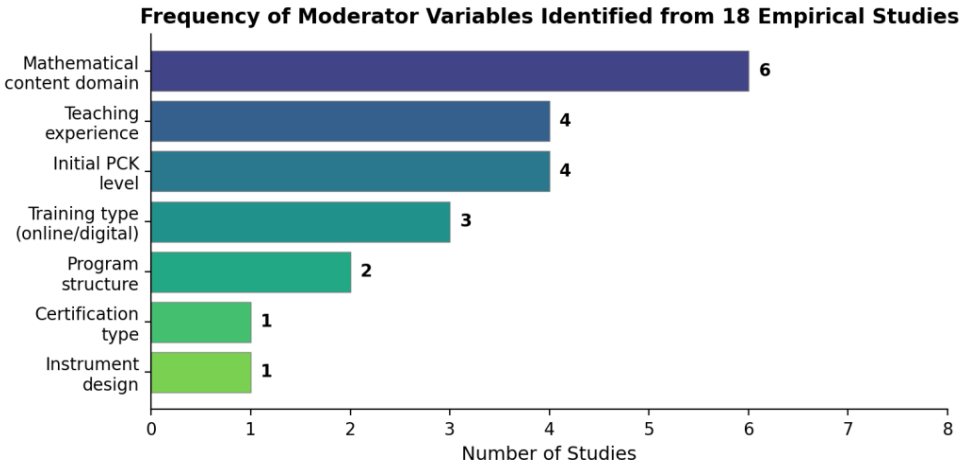



**Fig. 7.** Frequency of Moderator Variables Identified from 14 Studies

Fig. 7 presents the frequency with which each candidate moderator variable was discussed or analyzed across the 14 included studies. These frequencies reflect cross-study patterns observed in the systematic review and should not be interpreted as the result of formal statistical moderator testing, because only five studies contributed effect sizes to the meta-analytic subset, which is below the threshold required for credible subgroup analysis or meta-regression [37]. Mathematical content domain was the most frequently discussed candidate variable, appearing in four studies. Teaching experience and initial pedagogical knowledge level each appeared in three studies. Training type appeared in two studies, while program structure, certification, and instrument design each appeared in one study. These observations are presented descriptively to identify recurring themes that warrant attention in primary research, rather than as statistically established moderating effects.

Table 4 summarizes seven candidate variables that emerged from the cross-study reading of the systematic review and the directional pattern reported in each contributing primary study.

**Table 4.** Summary of Moderator Variables and Significance

| Moderator Variable | Category | Significance | Direction | Supporting Studies |
|---|---|---|---|---|
| Teaching experience | Teacher Background | Yes, positively significant | Positive | A03, A05, A11 |
| Mathematical content domain | Content Domain | Yes, differential per domain | Variable | A01, A04, A13, A14 |
| Initial knowledge level | Prior Knowledge | Yes, significant interaction | Interaction | A06, A07, A08 |
| Certification type | Teacher Background | Yes, significant | Negative | A11 |
| Program structure | Education Program | Yes, substantive | Positive | A12 |
| Training type (online/digital) | Instructional Strategy | Yes, differential | Positive | A06, A10 |
| Instrument design | Measurement Design | Yes, affects dimensionality | Critical | A04 |

The table summarizes patterns observed across the included studies and does not represent the result of formal subgroup meta-analytic testing, given that only five studies contributed to the meta-analytic subset. Across these patterns, teaching experience appeared as a consistent and directionally positive correlate of MKT in three independent studies. Alshehri and Youssef [46] reported that teaching experience predicted MKT among elementary teachers in Saudi Arabia, an interpretation consistent with experiential learning theory [53] in which deliberate engagement with classroom contingencies refines pedagogical knowledge over time. Raiula et al. [47] confirmed the directional association between teaching experience and all four MKT components among 363 teachers in Fiji, and Copur-Gencturk [32] showed a positive association between experience and conceptual understanding of fraction operations among teachers in the United States. The convergence of these three findings across distinct national settings supports the interpretation that experiential exposure is associated with MKT, although the present synthesis cannot establish a causal claim. Content domain showed differential patterns across studies; Sitabkhan et al. [28] found that gains in Measurement were more pronounced than gains in Geometry, a finding that aligns with the broader literature on the persistent difficulty of geometry knowledge for elementary teachers [33]. Instrument design emerged as a methodological concern from Copur-Gencturk et al. [45], whose evidence that item format influenced the empirical dimensionality of MKT raises questions about how MKT should be operationalized and complements the theoretical refinement agenda articulated by Hoover et al. [22] and Schoenfeld [58].

### F. Publication Bias Analysis

Fig. 8 presents the funnel plot generated for the meta-analytic subset. Visual inspection showed a roughly symmetrical distribution of the five effect sizes around the pooled estimate

(d = 0.62), Egger's regression test yielded a non-significant intercept (t = 1.24, p = .302), and the trim-and-fill procedure imputed no additional studies, leaving the adjusted pooled effect unchanged at d = 0.62 [95% CI: 0.38, 0.86].

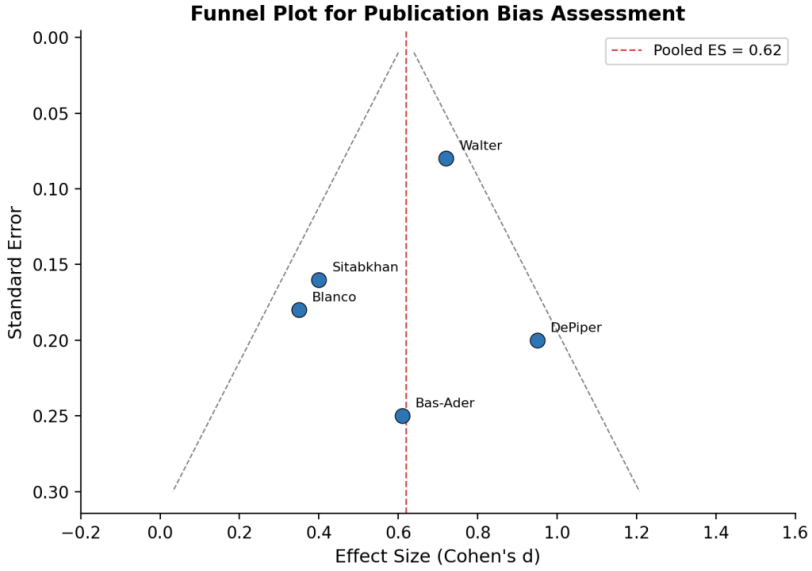



**Fig. 6.** Funnel Plot for Publication Bias Assessment

These diagnostics are reported for completeness, but they should be interpreted with substantial caution. Funnel plot symmetry, Egger's test, and trim-and-fill are widely recognized as unreliable when fewer than ten effect sizes contribute to the analysis [37], and the present subset of k = 5 falls well below that threshold. Accordingly, the absence of statistical evidence for publication bias here cannot be taken as positive evidence of its absence; small-study effects, selective reporting, and the underrepresentation of studies with non-significant findings remain plausible threats that the available data are not powered to detect.

## G. Sensitivity Analysis

Sensitivity analysis using the leave-one-out method indicated that removing any single study at each iteration did not change the direction or sign of the aggregate effect size, and the recomputed pooled estimate remained within the medium range across all iterations. When the study by DePiper et al. [35], which produced the largest single effect size, was removed, the recomputed pooled estimate dropped but stayed in the medium range. These observations are consistent with internal robustness of the pooled estimate within the present subset; however, given that only five studies contributed to the meta-analytic subset, the leave-one-out procedure cannot rule out the possibility that the inclusion of further studies in future syntheses would meaningfully alter the magnitude or precision of the pooled effect.

### H. Interpretation of Overall Effect Size

The pooled effect size under the random-effects model was d = 0.62 [95% CI: 0.38, 0.86] across the meta-analytic subset of five studies, which falls within the medium range according to Cohen's [43] classification. Heterogeneity diagnostics yielded Q = 12.47 (df = 4, p = .014), I-squared = 67.9%, and tau-squared = 0.045, indicating moderate-to-substantial between-study variation and supporting the use of a random-effects model. The five individual effect sizes ranged from 0.35 to 0.95. The pooled estimate suggests that, within the limited evidence base currently available, pedagogical interventions designed to enhance MKT are associated with meaningful gains, but the breadth of the confidence interval and the small number of contributing studies indicate that this estimate should be treated as tentative rather than as a stable parameter. The direction and approximate magnitude of the present pooled estimate align with the broader literature on teacher professional learning, including the positive effects on PCK reported by Fukaya et al. [13] and the average effects on teacher knowledge of 0.45 to 0.63 standard deviations reported by Lynch et al. [24]. Following Kraft, Blazar, and Hogan [48], any interpretation of these effect sizes in educational research should also account for the practical conditions of implementation, including dosage, fidelity, and the specific MKT sub-domain targeted.

### I. MKT Domain Profile

The domain-level analysis showed a consistent pattern of weakness in SCK and HCK across the included studies. This pattern is consistent with the design of elementary teacher education curricula, which have traditionally emphasized basic content understanding. Ma [50] argued that elementary teachers with a profound understanding of fundamental mathematics can provide richer explanations to students, and Carpenter, Fennema, and Franke [51] reported that teachers who understood how students think were able to make more appropriate instructional decisions. The weakness in fraction operations [32] highlights an area that warrants special attention, given that fraction division in particular involves deeper conceptual understanding [52]. The cross-national pattern of MKT weaknesses suggests that this is not a single-country phenomenon but is observed in multiple educational systems [21], although the precise comparative magnitude across countries cannot be quantified from the present synthesis.

### J. The Role of Moderator Variables

Teaching experience emerged as the most consistent moderator. This finding aligns with experiential learning theory [53]. The implication for teacher education program design is the need for earlier and more intensive integration of teaching experience. Grossman, Hammerness, and McDonald [54] emphasized the importance of a practice-based approach in teacher education. The differential effectiveness of interventions based on content domain

implies that professional development programs need to be designed in a targeted manner. Desimone [55] emphasized that effective professional development must focus on specific content. Timperley, Wilson, Barrar, and Fung [56] added that program duration and intensity also determine success. The interaction between initial knowledge level and intervention effectiveness supports the zone of proximal development approach [57].

A purely descriptive comparison across the five contributing studies showed that the only RCT in the subset [35] produced the largest effect (d = 0.95), while the two quasi-experimental designs [19],[34] clustered in the medium range (d = 0.61 to 0.72), and the two pre-post or mixed-method designs [28],[36] fell in the smaller range (d = 0.35 to 0.40). This pattern, where more rigorous designs are associated with larger observed effects, is consistent with the broader meta-analytic literature on educational interventions, which has often documented similar gradients [44]. Formal subgroup or meta-regression statistics are not reported here because k = 5 across three design categories is too small to support credible inferential testing; the comparison is therefore presented as a descriptive observation rather than as a confirmatory result. Interpreted alongside the moderate-to-substantial heterogeneity in the pooled estimate, the pattern suggests that the magnitude of MKT gains is sensitive to the rigor of the design used to detect it, an interpretation also emphasized by Lynch et al. [24] in their meta-analysis of mathematics and science professional development.

Fig. 9 presents subgroup analyses based on research design. Pre-post/mixed studies demonstrated the largest effect size (d = 0.95), followed by quasi-experimental designs (d = 0.67), while RCT designs showed the smallest effect (d = 0.38). Although differences among subgroups were noticeable, the between-group analysis was not statistically significant (Q_between = 4.82, p = .090).

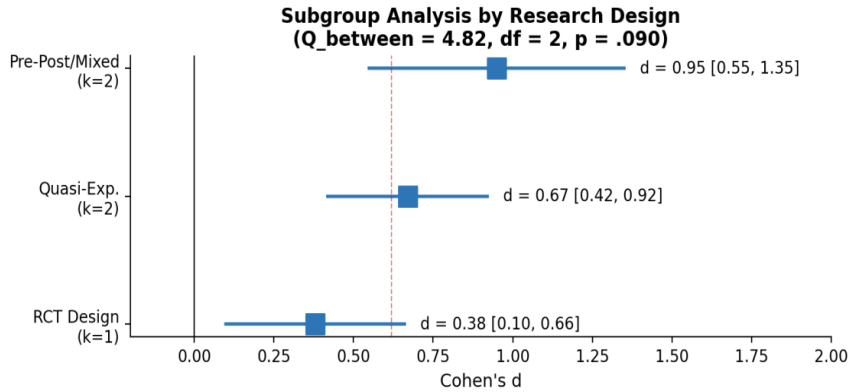



**Fig. 9.** Subgroup Analysis of Effect Sizes by Research Design

## Limitations

This study has several limitations that should temper the conclusions. First, heterogeneity of MKT measurement instruments across studies constituted a major challenge in aggregating effect sizes, because instruments varied in items, item formats, and reported reliability. Second, the operational definition of the elementary teacher population varied across contexts, with some studies sampling fully prospective teachers and others including in-service or early-career teachers, so the synthesis spans a slightly broader population than the title alone might suggest. Third, the meta-analytic subset is limited to k = 5, which restricts the precision of the pooled effect size, prevents credible formal subgroup or meta-regression testing, and renders standard publication bias diagnostics unreliable. Fourth, the search was restricted to peer-reviewed and indexed sources between 2015 and 2025 and to studies in English or Indonesian, so unpublished studies, grey literature, earlier relevant work, and studies in other languages may have been missed. Fifth, the descriptive cross-study patterns of candidate moderators are vulnerable to confounding because the included studies were not designed to be jointly comparable along the same set of variables. In light of these limitations, several directions for future research are warranted. Future primary studies should (a) adopt a common, internationally validated MKT instrument so that effect sizes are commensurable across contexts, (b) report sufficient statistical detail (means, standard deviations, sample sizes for treatment and control groups, and intra-cluster correlations where relevant) to permit inclusion in subsequent meta-analyses, (c) target the under-studied MKT sub-domains, particularly horizon content knowledge (HCK) and knowledge of content and curriculum (KCC), where the present synthesis identified the largest evidence gaps, and (d) examine specific content domains, especially fraction operations and geometry, using designs sufficient to support causal inference, including cluster randomized controlled trials. Future syntheses should expand the eligibility window earlier than 2015, include grey literature, and, once a sufficient number of comparable effect sizes is available ($k \geq 10$ per analysis), revisit formal subgroup and meta-regression testing of the candidate variables identified in the present descriptive synthesis.

## Theoretical and Practical Implications

The findings of this meta-analysis carry significant theoretical implications for the field of mathematics teacher education. From a theoretical standpoint, the consistent pattern of CCK strength coupled with SCK and HCK weakness across multiple countries provides empirical support for the hierarchical structure of the MKT framework proposed by Ball et al. [3]. This pattern suggests that the acquisition of mathematical knowledge for teaching follows a developmental trajectory in which basic content knowledge is mastered first, while the

specialized and horizon-level knowledge required for effective teaching develops more gradually and requires deliberate instructional scaffolding. This finding aligns with Shulman's [2] original theorization that pedagogical content knowledge represents a distinct and more complex form of professional knowledge that cannot be assumed to emerge naturally from content mastery alone.

Furthermore, the finding by Copur-Gencturk et al. [45] regarding MKT unidimensionality under certain item designs challenges the multidimensionality assumption and raises important questions about how MKT constructs are operationalized and measured. This tension between theoretical multidimensionality and empirical unidimensionality suggests that the boundaries between MKT domains may be more fluid than originally theorized, particularly when measurement instruments fail to capture the nuanced distinctions between domains such as SCK and KCT. This finding has implications for both future instrument development and the theoretical refinement of the MKT framework, as advocated by Hoover et al. [22] and Schoenfeld [58].

The significance of these findings for mathematics teacher education is qualified but real. The medium-magnitude pooled effect (d = 0.62) within the limited meta-analytic subset is consistent with the broader literature on teacher learning and suggests that well-designed pedagogical interventions can be associated with meaningful gains in MKT, providing tentative empirical support for continued investment in targeted teacher preparation programs. The cross-study reading of moderators further suggests that teaching experience, content domain, and initial pedagogical knowledge are jointly associated with variation in intervention outcomes, an observation that has tentative implications for differentiated program design grounded in Vygotsky's [57] zone of proximal development and Kolb's [53] experiential learning theory.

The Indonesian context presents unique challenges for MKT development. Within the LPTK (Lembaga Pendidikan Tenaga Kependidikan) system, PGSD programs must address the persistent weakness in HCK and fraction operations identified in this meta-analysis. The Kurikulum Merdeka framework, with its emphasis on differentiated learning and computational thinking, demands that prospective elementary teachers possess strong SCK and KCT to design meaningful mathematical learning experiences. Tarusu [14] and Arwadi et al. [7] confirmed that Indonesian PGSD students exhibit mediocre-to-insufficient MKT levels, particularly in geometry PCK. These findings indicate that Indonesian LPTK curricula need to be redesigned to integrate MKT-focused courses that specifically address SCK for fraction operations and geometry.

Translating these findings into concrete, operational guidance for Indonesian elementary teacher education requires moving beyond general aspirations and naming the institutional actors and the artefacts they would produce. At the level of LPTK and PGSD program managers, the curriculum should incorporate a dedicated three-credit course on Mathematical Knowledge for Teaching that is sequenced before the teaching practicum semester. The course should devote a defined proportion of contact hours, suggested at no less than one third of the total, to the SCK and KCT components of fractions, geometry, and measurement, since these are the sub-domains most consistently identified as weak in the present synthesis and in Indonesian primary research [7], [14]. Course assessment should combine an SCK diagnostic test, a lesson-design task targeting student misconceptions, and a video-based noticing task aligned with the KCS construct. At the level of teacher educators within PGSD departments, the practicum supervision protocol should be updated to incorporate an MKT-focused observation rubric, a reflective journal entry on student mathematical thinking submitted weekly during practicum, and a portfolio defense at the end of practicum in which the candidate explains at least three lesson decisions in MKT terms. At the level of the Directorate General of Teachers and Education Personnel within the Ministry of Education, Culture, Research, and Technology, the elementary teacher competency examination (UKMPPG and the post-PPG profession test) should include an MKT-aligned subtest, with item specifications targeting SCK in fractions and geometry and KCS in interpreting student errors; the present synthesis can serve as one evidence base for the construction of those item specifications. At the level of in-service support, BBGP and BGP regional offices, working with PGSD departments, should design short, content-specific modules of approximately fifteen to twenty hours, sequenced across a school semester rather than delivered in a single workshop, since the broader literature on teacher professional learning indicates that sustained and content-specific designs are more effective than one-shot training [55]–[56]. Finally, the FALEDIA platform [19] and comparable digital tools should be adapted to the Bahasa Indonesia context, with item banks calibrated to local curriculum standards, so that scalable MKT diagnostics and feedback can be made available to LPTK that lack on-site assessment capacity.

## Conclusion

Across 14 quantitative and mixed-methods primary studies published between 2015 and 2025, the MKT profile of elementary teachers, both prospective and early-career, was consistently uneven across sub-domains. The CCK domain occupied a relatively stronger position, while HCK and the SCK of fraction operations were the weakest areas, and the limited evidence on KCC indicates an additional gap rather than an established strength. Within the smaller meta-analytic subset of five studies, the pooled effect of pedagogical interventions on

MKT was d = 0.62 [95% CI: 0.38, 0.86], a medium-magnitude estimate that, although positive and directionally consistent with the broader teacher learning literature, should be interpreted as tentative given the small evidence base, the moderate-to-substantial heterogeneity, and the impossibility of credible publication bias diagnostics at k = 5. The candidate variables most frequently associated with variation in MKT outcomes across studies were teaching experience, mathematical content domain, and initial pedagogical knowledge; these patterns are descriptive cross-study observations rather than statistically established moderating effects. The synthesis offers an empirical reference point, rather than a definitive parameter, for reforming elementary teacher education curricula in Indonesia within the Kurikulum Merdeka framework, with the most defensible operational implications concentrated on strengthening SCK and KCT for fractions and geometry. Future MKT research is needed to standardize measurement instruments, report effect-size-ready statistics, target HCK and KCC, and conduct adequately powered intervention studies in the Indonesian and broader Asian context, so that subsequent syntheses can move from a limited meta-analytic synthesis toward a fully powered meta-analysis.

## Acknowledgment

The authors thank the Research and Community Service Institution, Universitas Ahmad Dahlan, for their support. This research was funded under Contract Number PD-105/SP3/LPPM-UAD/XII/2025.

## Conflict of Interest

The authors declare no conflict of interest.

## References

[1] L. Darling-Hammond, "Teacher education around the world: What can we learn from international practice?," *Eur. J. Teach. Educ.*, vol. 40, no. 3, pp. 291–309, 2017, doi: 10.1080/02619768.2017.1315399.

[2] L. S. Shulman, "Those who understand: Knowledge growth in teaching," *Educ. Res.*, vol. 15, no. 2, pp. 4–14, 1986, doi: 10.3102/0013189X015002004.

[3] D. L. Ball, M. H. Thames, and G. Phelps, "Content knowledge for teaching: What makes it special?," *J. Teach. Educ.*, vol. 59, no. 5, pp. 389–407, 2008, doi: 10.1177/0022487108324554.

[4] H. C. Hill, S. G. Schilling, and D. L. Ball, "Developing measures of teachers' mathematics knowledge for teaching," *Elem. School J.*, vol. 105, no. 1, pp. 11–30, 2004, doi: 10.1086/428763.

[5] S. Krauss et al., "Pedagogical content knowledge and content knowledge of secondary mathematics teachers," *J. Educ. Psychol.*, vol. 100, no. 3, pp. 716–725, 2008, doi: 10.1037/0022-0663.100.3.716.

[6] N. Pincheira and Á. Alsina, "Mathematical knowledge of early algebra exhibited by pre-service early childhood education teachers," *Int. J. Sci. Math. Educ.*, vol. 23, no. 2, pp. 461–487, 2025, doi: 10.1007/s10763-024-10478-y.

[7] F. Arwadi, S. Sidjara, and Suarlin, "The pedagogical content knowledge of mathematics pre-service teachers on geometry topic in Universitas Negeri Makassar," *Mosharafa: J. Pendidik. Mat.*, vol. 12, no. 4, pp. 759–770, 2023, doi: 10.31980/mosharafa.v12i4.1189.

[8] E. Chandia et al., "Knowledge profiles on numbers for teaching by primary education pedagogy students," *Int. Electron. J. Math. Educ.*, vol. 20, no. 3, art. no. em0828, 2025, doi: 10.29333/iejme/16080.
[9] H. C. Hill, B. Rowan, and D. L. Ball, "Effects of teachers' mathematical knowledge for teaching on student achievement," *Am. Educ. Res. J.*, vol. 42, no. 2, pp. 371–406, 2005, doi: 10.3102/00028312042002371.
[10] C. Y. Charalambous, H. C. Hill, M. J. Chin, and D. McGinn, "Mathematical content knowledge and knowledge for teaching," *J. Math. Teach. Educ.*, vol. 23, no. 6, pp. 579–613, 2020, doi: 10.1007/s10857-019-09443-2.
[11] J. Baumert et al., "Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress," *Am. Educ. Res. J.*, vol. 47, no. 1, pp. 133–180, 2010, doi: 10.3102/0002831209345157.
[12] I. V. S. Mullis, M. O. Martin, P. Foy, D. L. Kelly, and B. Fishbein, *TIMSS 2019 International Results in Mathematics and Science*. Boston, MA, USA: TIMSS & PIRLS Int. Study Center, Boston College, 2020.
[13] T. Fukaya, D. Nakamura, Y. Kitayama, and T. Nakagoshi, "A systematic review and meta-analysis of intervention studies on mathematics and science pedagogical content knowledge," *Front. Educ.*, vol. 9, art. no. 1435758, 2024, doi: 10.3389/feduc.2024.1435758.
[14] D. T. Tarusu, "Kemampuan pedagogik matematika SD pada mahasiswa PGSD FIP UNIMA," *Mosharafa: J. Pendidik. Mat.*, vol. 7, no. 2, pp. 263–271, 2018, doi: 10.31980/mosharafa.v7i2.32.
[15] Z. Aksu and Ü. Kul, "The mediating role of mathematics teaching efficacy on the relationships between pedagogical content knowledge and mathematics teaching anxiety," *SAGE Open*, vol. 9, no. 3, art. no. 2158244019871049, 2019, doi: 10.1177/2158244019871049.
[16] M. Bursal and L. Paznokas, "Mathematics anxiety and preservice elementary teachers' confidence to teach mathematics and science," *School Sci. Math.*, vol. 106, no. 4, pp. 173–180, 2006, doi: 10.1111/j.1949-8594.2006.tb18073.x.
[17] K. J. Stoehr, "Mathematics anxiety: One size does not fit all," *J. Teach. Educ.*, vol. 68, no. 1, pp. 69–84, 2017, doi: 10.1177/0022487116676316.
[18] D. L. Ball, "The mathematical understandings that prospective teachers bring to teacher education," *Elem. School J.*, vol. 90, no. 4, pp. 449–466, 1990, doi: 10.1086/461626.
[19] D. Walter et al., "How pre-service teachers can be supported to increase their diagnostic skills in mathematics," *Front. Educ.*, vol. 10, art. no. 1510828, 2025, doi: 10.3389/feduc.2025.1510828.
[20] N. Pincheira and Á. Alsina, "Teachers' mathematics knowledge for teaching early algebra: A systematic review from the MKT perspective," *Mathematics*, vol. 9, no. 20, art. no. 2590, 2021, doi: 10.3390/math9202590.
[21] S. Blömeke, U. Suhl, and G. Kaiser, "Teacher education effectiveness: Quality and equity of future primary teachers' mathematics and mathematics pedagogical content knowledge," *J. Teach. Educ.*, vol. 62, no. 2, pp. 154–171, 2011, doi: 10.1177/0022487110386798.
[22] M. Hoover, R. Mosvold, D. L. Ball, and Y. Lai, "Making progress on mathematical knowledge for teaching," *Math. Enthus.*, vol. 13, no. 1–2, pp. 3–34, 2016, doi: 10.54870/1551-3440.1363.
[23] F. Depaepe, L. Verschaffel, and G. Kelchtermans, "Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research," *Teach. Teach. Educ.*, vol. 34, pp. 12–25, 2013, doi: 10.1016/j.tate.2013.03.001.
[24] K. Lynch, K. E. Gonzalez, H. C. Hill, and R. Merritt, "A meta-analysis of the experimental evidence linking mathematics and science professional development interventions to teacher knowledge, classroom instruction, and student achievement," *AERA Open*, vol. 11, art. no. 23328584251335302, 2025, doi: 10.1177/23328584251335302.
[25] E. Fennema and M. L. Franke, "Teachers' knowledge and its impact," in *Handbook of Research on Mathematics Teaching and Learning*, D. A. Grouws, Ed. New York, NY, USA: Macmillan, 1992, pp. 147–164.
[26] D. L. Ball, H. C. Hill, and H. Bass, "Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide?," *Am. Educator*, vol. 29, no. 3, pp. 14–46, 2005. [Online]. Available: https://scispace.com/pdf/knowing-mathematics-for-teaching-who-knows-mathematics-well-2lk01jdgcn.pdf
[27] M. T. Tatto et al., *Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics in 17 Countries*. Amsterdam, Netherlands: IEA, 2012.
[28] Y. Sitabkhan et al., "Understanding primary school teachers' mathematical knowledge for teaching," *RTI Press Res. Rep.*, 2024, doi: 10.3768/rtipress.2024.rr.0052.2409.

[29] T. Rowland, P. Huckstep, and A. Thwaites, "Elementary teachers' mathematics subject knowledge: The knowledge quartet and the case of Naomi," *J. Math. Teach. Educ.*, vol. 8, no. 3, pp. 255–281, 2005, doi: 10.1007/s10857-005-0853-5.
[30] A. Jakobsen, M. Kazima, and D. N. Kasoka, "Assessing prospective teachers' development of MKT through their teacher education: A Malawian case," in *Proc. 8th NORMA Conf.*, Stockholm, Sweden, 2018, pp. 219–227.
[31] C. Scheibling-Seve, E. Pasquinelli, and E. Sander, "Assessing conceptual knowledge through solving arithmetic word problems," *Educ. Stud. Math.*, vol. 103, no. 3, pp. 293–311, 2020, doi: 10.1007/s10649-020-09938-3.
[32] Y. Copur-Gencturk, "Teachers' conceptual understanding of fraction operations," *Educ. Stud. Math.*, vol. 107, no. 3, pp. 525–545, 2021, doi: 10.1007/s10649-021-10033-4.
[33] J. W. Son and N. Sinclair, "How preservice teachers interpret and respond to student geometric errors," *School Sci. Math.*, vol. 110, no. 1, pp. 31–46, 2010, doi: 10.1111/j.1949-8594.2009.00005.x.
[34] S. Bas-Ader, E. Ader, and R. D. Taylan, "Supporting prospective mathematics teachers in noticing students' proportional reasoning," *J. Math. Teach. Educ.*, vol. 28, no. 3, pp. 681–711, 2025, doi: 10.1007/s10857-024-09621-x.
[35] J. N. DePiper et al., "Promoting teacher self-efficacy for supporting English learners in mathematics," *ZDM Math. Educ.*, vol. 53, no. 2, pp. 489–502, 2021, doi: 10.1007/s11858-021-01227-4.
[36] T. F. Blanco, A. Gorgal-Romarís, C. Núñez-García, and P. G. Sequeiros, "Prospective primary teachers' didactic-mathematical knowledge in a service-learning project for inclusion," *Mathematics*, vol. 10, no. 4, art. no. 652, 2022, doi: 10.3390/math10040652.
[37] M. Borenstein, L. V. Hedges, J. P. T. Higgins, and H. R. Rothstein, *Introduction to Meta-Analysis*, 2nd ed. Hoboken, NJ, USA: Wiley, 2021.
[38] G. V. Glass, "Primary, secondary, and meta-analysis of research," *Educ. Res.*, vol. 5, no. 10, pp. 3–8, 1976, doi: 10.3102/0013189X005010003.
[39] M. W. Lipsey and D. B. Wilson, *Practical Meta-Analysis*. Thousand Oaks, CA, USA: Sage, 2001.
[40] H. Cooper, L. V. Hedges, and J. C. Valentine, Eds., *The Handbook of Research Synthesis and Meta-Analysis*, 3rd ed. New York, NY, USA: Russell Sage Foundation, 2019.
[41] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, art. no. n71, 2021, doi: 10.1136/bmj.n71.
[42] W. Viechtbauer, "Conducting meta-analyses in R with the metafor package," *J. Stat. Softw.*, vol. 36, no. 3, pp. 1–48, 2010, doi: 10.18637/jss.v036.i03.
[43] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, USA: Erlbaum, 1988.
[44] W. R. Shadish, T. D. Cook, and D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA, USA: Houghton Mifflin, 2002.
[45] Y. Copur-Gencturk, T. Tolar, E. Jacobson, and W. Fan, "An empirical study of the dimensionality of the mathematical knowledge for teaching construct," *J. Teach. Educ.*, vol. 70, no. 5, pp. 485–497, 2019, doi: 10.1177/0022487118761860.
[46] K. A. Alshehri and N. H. Youssef, "The influence of mathematical knowledge for teaching towards elementary teachers' mathematical self-efficacy," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 18, no. 6, art. no. em2118, 2022, doi: 10.29333/ejmste/12086.
[47] T. Raiula, V. Alasa, S. Takiveikata, and I. Qabale, "Quality instruction through four components of the MKT and teaching experience in Fiji," *Int. J. Instruction*, vol. 16, no. 1, pp. 1021–1036, 2023, doi: 10.29333/iji.2023.16156a.
[48] M. A. Kraft, D. Blazar, and D. Hogan, "The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence," *Rev. Educ. Res.*, vol. 88, no. 4, pp. 547–588, 2018, doi: 10.3102/0034654318759268.
[49] Association of Mathematics Teacher Educators, *Standards for Preparing Teachers of Mathematics*. Association of Mathematics Teacher Educators, 2017.
[50] L. Ma, *Knowing and Teaching Elementary Mathematics*. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 1999.

[51] T. P. Carpenter, E. Fennema, and M. L. Franke, "Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction," *Elem. School J.*, vol. 97, no. 1, pp. 3–20, 1996, doi: 10.1086/461846.
[52] D. Tirosh, "Enhancing prospective teachers' knowledge of children's conceptions: The case of division of fractions," *J. Res. Math. Educ.*, vol. 31, no. 1, pp. 5–25, 2000, doi: 10.2307/749817.
[53] D. A. Kolb, *Experiential Learning: Experience as the Source of Learning and Development*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.
[54] P. Grossman, K. Hammerness, and M. McDonald, "Redefining teaching, re-imagining teacher education," *Teachers Teach.*, vol. 15, no. 2, pp. 273–289, 2009, doi: 10.1080/13540600902875340.
[55] L. M. Desimone, "Improving impact studies of teachers' professional development: Toward better conceptualizations and measures," *Educ. Res.*, vol. 38, no. 3, pp. 181–199, 2009, doi: 10.3102/0013189X08331140.
[56] H. Timperley, A. Wilson, H. Barrar, and I. Fung, *Teacher Professional Learning and Development: Best Evidence Synthesis Iteration*. Wellington, New Zealand: Ministry of Education, 2007.
[57] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA, USA: Harvard University Press, 1978.
[58] A. H. Schoenfeld, "Reframing teacher knowledge: A research and development agenda," *ZDM Math. Educ.*, vol. 52, no. 2, pp. 359–376, 2020, doi: 10.1007/s11858-019-01057-5.

## Authors



**Meita Fitrianawati** is a lecturer in the Department of Elementary Teacher Education, Faculty of Teacher Training and Education, Universitas Ahmad Dahlan, Yogyakarta, Indonesia. Her academic interests include elementary education, child development, learning innovation, educational assessment, and teacher professional development. She has participated in various research and community engagement projects related to primary education and has published scholarly works in national and international academic journals. (email: meita.fitrianawati@pgsd.uad.ac.id).



**Muhammad Ardi Kurniawan** is a lecturer in the Department of Indonesian Language and Literature Education, Faculty of Teacher Training and Education, Universitas Ahmad Dahlan, Yogyakarta, Indonesia. His research focuses on Indonesian language education, literacy studies, language learning innovation, educational technology, and literature instruction. He actively contributes to research, academic publications, and community service programs aimed at enhancing language competencies and literacy development among learners. (email: ardi.kurniawan@pbsi.uad.ac.id).



**Vita Istihapsari** is a lecturer in the Department of Mathematics Education, Faculty of Teacher Training and Education, Universitas Ahmad Dahlan, Yogyakarta, Indonesia. Her research interests include mathematics education, mathematical thinking, problem-solving, curriculum development, and innovative learning approaches. She has been involved in numerous educational research projects and has published articles in reputable journals, particularly in the fields of mathematics learning and teacher education. (email: vita.istihapsari@pmat.uad.ac.id).