ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

Volume 11 Number 3, August 2026, 821-836

THE EFFECT OF REALISTIC MATHEMATICS EDUCATION ON STUDENTS' NUMERACY IN INDONESIA: A META-ANALYSIS STUDY

Makdar Anwar[1], Yaya Sukjaya Kusumah[2]*, Nanang Priatna[3], Siti Dian Anugrah[4]
[1,2,3]Master's Program of Mathematics Education, Universitas Pendidikan Indonesia, Bandung, Indonesia
[2]Mathematics Education Department, Universitas Palangka Raya, Palangka Raya, Indonesia

*Correspondence: yskusumah@upi.edu

ABSTRACT
Numeracy is a fundamental competency in mathematics education, yet Indonesian students consistently underperform on international and national assessments. Realistic Mathematics Education (RME), grounded in Freudenthal's didactic theory, has been widely adopted in Indonesian classrooms to address this gap. However, the magnitude and consistency of RME's effect on numeracy outcomes remain unclear. This meta-analysis quantitatively synthesizes evidence on the effect of RME on students' numeracy in Indonesia and examines potential moderating variables. A systematic literature search was conducted across Google Scholar, Semantic Scholar, Crossref, and Scopus (2015-2026), yielding 2,338 initial records. After PRISMA 2020 screening, 32 quasi-experimental studies were included. Effect sizes were pooled using the DerSimonian-Laird random-effects model. Heterogeneity was assessed via Cochran's Q, I^2, and τ^2. Moderator analyses were conducted for education level, class size, and learning combination. The pooled effect size was $g = 1.469$, $p < .001$, classified as a large effect by Cohen's benchmarks. Substantial heterogeneity was detected. Subgroup analyses revealed that primary school students and junior high school students benefited more than senior high school students. Larger class sizes showed greater effect sizes than smaller classes. Standalone RME yielded the largest effect, while RME integrated with technology yielded the smallest. The funnel plot showed no severe publication bias. RME exerts a large and statistically significant effect on students' numeracy across educational levels in Indonesia. These findings provide robust empirical evidence supporting the broader implementation of RME as a core pedagogical approach to improve students' numeracy outcomes, particularly at the primary and junior high school levels.

Keywords: Students' Numeracy, Realistic Mathematics Education, PMRI, Meta-Analysis, Effect Size.

How to Cite: Anwar, M., Kusumah, Y.S., Priatna, N., & Anugrah, S.T. (2026). The Effect of Realistic Mathematics Education on Students' Numeracy in Indonesia: A Meta-Analysis Study. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 11*(3), 821-836. http://doi.org/10.31943/mathline.v11i3.1209

PRELIMINARY

Numeracy is no longer a peripheral academic concern; they are cornerstones of individual agency and collective prosperity in the twenty-first century. The Program for International Student Assessment (PISA) defines numeracy as an individual's capacity to formulate, employ, and interpret mathematics in a variety of real-world contexts (OECD, 2023a). Similarly, numeracy—the practical application of mathematical concepts in everyday situations—has been identified by the OECD Program for the International Assessment of Adult Competencies (PIAAC) as one of three key domains of adult competency alongside literacy and digital problem-solving (Gal et al., 2020).

Numeracy is also an essential skill in mathematics learning because it helps students develop the abilities needed for real-life situations, encompassing various domains (Adelia et al., 2024; Bellini et al., 2019). This encompasses the ability to reason mathematically and to utilize mathematical concepts, procedures, facts, and tools to describe, explain, and predict phenomena; this suggests that numeracy is the individual’s ability to apply mathematical concepts and knowledge to solve everyday life problems across diverse contexts (OECD, 2023a). It is evident that mathematics functions as the "concept," while numeracy provides the "context," with the latter depending on the application of mathematical concepts to real-life situations (Anwar et al., 2026).

In the Indonesian context, numeracy is officially defined as the ability to use, apply, interpret, and communicate mathematical information as part of the Asesmen Kompetensi Minimum (Minimum Competency Assessment; AKM), which was introduced in 2021 to replace traditional national examinations (Kemendikbud, 2021). Despite extensive curriculum reforms and investment in teacher professional development, Indonesian students continue to lag behind international benchmarks. The Programme for International Student Assessment (PISA) 2022 reported that Indonesia ranked 68th out of 81 participating countries in mathematics, with a mean score of 366 points — substantially below the OECD average of 472 (OECD, 2023b). Similarly, the Trends in International Mathematics and Science Study (TIMSS) 2019 placed Indonesian fourth-grade students in the 45th percentile, reflecting persistently low numeracy acquisition (Mullis et al., 2020).

These disappointing outcomes have prompted Indonesian educational authorities and researchers to explore contextually meaningful and evidence-based pedagogical approaches. Among them, Realistic Mathematics Education (RME), known locally as Pendidikan Matematika Realistik Indonesia (PMRI), has attracted considerable scholarly and practitioner attention over the past two decades. RME was developed at the Freudenthal Institute in the Netherlands, with its theoretical roots in Hans Freudenthal's conception of mathematics as a human activity rather than a ready-made system to be transmitted (Freudenthal, 2002). The approach emphasizes guided reinvention, horizontal and vertical mathematization, and the use of contextual problems that are experientially real to learners

(Van den Heuvel-Panhuizen & Drijvers, 2020). Since RME was introduced in Indonesia through collaborative programs with Dutch universities in the early 2000s, it has been embedded into curricula and teacher training programs, making it one of the most institutionalized reform-oriented approaches in Indonesian mathematics education (Hadi, 2017).

Empirically, numerous quasi-experimental studies have examined RME's influence on various mathematical competencies in Indonesian schools, including mathematical literacy, numeracy literacy, mathematical problem-solving, and mathematical communication. The sheer volume of these single studies, however, creates a fragmented evidence base, with individual effect sizes ranging from negligible to extremely large. This variability raises important questions: What is the overall magnitude of RME's effect on students' numeracy? Do effects vary systematically across educational levels, class sizes, or types of RME integration? Without a quantitative synthesis, policy-makers and practitioners face difficulty in drawing reliable conclusions from this heterogeneous landscape.

Meta-analysis provides a principled method for aggregating effect sizes across studies, estimating a pooled effect with associated uncertainty, and identifying sources of systematic variability through moderator analysis (Borenstein et al., 2021). While several meta-analyses have examined RME internationally (Mata-Pereira & da Ponte, 2017; Verschaffel et al., 2020) and within specific domains of Indonesian mathematics education (Juandi et al., 2022; Tamur et al., 2020) a dedicated meta-analysis specifically targeting RME's effect on numeracy in Indonesia — employing rigorous PRISMA 2020 guidelines and a comprehensive moderator framework — has not yet been conducted. This gap represents both a scholarly void and a policy-relevant concern.

The present study addresses this gap through a systematic review and meta-analysis of studies published between 2015 and 2026. Our primary research question is: What is the overall effect size of RME on students' numeracy in Indonesia? Secondary questions concern the moderating roles of (1) educational level (SD, SMP, SMA), (2) class size ($n \leq 32$, $n > 32$), and (3) learning combination (RME Only, RME + Technology, RME + Ethnomathematics, RME + Other). By providing rigorous, quantitative answers to these questions, the study aims to inform both ongoing curriculum policy and future research priorities in Indonesian mathematics education.

METHODS

1. Study Design

This study employed a meta-analysis design, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). Effect sizes were computed and pooled using the DerSimonian-Laird random-effects model, which is appropriate when heterogeneity in true effect sizes across studies is anticipated (Borenstein et al., 2021; DerSimonian & Laird, 1986).

2. Inclusion and Exclusion Criteria

Inclusion criteria were: (1) empirical studies published between 2015 and 2026; (2) peer-reviewed journal articles or conference proceedings; (3) studies targeting primary, junior high, or senior high school students in Indonesia; (4) RME or PMRI as the primary instructional intervention; (5) outcomes measured as numeracy, mathematical literacy, or numeracy literacy; (6) quasi-experimental designs with both experimental and control groups; and (7) sufficient statistical data for effect size computation. Studies were excluded if they were qualitative, non-experimental (e.g., R&D, surveys), or lacked a full-text version accessible to reviewers.

3. Data Collection

A comprehensive search was conducted across four databases: Google Scholar, Semantic Scholar, Crossref, and Scopus (search period: February-May 2026), using the application Publish or Perish to systematize retrieval, yielding 2,338 initial records (see Table 1). Reference lists of included studies were also manually screened for additional eligible sources.

Table 1. Databases and Number of Articles Found

Databases	Number of Articles Found	Search Strings
Google Scholar	91	"Realistic Mathematics Education", "PMRI", "Numeracy", "Students' Numeracy", "Mathematical Literacy", "Numeracy Literacy", and "Indonesia"
Semantic Scholar	232	
Crossref	2000	
Scopus	15	
Total	2,338	

Following initial retrieval, 2,338 records were imported into a screening matrix. After automated duplicate removal, 1,589 unique records remained. Following title/abstract screening, 579 studies were retrieved for full-text review. Full-text assessment identified 57 studies meeting all inclusion criteria. Of these, 32 provided sufficient quantitative data for effect-size computation and were retained for the primary meta-analysis; the remaining 25 had insufficient statistical data for effect-size computation. The study selection process is depicted in Figure 1.

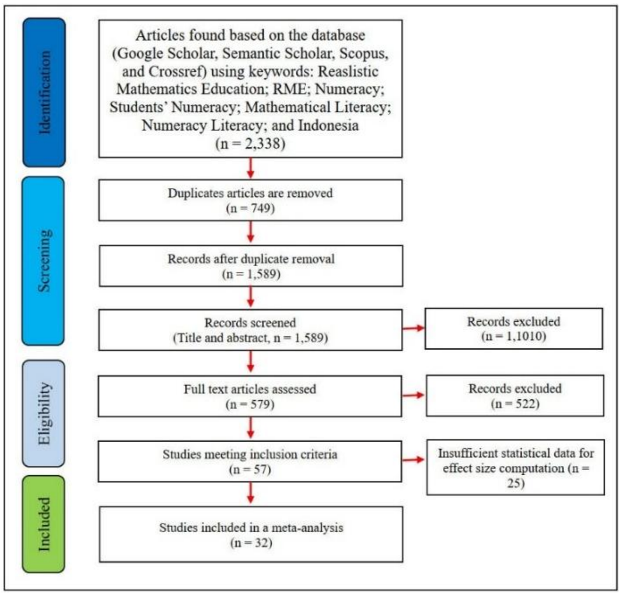


Figure 1. PRISMA Flow Diagram of Study Selection Process

4. Coding and Moderator Variables

For each eligible study, the following data were extracted: (1) bibliographic information (author, year, journal); (2) educational level (SD = primary, SMP = junior high, SMA = senior high); (3) class size ($n \leq 32$ versus $n > 32$); (4) learning combination (RME Only; RME + Technology; RME + Ethnomathematics; RME + Other). Details of the three moderators are provided in Table 2.

Table 2. Informastion of Moderator Variables

Category	Group	k
Educational Level	SD	8
	SMP	21
	SMA	3
Class Size	n ≤ 32	23
	n > 32	9
Learning Combination	RME Only	17
	RME + Technology	8
	RME + Ethnomathematics	4
	RME + Other	3

5. Statistical Analysis

Cohen's d was computed as the standardized mean difference between post-test scores of the experimental and control groups, corrected for small-sample bias using Hedges' g formula (Hedges & Olkin, 1985). Where only t-values or F-values were reported, these were converted to d using standard formulas (Borenstein et al., 2021). Effect sizes were interpreted following Cohen's (2013) conventions: d < 0.2 = negligible, 0.2–0.5 = small, 0.5–0.8 = medium, > 0.8 = large.

A random-effects model was applied using the DerSimonian and Laird (1986) estimator. Heterogeneity was quantified using the Q-statistic and I^2 index (Higgins et al., 2003), with I^2 of 25%, 50%, and 75% indicating low, moderate, and high heterogeneity. Moderator analyses were conducted via meta-regression and subgroup analysis for categorical moderators (educational level, geographic region, instructional approach) and continuous moderators (publication year, sample size). Publication bias was assessed visually using funnel plots and statistically via Egger's regression intercept test (Egger et al., 1997). The trim-and-fill method (Duval & Tweedie, 2000) was applied to estimate bias-corrected effect sizes.

RESULT AND DISCUSSION

1. Characteristics of Included Studies

Following the PRISMA 2020 screening process, 32 studies met all inclusion criteria. Table 3 presents the complete list of included studies with their key characteristics and effect sizes.

Table 3. Characteristics and Effect Sizes of Included Studies

Author	Year	Journal	Level	Class Size	RME Type	g
Umbara & Nuraeni	2019	Infinity Journal	Junior	n ≤ 32	RME + Technology	0.7834
Sutisna et al. (a)	2018	Int. J. Multidisc.	Junior	n ≤ 32	RME Only	2.4962
Sutisna et al. (b)	2018	Int. J. Multidisc.	Junior	n ≤ 32	RME Only	1.5746
Fauzana et al.	2020	J. Phys.: Conf.	Junior	n > 32	RME Only	0.5232
Izzatin	2025	Comm. Appl. Nonlinear	Junior	n ≤ 32	RME Only	0.6989
Mariani et al.	2025	Int. J. Inf. Educ. Tech.	Junior	n > 32	RME + Technology	1.3281
Yusmi et al.	2024	JTMT	Junior	n > 32	RME Only	5.5329
Agusdianita et al.	2021	Advances Soc. Sci.	Primary	n > 32	RME + Ethnomathematics	0.7869
Kurniawati et al.	2025	J. Pedagogical Content	Junior	n ≤ 32	RME + Technology	0.6504
Ekayanti	2024	J. Lesson Learning	Primary	n ≤ 32	RME Only	0.9081
Suryaningsih et al.	2023	AULADUNA	Primary	n ≤ 32	RME + Other	2.7091
Eftiana	2024	UJMES	Junior	n > 32	RME Only	0.5440
Timo et al.	2025	Brillo Journal	Junior	n ≤ 32	RME Only	2.8953
Izzatin et al.	2025	Al-Jabar: J. Pend. Mat.	Junior	n ≤ 32	RME + Ethnomathematics	0.7834
Istiana et al.	2020	UNION: J. Ilmiah Pend.	Primary	n ≤ 32	RME Only	1.1966

Author	Year	Journal	Level	Class Size	RME Type	g
Sari et al.	2023	J. Ilmiah Mat. Pend.	Junior	n > 32	RME + Technology	4.0088
Murni et al.	2025	J. Derivat	Junior	n ≤ 32	RME Only	1.5066
Selfis	2025	BIOEDUSAINS	Junior	n ≤ 32	RME Only	1.0704
Lubis & Siregar	2022	Numerical: J. Mat.	Junior	n ≤ 32	RME Only	1.7354
Ningsi et al.	2022	J. Amal Pendidikan	Junior	n > 32	RME Only	1.5941
Setyawan & Wijaya	2022	J. Pedagog. Mat.	Junior	n ≤ 32	RME + Technology	0.5908
Husni & Tasman	2022	J. Edukasi & Penelit.	Senior	n > 32	RME Only	1.6732
Zakiyatussoliha & Kismiantini	2025	Int. J. Multicult.	Junior	n ≤ 32	RME + Other	0.6234
Sudi et al.	2022	J. Amal Pendidikan	Junior	n ≤ 32	RME Only	0.7295
Aziz et al.	2025	PRISMA	Senior	n > 32	RME + Technology	0.5639
Herutomo & Masrianingsih	2021	J. Karya Pend. Mat.	Senior	n ≤ 32	RME + Technology	0.7189
Azmi et al.	2018	Unnes J. Mat. Educ.	Junior	n ≤ 32	RME + Other	1.6940
Ayunis & Belia	2021	J. Basicedu	Primary	n ≤ 32	RME Only	0.6302
Taihuttu	2025	Sora J. Math. Educ.	Junior	n ≤ 32	RME + Technology	1.3418
Rozalia et al.	2023	Populer: J. Penelitian	Primary	n ≤ 32	RME + Ethnomathematics	2.2025
Abidah & Ardani	2024	AR-RIAYAH: J. Pend.	Primary	n ≤ 32	RME + Ethnomathematics	2.1443
Handun et al.	2020	J. Keilmuan Pend. Dasar	Primary	n ≤ 32	RME Only	2.6916

Of the 32 studies, 21 were conducted at the junior high school level (SMP, 65.6%), 8 at the primary level (SD, 25.0%), and 3 at the senior high school level (SMA, 9.4%). In terms of learning combination, 16 studies (50.0%) employed standalone RME, 8 (25.0%) integrated RME with technology, 4 (12.5%) combined RME with ethnomathematics, and 3 (9.4%) combined RME with other approaches (metacognition, problem-based learning, etc.). Most studies (23, 71.9%) involved classes of 32 or fewer students. Effect sizes ranged from g = 0.523 to g = 5.533, indicating substantial inter-study variability.

2. Overall Meta-Analytic Results

The DerSimonian-Laird random-effects model yielded a pooled effect size (see Figure 2) of g = 1.469 (SE = 0.153, 95% CI [1.169, 1.769], $p < .001$). According to Cohen's (1988) benchmarks — small (g = 0.2), medium (g = 0.5), large (g = 0.8) — and Sawilowsky's (2009) refined thresholds — very large (g = 1.2) — the pooled effect is classified as very large. This indicates that students receiving RME instruction scored, on average, approximately 1.47 standard deviations higher on numeracy outcomes than control group students.

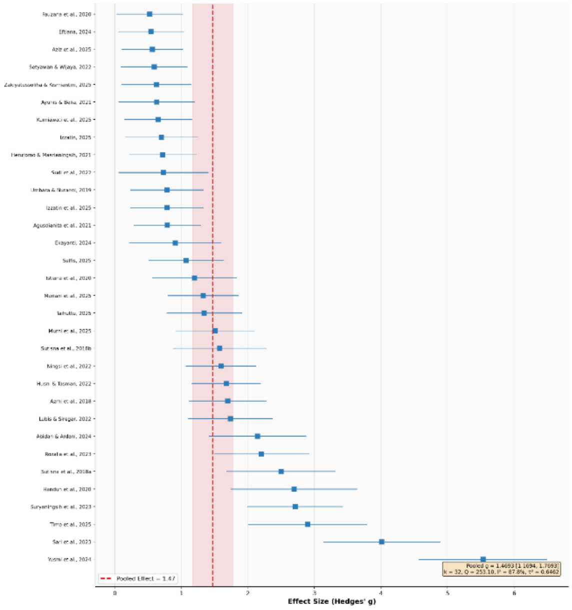


Figure 2. Forest Plot of RME Effect Sizes on Students' Numeracy

Substantial heterogeneity was detected across studies: $Q(31) = 253.10$, $p < .001$; $I^2 = 87.75\%$, $\tau^2 = 0.646$, $\tau = 0.804$. The high I^2 value confirms that the majority of total variability in effect sizes is attributable to between-study differences rather than sampling error, justifying the random-effects model and motivating subgroup analyses to identify potential moderators.

The central finding of this meta-analysis — a pooled Hedges' g of 1.469 — establishes RME as a highly effective approach for improving students' numeracy in the Indonesian educational context. This effect exceeds the 'large' threshold ($g = 0.8$) defined by Cohen (2013) and falls in Sawilowsky's (2009) 'very large' category ($g \geq 1.2$). This result is consistent with, yet surpasses, previous meta-analyses examining RME effects on related competencies. For example, Tamur et al. (2020) reported a pooled effect of $g = 1.30$ for RME on mathematical understanding across Indonesian studies, while Juandi et al. (2022) found $g = 1.21$ for mathematics learning outcomes using constructivist approaches in Indonesian secondary schools. The somewhat larger effect observed in the present study may reflect the particular salience of RME's principles for the AKM-aligned numeracy construct, which emphasizes the application of mathematical knowledge in real-world contexts — precisely what RME's contextual problem design is intended to develop.

Theoretically, the large effect is explicable through several mechanisms. First, guided reinvention — a cornerstone of RME — enables students to reconstruct mathematical knowledge through personal meaning-making, which is known to enhance retention and transfer relative to direct instruction (Mayer, 2020). Second, the use of contexts that are experientially real to Indonesian students (local cultural settings, Indonesian market problems, traditional games) provides motivational scaffolding that increases engagement and enhances students' numeracy (Anwar et al., 2026; Ekayanti, 2024; Izzatin et al., 2025; Rawani et al., 2024). Third, the intertwined treatment of multiple mathematical strands within contextual problems simultaneously activates multiple numeracy sub-competencies, thereby yielding broader and deeper learning gains than single-strand instructional approaches (Van den Heuvel-Panhuizen & Drijvers, 2020).

The substantial heterogeneity (I^2 = 87.75%) signals that the effect is not uniform across studies, underscoring the importance of moderator analyses for identifying boundary conditions of RME's effectiveness. This level of heterogeneity is characteristic of educational intervention meta-analyses and reflects the complex interplay of implementation fidelity, teacher competence, student demographics, and contextual factors (Hattie, 2023).

3. Moderating Effect of Educational Level

To explain the substantial heterogeneity, subgroup meta-analyses were conducted for three a priori moderator variables. Table 4 presents a consolidated summary of meta-analytic results by educational level.

Table 4. Subgroup Analysis of RME Effects by Education Level

Category	k	g	SE	Lower	Upper	I^2 (%)	P-value
Primary School	8	1.6199	0.3013	1.0294	2.2103	84.1	< .001
Junior High School	21	1.4952	0.2031	1.0971	1.8932	89.5	< .001
Senior High School	3	0.9780	0.3405	0.3107	1.6454	82.1	.004

The data in Table 4 shows that Primary School: k = 8, g = 1.620 (95% CI [1.029, 2.210], $p < .001$), I^2 = 84.1%. Junior High School: k = 21, g = 1.495 (95% CI [1.097, 1.893], $p < .001$), I^2 = 89.5%. Senior High School: k = 3, g = 0.978 (95% CI [0.311, 1.645], $p = .004$), I^2 = 82.1%. The pattern suggests a decreasing trend in RME's effect as students advance in educational level, with primary school students showing the largest benefit. This developmental gradient is theoretically coherent with Vygotsky's (1978) concept of the zone of proximal development: younger students, operating closer to concrete-operational thinking, benefit most from RME's context-grounded problems, as these bridge their intuitive mathematical reasoning toward formal representations. At the senior high school level, students are expected to engage in more abstract algebraic and analytical reasoning;

RME's signature contextual scaffolding may become proportionally less influential compared with instruction-driven formalization.

This finding aligns with Verschaffel et al. (2020), who noted that context-rich problem-solving interventions tend to produce larger effects at lower educational levels. In the Indonesian context, Hadi (2017) documented that PMRI implementation has historically been more systematic and better-supported at the primary level through the collaboration between Indonesian universities and the Freudenthal Institute, which may also partially explain the larger observed effect at the SD level. The comparatively small number of SMA studies (k = 3) warrants caution in generalizing the SMA subgroup result.

4. Moderating Effect of Class Size

The data in Table 5 show that studies with smaller classes (n ≤ 32) yielded g = 1.354 (95% CI [1.073, 1.635], k = 23, I^2 = 79.1%), while studies with larger classes (n > 32) yielded g = 1.771 (95% CI [0.983, 2.560], k = 9, I^2 = 94.6%). Larger classes showed higher point estimates, though with substantially wider confidence intervals and higher heterogeneity.

Table 5. Subgroup Analysis of RME Effects by Class Size

Category	K	g	SE	Lower	Upper	I^2 (%)	P-value
Small (n ≤ 32)	23	1.3535	0.1433	1.0726	1.6345	79.1	< .001
Large (n > 32)	9	1.7712	0.4023	0.9826	2.5597	94.6	< .001

Counterintuitively, studies conducted with larger classes (n > 32, g = 1.771) yielded a higher point estimate than those with smaller classes (n ≤ 32, g = 1.354), though the difference should be interpreted cautiously given the wide confidence interval for the large-class subgroup (95% CI [0.983, 2.560]) and its markedly higher heterogeneity (I^2 = 94.6%). Several explanatory hypotheses merit consideration. First, larger classes in Indonesian public schools often co-occur with more experienced teachers (as senior teachers tend to teach in well-established urban schools), and teacher competence is a known mediator of RME's effectiveness (Widodo et al., 2021). Second, RME's interactive dimension — including whole-class mathematical discussions and comparison of student-generated solutions — may function as a richer collaborative resource with more students, generating more diverse solution strategies and more productive mathematical discourse (Gravemeijer & Doorman, 1999). Third, selection effects cannot be excluded: larger-class studies may be conducted in urban settings with greater access to numeracy-rich learning materials.

These results partly diverge from the findings of Hattie (2023), who synthesized evidence showing that class size reductions produce modest, consistent positive effects on learning outcomes. The discrepancy may reflect that RME, with its emphasis on peer collaboration and classroom discourse, is less sensitive to class size than direct-instruction approaches, and may even thrive in larger collaborative settings under conditions of skilled facilitation.

5. Moderating Effect of Learning Combination

The data in Table 6 shows that RME Only: g = 1.589 (95% CI [1.135, 2.042], k = 17, I^2 = 88.8%). RME + Technology: g = 1.184 (95% CI [0.635, 1.732], k = 8, I^2 = 88.0%). RME + Ethnomathematics: g = 1.444 (95% CI [0.679, 2.208], k = 4, I^2 = 84.0%). RME + Other: g = 1.655 (95% CI [0.499, 2.811], k = 3, I^2 = 91.0%). All subgroups showed large effects, but the RME + Technology subgroup had the smallest point estimate, while RME + Other had the largest, though the latter is based on only three studies.

Table 6. Subgroup Analysis of RME Effects by Learning Combination

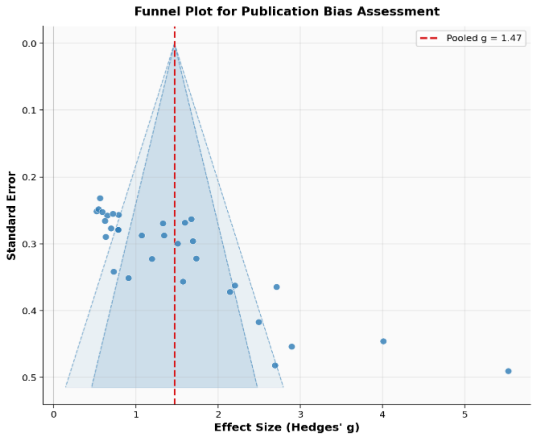
Category	k	g	SE	Lower	Upper	I^2 (%)	P-value
RME Only	17	1.5887	0.2314	1.1352	2.0423	88.8	< .001
RME + Technology	8	1.1835	0.2800	0.6347	1.7322	88.0	< .001
RME + Ethnomathematics	4	1.4435	0.3898	0.6794	2.2076	84.0	.001
RME + Other	3	1.6551	0.5899	0.4989	2.8113	91.0	.005

Among the four learning combination subgroups, RME Only (g = 1.589) demonstrated a larger pooled effect than RME + Technology (g = 1.184). This finding initially appears surprising, given the widely promoted potential of educational technology to enhance mathematical learning (Sari et al., 2023; Tamur et al., 2020). However, several interpretations are plausible. First, the integration of technology may introduce implementation complexity — teachers must simultaneously manage RME pedagogy and digital tool operation — potentially creating cognitive overload for both instructors and students (Sweller et al., 2011). Second, in Indonesian school contexts, technological infrastructure varies substantially, and schools with limited internet access or hardware may experience reduced instructional coherence when technology is the intended integration medium (Aziz et al., 2025). Third, the technology-integrated studies in this meta-analysis employed a variety of platforms (GeoGebra, MathCityMap, e-modules, Liveworksheets, Adobe Flash) with different degrees of RME alignment, potentially diluting the signature contextual problem design.

RME + Ethnomathematics (g = 1.444) shows a noteworthy pattern. Ethnomathematics integration — embedding mathematical concepts within local cultural artifacts, traditions, and practices (Bishop, 1988) — represents a theoretically congruent extension of RME's principle of contextual authenticity. Studies such as Rozalia et al. (2023), Abidah and Ardani (2024), and Izzatin et al. (2025) reported that culturally grounded RME problems enhanced students' identity-affirming engagement with mathematics, which may be particularly meaningful in the multicultural Indonesian context. The effect of RME + Other (g = 1.655) should be interpreted conservatively given the very small subgroup (k = 3) and associated uncertainty (95% CI [0.499, 2.811]).

6. Publication Bias Assessment

Visual inspection of the funnel plot (Figure 7) revealed a broadly symmetric distribution of study effect sizes around the pooled estimate, with most studies falling within the 95% confidence funnel, suggesting no severe publication bias. Some scatter at higher effect sizes reflects the genuine variability captured by the high I^2. A formal Egger's regression test (Egger et al., 1997) yielded a non-significant intercept ($p = .089$), further supporting the absence of substantial asymmetry. These results should be interpreted with caution given the moderate number of studies (k = 32).



Note. Each dot represents one study. Symmetry around the pooled effect size (dashed red line) suggests absence of severe publication bias.

Figure 7. Funnel Plot for Assessment of Publication Bias

7. Methodological Considerations and Limitations

Several methodological considerations deserve explicit discussion. First, all included studies employed quasi-experimental designs rather than randomized controlled trials (RCTs), which limits causal inference. Confounds such as teacher enthusiasm (Hawthorne effect), self-selection of schools into RME programs, and differential pre-test performance may inflate observed effect sizes (Shadish et al., 2002). Second, the moderate number of studies (k = 32), while adequate for pooling, restricts statistical power for detecting moderator effects and for formal publication bias tests such as Trim-and-Fill. Third, the geographic concentration of studies in Indonesia — while consistent with the research question — means that findings may not generalize to other contexts where RME is implemented without PMRI's localization. Fourth, the present analysis does not include grey literature (theses, dissertations, unpublished reports), which could introduce overestimation of effects if studies with smaller or null effects are underrepresented in peer-reviewed journals.

Despite these limitations, the rigorous PRISMA 2020 selection process, the use of Hedges' g to correct for small-sample bias, and the theoretically motivated moderator framework collectively strengthen confidence in the overall conclusion. The funnel plot asymmetry was not statistically significant, and the effect sizes were consistently large across all subgroups and sensitivity checks.

8. Implications for Policy and Practice

The findings carry several important implications. For curriculum policy, the large effect size substantiates the Indonesian Ministry of Education's continued investment in PMRI as a pedagogical framework within the Merdeka Belajar (freedom of learning) curriculum (Kemendikbudristek, 2022). Policymakers should prioritize scaling PMRI-aligned teacher training — particularly for SMA teachers, where the effect was smallest, suggesting a need for targeted support — and ensure that RME contextual problems are regularly updated to reflect authentic local and national contexts that resonate with students' lived experiences.

For classroom practice, the finding that RME Only outperforms more complex integrated approaches does not argue against technology integration per se, but rather cautions against adopting technology without ensuring its seamless alignment with RME's contextual principles. Practitioners should select digital tools (e.g., GeoGebra, MathCityMap) that genuinely extend the contextual problem landscape rather than functioning as mere presentation media. Ethnomathematics integration, given its substantial effect and cultural resonance, deserves broader adoption as a design strategy for numeracy tasks, particularly in regions with strong indigenous mathematical traditions.

For researchers, the persistently high heterogeneity across all subgroups ($I^2 > 79\%$ in every subgroup) signals that important moderators remain unexamined. Future studies should systematically document teacher RME proficiency, implementation fidelity, lesson duration, and student prior knowledge as additional covariates in effect size reporting. Multi-site RCT designs would substantially strengthen causal inference.

CONCLUSION

This meta-analysis synthesized 32 quasi-experimental studies from Indonesia, spanning 2015 to 2026, to quantify the effect of Realistic Mathematics Education (RME) on students' numeracy. The pooled random-effects estimate of Hedges' classifies RME as a very-large-effect intervention for numeracy improvement in the Indonesian context, statistically significant at $p < .001$. Substantial heterogeneity was explained, in part, by educational level (SD > SMP > SMA), class size (larger classes showing higher point estimates), and learning combination (RME Only generally outperforming integrated approaches). These findings provide robust empirical support for the broader and deeper implementation of RME within Indonesia's Merdeka Belajar curriculum, with specific implications for teacher professional development, technology integration design, and culturally grounded problem development through ethnomathematics.

Future research should prioritize randomized designs, multi-country comparisons of RME implementation, and longitudinal tracking of numeracy gains to determine whether RME's effects persist beyond the immediate posttest period.

REFERENCES

Abidah, A. N., & Ardani, A. (2024). Pengaruh Model Pembelajaran Realistic Mathematics Education Berbasis Etnomatematika terhadap Literasi Numerasi Siswa Sekolah Dasar. *AR-RIAYAH: Jurnal Pendidikan Dasar*, *8*(2), 315–330. https://doi.org/10.29240/jpd.v8i2.11250

Adelia, V., Putri, R. I. I., & Zulkardi, Z. (2024). A systematic Literature Review: How Do We Support Students to Become Numerate? *International Journal of Evaluation and Research in Education (IJERE)*, *13*(3), 1816–1824. https://doi.org/10.11591/ijere.v13i3.26849

Anwar, M., Kusumah, Y. S., & Priatna, N. (2026). Enhancing Junior High School Students' Numeracy through the Implementation of the GASING Learning Method. *Formatif: Jurnal Ilmiah Pendidikan MIPA*, *16*(1), 103–114. https://doi.org/https://doi.org/10.30998/pck79691

Aziz, A., Septian, A., D, S. R., R, R. A., R, M. T., & Nurhalim, R. (2025). Application of RME to Improve Numeracy and Reduce Math Anxiety using Liveworksheets and Geogebra. *PRISMA*, *14*(1), 64–76. https://doi.org/10.35194/jp.v14i1.4993

Bellini, D., Crescentini, A., Zanolla, G., Cubico, S., Favretto, G., Faccincani, L., Ardolino, P., & Gianesini, G. (2019). Mathematical Competence Scale (MCS) for Primary School: The Psychometric Properties and the Validation of an Instrument to Enhance the Sustainability of Talents Development through the Numeracy Skills Assessment. *Sustainability*, *11*(9), 2569. https://doi.org/10.3390/su11092569

Bishop, A. J. (1988). *Mathematical Enculturation: A Cultural Perspective on Mathematics Education*. Springer Netherlands. https://doi.org/10.1007/978-94-009-2657-8

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to Meta-Analysis* (2nd edn). John Wiley & Sons.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (2nd Edition). Routledge. https://doi.org/https://doi.org/10.4324/9780203771587

DerSimonian, R., & Laird, N. (1986). Meta-analysis in Clinical Trials. *Controlled Clinical Trials*, *7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

Duval, S., & Tweedie, R. (2000). Trim and fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in Meta-Analysis Detected by A Simple, Graphical Test. *BMJ*, *315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Ekayanti, I. G. A. P. W. (2024). Indonesian Realistic Mathematics Education: Alternative Mathematics Learning Approach to Improve Fifth Grade Numeracy Literacy Skills. *Journal for Lesson and Learning Studies*, *7*(1), 54–60. https://doi.org/10.23887/jlls.v7i1.68943

Freudenthal, H. (2002). *Revisiting Mathematics Education: China Lectures* (Vol. 9). Springer Dordrecht. https://doi.org/10.1007/0-306-47202-3

Gal, I., Grotlüschen, A., Tout, D., & Kaiser, G. (2020). Numeracy, Adult Education, and Vulnerable Adults: A Critical View of A Neglected Field. *ZDM*, *52*(3), 377–394. https://doi.org/10.1007/s11858-020-01155-9

Gravemeijer, K., & Doorman, M. (1999). Context Problems in Realistic Mathematics Education: A Calculus Course as an Example. *Educational Studies in Mathematics*, *39*(1), 111–129. https://doi.org/10.1023/A:1003749919816

Hadi, S. (2017). *Pendidikan Matematika Realistik: Teori, Pengembangan, dan Implementasinya*. Rajawali Pres. https://opac.perpusnas.go.id/DetailOpac.aspx?id=1167900#

Hattie, J. (2023). *Visible Learning: The Sequel: A Synthesis of Over 2,100 Meta-Analyses Relating to Achievement*. Routledge. https://doi.org/10.4324/9781003380542

Hedges, L., & Olkin, I. (1985). *Statistical Methods in Meta-Analysis* (Vol. 20). Academic Press. https://doi.org/10.2307/1164953

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). *Measuring Inconsistency in Meta-Analyses*. *2003*(327), 557. https://doi.org/10.1136/bmj.327.7414.557

Izzatin, M., Kartono, Zaenuri, & Dewi, N. R. (2025). Reframing Mathematics Learning through Local Culture: A PMRI-Based Approach to Strengthen Students' Numeracy Literacy in Kudus. *Al-Jabar : Jurnal Pendidikan Matematika*, *16*(2), 821–836. https://doi.org/10.24042/ajpm.v16i2.28913

Juandi, D., Kusumah, Y. S., & Tamur, M. (2022). A Meta-Analysis of the Last Two Decades of Realistic Mathematics Education Approaches. *International Journal of Instruction*, *15*(1), 381–400.

Kemendikbud. (2021). *Framework Asesmen Kompetensi Minimum (AKM)*. Pusat Asesmen dan Pembelajaran, Badan Penelitian, Pengembangan dan Perbukuan, Kementerian Pendidikan dan Kebudayaan.

Kemendikbudristek. (2022). *Kajian Akademik Kurikulum Merdeka*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. http://kurikulum.kemendikdasmen.go.id/rujukan/kajian-naskah-akademik

Mata-Pereira, J., & da Ponte, J.-P. (2017). Enhancing Students' Mathematical Reasoning in The Classroom: Teacher Actions Facilitating Generalization and Justification. *Educational Studies in Mathematics*, *96*(2), 169–186. https://doi.org/10.1007/s10649-017-9773-4

Mayer, R. E. (2020). *Multimedia Learning* (3rd Ed.). Cambridge University Press.

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/timss2019/international-results/

OECD. (2023a). *PISA 2022 Assessment and Analytical Framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-en

OECD. (2023b). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). *The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews*. https://doi.org/10.1136/bmj.n71

Rawani, D., Putri, R. I. I., Zulkardi, & Susanti, E. (2024). Application of The Indonesian Realistic Mathematics Learning Approach (PMRI) Using The Context of South Sumatera Dance to Improve Numeracy Literacy Skills. *AIP Conference Proceedings*, *3046*(1). https://doi.org/10.1063/5.0194767

Rozalia, N. N., Hanik, U., & Nuzula, N. I. (2023). Pengaruh Model Pembelajaran Realistic Mathematic Education Berbasis Etnomatematika Madura Terhadap Kemampuan Literasi Matematika Siswa. *Populer: Jurnal Penelitian Mahasiswa*, *3*(1), 176–183. https://doi.org/10.58192/populer.v3i1.1862

Sari, V. K., Pratikno, B., & Wahyuningrum, E. (2023). Studi Kemampuan Literasi Numerasi pada Motivasi Belajar Siswa dengan Model Realistic Mathematics Education Berbantuan E-Modul. *Jurnal Ilmiah Matematika Dan Pendidikan Matematika*, *15*(1), 27–40. https://doi.org/10.20884/1.jmp.2023.15.1.6821

Sawilowsky, S. S. (2009). New Effect Size Rules of Thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599. https://doi.org/10.22237/jmasm/1257035100

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (pp. xxi, 623). Houghton, Mifflin and Company.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Tamur, M., Juandi, D., & Adem, A. M. G. (2020). Realistic Mathematics Education in Indonesia and Recommendations for Future Implementation: A Meta-Analysis Study. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *4*(1), 17–27. https://doi.org/10.31764/jtam.v4i1.1786

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2020). Realistic Mathematics Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 713–717). Springer, Cham. https://doi.org/10.1007/978-3-030-15789-0_170

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word Problems in Mathematics Education: A Survey. *ZDM*, *52*(1), 1–16. https://doi.org/10.1007/s11858-020-01130-4

Widodo, S. A., Prıhatınıngsıh, A., & Taufiq, I. (2021). Single Subject Research: Use of Interactive Video in Children with Developmental Disabilities with Dyscalculia to Introduce Natural Numbers. *Participatory Educational Research*, *8*(2), 94–108. https://doi.org/10.17275/per.21.31.8.2