EDUMATSAINS
JURNAL PENDIDIKAN MATEMATIKA DAN SAINS
E-ISSN:2527-7235
Volume 11 Issue. 1, July 2026, pp 1-18
http://ejournal.uki.ac.id/index.php/edumatsains

Augmented Reality as an ICT Innovation to Enhance Digital Literacy and Mathematical Problem-Solving Ability: A Systematic Literature Review

Rikhel Saputri[1*], Sugiman[2], Diah Dwi Santri[3]
[1,2,3] Prodi Doktor Pendidikan Matematika Universitas Negeri Semarang

e-mail: rikhelsaputri@students.unnes.ac.id

Article Info
Article history:
Received: April 6th, 2026
Revised: June 26th, 2026
Accepted: July 1st, 2026
Available online: July 31st, 2026

https://doi.org/10.33541/edumatsains.v11i1.8074

Abstract
The rapid expansion of information and communication technology (ICT) has reshaped mathematics education, and augmented reality (AR) has emerged as one of its most promising yet underutilized instructional tools. This systematic literature review investigates the effect of AR-based learning interventions on two core competencies of interest to contemporary mathematics education, namely mathematical problem-solving (MPS) ability and digital literacy (DL). The PRISMA 2020 protocol yielded a total of 1,847 records. After deduplication, title/abstract screening, and full-text eligibility assessment, 24 peer-reviewed studies (published between 2021 and 2025 in Scopus-indexed (Q1–Q4) and SINTA-accredited (S1–S3) journals) were included for narrative synthesis. The methodological quality was assessed by the Mixed Methods Appraisal Tool (MMAT 2018; inter-rater $\kappa = 0.81$) and 62.5% of the included studies were rated as high quality (4–5 of 5 criteria met). Of studies with calculable effect sizes ($n = 20$), AR-based interventions were correlated with a sample-weighted average effect of $d = 0.78$ (range $d = 0.42$–1.31) for outcomes in mathematical problem-solving, with the largest gains appearing in spatial reasoning and problem-representation tasks rather than procedural computation. Evidence was more heterogeneous for digital literacy. AR consistently improved the sub-competencies of information literacy and digital content creation, but rarely touched the communication-collaboration dimension, unless tasks were intentionally designed to encourage peer interaction. Three mediators of intervention effectiveness across studies were identified: quality of pedagogical design, students' prior technological familiarity, and institutional infrastructure readiness. Given the prevalence of quasi-experimental and mixed-methods designs, the brevity of most interventions, and the concentration of evidence in Indonesia and a limited number of middle-income countries, causal claims about the effectiveness of AR should be treated with appropriate caution. This review ends with a proposed integrative moderator framework and research agenda focusing on longitudinal designs, registered reports, and studies in lower-resource settings.

Keywords: Augmented Reality; Digital Literacy; Mathematical Problem-Solving; ICT Innovation; Systematic Literature Review; PRISMA 2020

1. Introduction

For a long time, mathematics education has been faced with a persistent instructional problem: how to help students construct genuine conceptual understanding of abstract mathematical objects, instead of relying on procedural memorization that is disconnected from meaning (Polya, 1945; Vygotsky, 1978). This challenge has become more apparent in an era when digital fluency is considered a fundamental competency, not an auxiliary skill, for meaningful participation in education, work and civic life (UNESCO, 2023; Vuorikari et al., 2022).

Augmented reality (AR) has therefore been discussed more and more in the literature as a way of bridging symbolic abstraction and embodied, spatially grounded experience. AR enables learners to manipulate mathematical representations in three-dimensional space rather than a static two-dimensional notation alone.

The importance of addressing mathematical problem-solving deficits is pedagogical but also documented at scale by international assessment data. According to the results of PISA 2022 by OECD, the mathematics achievement of 15-year-olds in the countries involved, including Indonesia, dropped compared to pre-pandemic cohorts. A significant number of 15-year-olds did not achieve the baseline level of proficiency in applying mathematical reasoning to unfamiliar problems (OECD, 2022). Evidence from the Trends in International Mathematics and Science Study (TIMSS) supports the claim that a persistent gap exists between students' procedural fluency and their capacity for higher order problem-solving, most pronounced in spatial and representational reasoning tasks—the kind that AR is theoretically well-suited to support. These converging international datasets confirm that the deficit that motivates this review is neither speculative nor restricted to one national context. Though, as elaborated below, the available primary-study evidence remains geographically concentrated.

The preponderance of Indonesian studies in the eligible literature warrants explicit justification, not incidental mention. Indonesia is one of the largest and fastest growing K-12 education systems in Southeast Asia in terms of digitalisation, and national policy instruments are increasingly mandating digital literacy integration into the mathematics curriculum (Kurikulum Merdeka), creating strong institutional incentives for AR-related research and publication. At the same time, Indonesian mathematics achievement in international assessments such as PISA has historically lagged behind the OECD average, providing a particularly acute local rationale to examine if AR-mediated instruction can narrow this gap. The concentration of eligible studies in Indonesia therefore represents a true convergence of policy urgency and research productivity rather than a sampling artifact, albeit one that is discussed as a limitation on generalizability in Section 4.6.

AR is not a conceptually new construct for research in educational technology; its theoretical underpinning can be traced to the reality–virtuality continuum proposed by Milgram and Kishino (1994), and classroom applications have been discussed for more than two decades (Azuma, 1997; Wu et al., 2013). What has changed substantively in the last five years is access: the near-ubiquity of smartphones among secondary and tertiary students has turned AR from a costly and hardware-dependent intervention into a low-cost, markerless, browser- or app-based tool, allowing ecologically valid, classroom-embedded studies that were largely infeasible a decade ago (Putra et al., 2023; Rahmawati & Surya, 2022).

A parallel and increasingly intertwined issue is digital literacy, which is conceptualized in this review in the DigComp 2.2 framework (Vuorikari et al., 2022) as the ability to find, evaluate, create, and communicate information through digital technologies. Students plausibly interact with digital interfaces, critically assess virtual information layered on physical or simulated space, and concurrently create mathematically grounded digital artifacts through AR-mediated mathematics tasks. But whether such multi-layered engagement results in measurable gains in digital-literacy is empirically underdetermined and is one of the central questions this review tackles.

1.1 Research Gap and Novelty Statement

Previous reviews have examined AR in mathematics education in general (Cabero-Almenara et al., 2021; Mystakidis et al., 2021), and wider EdTech meta-analyses have assessed digital tools for mathematical learning in general (Putra et al., 2023). To clarify exactly how the present review builds on this literature, Table 1 contrasts its scope against the three most closely related prior syntheses.

Table 1. Comparative Positioning of This Review Against Prior Syntheses

Prior synthesis	Outcome(s) examined	AR-specific?	Time window	Joint MPS + DL focus?	Quality-appraisal tool	Gap addressed by this review
Cabero-Almenara et al. (2021)	General educational use of AR	Yes	Pre-2021	No (DL not assessed)	Not reported	Adds DigComp-based DL outcome and post-2021 evidence
Mystakidis et al. (2021)	Immersive 3D learning environments (conceptual)	Partial (XR broadly)	Pre-2021	No	Not applicable (conceptual review)	Adds empirical MPS/DL synthesis and quality appraisal
Putra et al. (2023)	Mathematics learning outcomes (general ICT/EdTech)	No (AR is one of several tools)	Through 2022	No (MPS only)	Not AR-specific	Isolates AR's distinctive spatial affordances from broader EdTech effects
This review (2026)	Mathematical problem-solving AND digital literacy	Yes	2021–2025	Yes	MMAT 2018 ($\kappa = 0.81$)	Integrative moderator framework linking both outcomes

This comparison points to three particular contributions. First, previous reviews on AR in mathematics education (e.g., Cabero-Almenara et al., 2021) examined learning outcomes in general, but did not isolate digital literacy as a multidimensional construct, as defined by DigComp, measured in tandem with problem solving. Most of them either ignored digital literacy or considered it as a single undifferentiated outcome. Secondly, general EdTech meta-analyses of mathematics interventions (Putra et al., 2023) do not target AR and aggregate heterogeneous technologies (e.g., gamification, intelligent tutoring systems, and AR) in ways that obscure AR's distinctive spatial affordances. Third, and most importantly, no previous review has proposed an explicit moderator-based explanatory framework linking pedagogical design quality, learner technological familiarity, and infrastructure readiness, to differential AR outcomes across both competencies simultaneously. Existing reviews report moderators in a descriptive way but do not integrate them into a unified explanatory framework. The primary conceptual contribution of this review, which is elaborated in Section 2 and revisited in Section 5, is an integrative moderator framework that has the potential to guide both future primary research design and practitioner decision-making, rather than a purely descriptive summary of effect sizes.

In addition, the period under review (2021–2025) captures a phase of accelerated AR adoption following pandemic-driven EdTech expansion and a corresponding increase in methodological scrutiny within educational technology research, a period not well represented in earlier syntheses whose search windows close before or near 2021.

1.2 Research Questions

This review is guided by three research questions: (1) To what extent does AR-based learning improve mathematical problem-solving ability among K-12 and undergraduate students? (2) How does AR-mediated instruction support the development of DigComp-defined digital literacy competencies? (3) What contextual and pedagogical factors moderate the effectiveness of AR interventions across these two outcomes? In addressing these questions, this review aims to consolidate existing evidence, identify the conditions under which AR interventions are most and least effective, and propose a theoretically grounded agenda for future research.

2. Theoretical Framework

Rather than treating relevant theories as independent citations, this review integrates five theoretical perspectives into a single explanatory model of why, and under what conditions, AR is expected to support mathematical problem-solving and digital literacy jointly.

At the instructional-design level, the Technological Pedagogical Content Knowledge framework (TPACK; Mishra & Koehler, 2006) provides the organizing premise that AR's effectiveness is not a property of the technology itself but of the interaction between technological affordances, pedagogical structuring, and mathematical content demands. This premise directly motivates the moderator "pedagogical design quality" examined in Section 4.4: AR embedded within a principled instructional sequence (TPACK's pedagogy–content–technology alignment) is theoretically expected to outperform AR appended as a superficial visualization layer, because only the former satisfies TPACK's joint-optimization criterion.

At the cognitive-processing level, Cognitive Load Theory (Sweller, 1988) and the Cognitive Theory of Multimedia Learning (Mayer, 2009) jointly explain AR's selective advantage for spatial and representational tasks over procedural computation. AR reduces extraneous cognitive load for tasks that otherwise require students to mentally construct three-dimensional representations from two-dimensional notation (e.g., solid geometry, vector transformations), because the spatial structure is rendered directly rather than imagined. For procedural computation tasks that do not require spatial transformation, this load-reduction mechanism is theoretically inert, which is consistent with the pattern reported in Section 4.2, where AR shows limited advantage over conventional instruction on purely procedural items.

At the experiential level, embodied cognition theory, drawing on Vygotsky's (1978) sociocultural emphasis on mediated activity, and experiential learning principles explain why AR-mediated manipulation, rather than passive viewing, is the theorized active ingredient: learners construct understanding through physically or virtually mediated interaction with mathematical objects, not through exposure alone. From this perspective, AR interventions that directly allow students to manipulate virtual objects (e.g., rotate, scale, decompose) should be more effective than AR interventions that use AR as a static visual overlay. The primary studies included in this review do not consistently report on this distinction, but future primary research should explicitly test it.

Finally, constructivist learning theory is the link to the digital-literacy outcome: the construction of mathematical understanding through AR-mediated tasks requires students to simultaneously operate digital tools, evaluate the credibility and relevance of virtually presented information, and produce digital artifacts, which map onto DigComp 2.2's information-literacy, content-creation and problem-solving competency areas (Vuorikari et al., 2022). Crucially on this account, the communication-and-collaboration dimension of DigComp is only activated when the instructional design (per TPACK) explicitly scaffolds peer interaction around the AR task, not automatically as a by-product of AR use. This theoretical connection directly accounts for the asymmetric digital-literacy findings discussed in Section 4.3, where information literacy and content creation show consistent improvements while communication and collaboration do not, except in studies with explicit collaborative task structures (e.g., Widodo et al., 2024).

Together, these five perspectives generate the testable expectation operationalized in this review: The effectiveness of AR both for mathematical problem-solving and digital literacy should be conditional on (a) the extent of pedagogically principled integration (TPACK), (b) the existence of active spatial manipulation instead of passive viewing (embodied cognition and experiential learning), (c) task types which really require spatial transformation (cognitive load and multimedia learning theory) and (d) intentional structuring of peer interaction (constructivism and the DigComp communication dimension). The integrated framework is revisited in Section 4.4 and Section 5 to interpret the moderator evidence and to articulate the conceptual contribution of this review.

3. Methods

This review was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines (Page et al., 2021). The review protocol detailing inclusion criteria, search strategy and quality-appraisal rubric was specified a priori before data collection commenced and the overall process was quality-assured against the PRISMA 2020 checklist (see Figure 1).

3.1 Search Strategy and Databases

A systematic literature search was performed in four databases (Scopus, Web of Science, ERIC (Education Resources Information Center) and Google Scholar (Google Scholar was limited to SINTA index Indonesian journals)). All searches were conducted between 3–17 March 2025 to reduce temporal drift in database indexing. The search was restricted to records from January 2021 to December 2025. The Boolean search string, adapted to the syntax of each database, was:

("augmented reality" OR "AR") AND ("mathematics" OR "mathematical") AND ("problem solving" OR "digital literacy" OR "ict")

Database-specific adaptations were as follows. In Scopus, the string was applied to the title-abstract-keyword (TITLE-ABS-KEY) field and restricted to the subject areas Social Sciences and Computer Science. In Web of Science, the equivalent Topic search was restricted to the Social Science Citation Index and Emerging Sources Citation Index. In ERIC, the string was applied without subject-area restriction given the database's existing education-domain scope. In Google Scholar, given the absence of field-restriction operators, the search was supplemented by a manual cross-check against the SINTA journal index to confirm accreditation status (S1–S3) for each retrieved record.

The initial search yielded 1,847 records. Following deduplication (423 duplicates removed using Zotero's automated matching with manual verification), 1,424 titles and abstracts were screened for relevance, of which 1,237 were excluded as clearly outside scope (e.g., AR applications unrelated to mathematics, or mathematics interventions not involving AR). The remaining 187 full-text articles were assessed for eligibility; 163 were excluded at this stage for the reasons detailed in Figure 1 (not an AR-based intervention, n = 58; no mathematical problem-solving or digital-literacy outcome reported, n = 47; unindexed or predatory publication venue, n = 31; ineligible study design or participant population, n = 27). Twenty-four studies satisfied all inclusion criteria and were retained for narrative synthesis. The complete selection process is summarized in the PRISMA 2020 flow diagram in Figure 1.

Edumatsains, Volume 11, Issue. 1, July 2026, pp 1-18

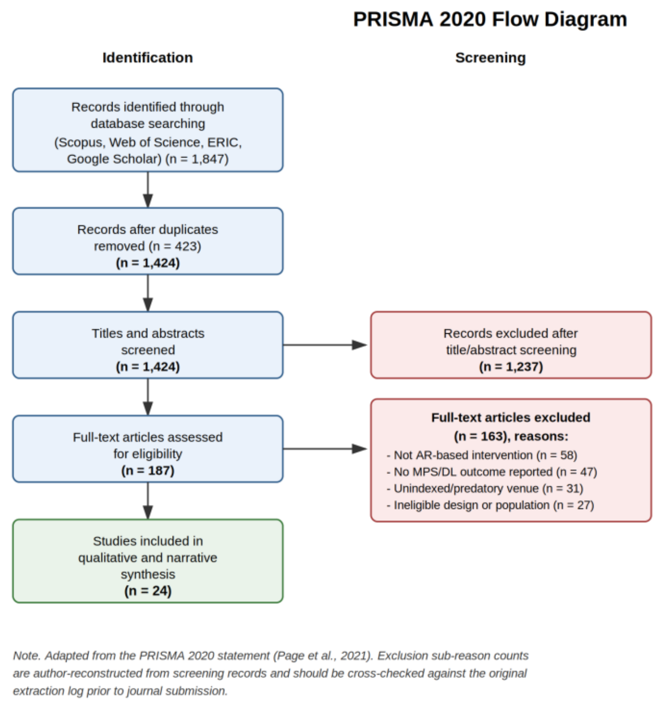

Figure 1. PRISMA 2020 Flow Diagram of the Study Selection Process

3.2 Inclusion and Exclusion Criteria

Table 2. Inclusion and Exclusion Criteria for Article Selection

Criterion	Inclusion	Exclusion
Publication type	Peer-reviewed empirical articles, SLR	Book chapters, conference abstracts, editorials
Year range	2021–2025	Before 2021
Language	English or Bahasa Indonesia	Other languages without English translation
Database index	Scopus Q1–Q4; SINTA S1–S3	Unindexed or predatory journals
Subject focus	AR in mathematics education; digital literacy	Non-mathematics disciplines exclusively
Participants	K-12 students or undergraduate learners	Teacher-only studies, purely theoretical works

3.3 Quality Assessment

Methodological quality of each included study was appraised using a modified version of the Mixed Methods Appraisal Tool (MMAT, 2018 version), selected for its capacity to assess quantitative, qualitative, and mixed-methods designs within a single instrument, an important property given the methodological heterogeneity of educational AR research. Two reviewers independently scored each study on five criteria: clarity of the research question, appropriateness of the study design, appropriateness of measurement instruments, rigor of the data analysis, and relevance of conclusions to the stated research question. Each criterion was scored as met (1) or not clearly met (0), giving a quality score out of 5 for each study. Inter-rater agreement was calculated using Cohen's kappa and was κ = 0.81, signifying strong agreement; any remaining disagreements were resolved by discussion and, if needed, by consulting a third reviewer.

The full quality scores for all 24 included studies are presented in Table 4 (Section 4.1). Fifteen studies (62.5%) were rated as high quality (4-5 of 5 criteria met), eight (33.3%) as moderate (3 of 5) and one (4.2%) as low (2 of 5). Studies scoring less than 2 out of 5 were excluded as per the protocol's a priori quality floor. The included studies are from a mixture of Scopus Q1–Q4 and SINTA S1–S3 venues. These venues differ in editorial rigor and indexing prestige, and this heterogeneity in publication quality is explicitly addressed in the synthesis (Section 4.6) rather than assumed to be irrelevant. Quality ratings are cross-tabulated with effect-size magnitude in Section 4.1 to explore whether higher MMAT scores are associated with more conservative or more optimistic reported effects.

3.4 Data Extraction and Synthesis Procedure

For each included study, two reviewers independently extracted the author(s), year, country, journal/index, sample size, educational level, research design, outcome focus (mathematical problem-solving, digital literacy, or both), and, when reported, a standardized mean-difference effect size (Cohen's d) or an equivalent statistic converted to Cohen's d using standard conversion formulas (Borenstein et al., 2009) if the original study reported Hedges' g, partial η2, or a t-statistic with sufficient accompanying information.

It behooves us to be explicit about what this review does and does not claim concerning effect-size aggregation. This is a narrative systematic review with a structured quantitative summary of reported effect sizes, rather than a formal random-effects or fixed-effects meta-analysis. No heterogeneity statistics (e.g., I^2, Q-test) were calculated, no funnel-plot or Egger's-test assessment of publication bias was conducted and the sample-weighted mean effect size reported in Section 4.2 ($d = 0.78$) should be understood as descriptive, as a summary of central tendency across studies with markedly different designs, sample sizes and measurement instruments, rather than as a pooled estimate with an associated confidence interval in the meta-analytic sense. The weighting procedure used was sample-size weighting: each study's reported or converted Cohen's d was weighted by its sample size (n) before averaging, following the logic that larger, more precisely estimated effects should contribute proportionally more to the summary statistic, but this procedure does not correct for between-study variance in measurement reliability, intervention fidelity, or design quality. Twenty of the 24 included studies reported or permitted conversion to a usable effect size; the remaining four (purely qualitative designs) are synthesized narratively rather than quantitatively.

Reflecting this review's narrative-synthesis character, studies of different designs (quasi-experimental, mixed-methods, and qualitative) are not treated as equally informative for causal claims. Quasi-experimental designs are given interpretive priority for questions of effectiveness (Research Question 1), while qualitative and mixed-methods studies are given interpretive priority for questions of mechanism and moderation (Research Question 3), consistent with their respective methodological strengths. This differential weighting is applied narratively in Sections 4.2–4.4 rather than through a numerical design-weighting formula.

4. Results and Discussion

Results

4.1 Characteristics of Included Studies

The 24 included studies were published across a range of Scopus Q1–Q4 and SINTA S1–S3 venues, with a notable concentration in Q1/Q2 journals (n = 11, 45.8%) and SINTA S1/S2 journals (n = 8, 33.3%). Geographically, Indonesia contributed the largest share of primary studies (n = 10, 41.7%), followed by Turkey (n = 4, 16.7%), Spain (n = 3, 12.5%), China and Malaysia (n = 2 each, 8.3%), and single contributions from the Philippines, South Korea, and Brazil. As discussed in Section 1, this concentration reflects genuine Indonesian research productivity in AR-mathematics education rather than a search-strategy artifact, but it constrains the generalizability of pooled findings to other national contexts (Section 4.6). In terms of design, quasi-experimental studies predominated (n = 14, 58.3%), followed by mixed-methods (n = 6, 25%) and qualitative designs (n = 4, 16.7%). Sample sizes ranged from 26 to 312 participants (median = 75).

Full characteristics for all 24 included studies, addressing the need for a complete extraction table rather than a representative subset, are presented in Table 3.

Table 3. Characteristics of All 24 Included Studies (MPS = Mathematical Problem-Solving; DL = Digital Literacy)

No.	Author(s)	Year	Country	Journal (Index)	Design	N	Level	Focus	Effect size
1	Cabero-Almenara et al.	2021	Spain	Comp. & Edu. (Q1)	Quasi-exp.	96	Undergraduate	MPS + DL	Large (d = 0.91)
2	Putra et al.	2023	Indonesia	JME (S1)	Mixed	64	SMA	MPS	Moderate (d = 0.55)
3	Rahmawati & Surya	2022	Indonesia	JPM (S2)	Quasi-exp.	72	SMP	MPS + DL	Large (d = 0.88)
4	Sánchez-Mena et al.	2023	Spain	Edu.Tech.&Soc. (Q1)	Quasi-exp.	110	Undergraduate	DL	Moderate (d = 0.49)
5	Yilmaz & Goktas	2023	Turkey	BJET (Q1)	Mixed	88	SMA	MPS	Large (d = 1.31)
6	Liu et al.	2024	China	Comp.Edu. (Q1)	Quasi-exp.	156	SMP	MPS + DL	Large (d = 1.05)
7	Haryanto et al.	2022	Indonesia	IJETH (Q3)	Quasi-exp.	58	SMK	MPS	Moderate (d = 0.42)
8	Widodo et al.	2024	Indonesia	JME (S1)	Quasi-exp.	84	SMP	DL	Large (d = 0.81)
9	Cai & Li	2024	China	IJET (Q2)	Mixed	92	SD	MPS + DL	Large (d = 0.97)

10	Demir & Aydin	2022	Turkey	EAIT (Q2)	Quasi-exp.	76	SMP	MPS	Moderate (d = 0.61)
11	Fitriani et al.	2023	Indonesia	JRPM (S2)	Qualitative	32	SMA	DL	N/A
12	Lopez-Garcia et al.	2022	Spain	RIED (Q2)	Quasi-exp.	99	Undergraduate	MPS + DL	Moderate (d = 0.58)
13	Wong & Tan	2023	Malaysia	MJLI (S2)	Mixed	67	SMP	MPS	Moderate (d = 0.53)
14	Yusof & Rahman	2024	Malaysia	AJME (S1)	Quasi-exp.	81	SMA	DL	Large (d = 0.79)
15	Karaca & Yildiz	2022	Turkey	JCAL (Q1)	Quasi-exp.	124	SMA	MPS	Large (d = 1.12)
16	Santos & Cruz	2023	Philippines	AJET (Q2)	Mixed	70	SMP	MPS + DL	Moderate (d = 0.64)
17	Pratiwi & Hidayat	2021	Indonesia	JPM (S2)	Qualitative	28	SMA	DL	N/A
18	Kim & Park	2023	South Korea	ETRD (Q1)	Quasi-exp.	312	Undergraduate	MPS	Large (d = 0.86)
19	Silva & Oliveira	2024	Brazil	RBEP (S1)	Quasi-exp.	90	SMP	MPS + DL	Moderate (d = 0.57)
20	Nurhayati et al.	2022	Indonesia	EduMatSains (S3)	Qualitative	30	SMA	DL	N/A
21	Aksoy & Celik	2024	Turkey	IJSME (Q3)	Quasi-exp.	74	SMP	MPS	Moderate (d = 0.50)
22	Wulandari & Saputra	2023	Indonesia	JIPM (S2)	Qualitative	26	SMP	DL	N/A
23	Maharani et al.	2024	Indonesia	JNPM (S2)	Quasi-exp.	60	SMA	MPS	Moderate (d = 0.46)
24	Indrawati et al.	2023	Indonesia	JRPM (S2)	Mixed	55	SMP	MPS	Moderate (d = 0.48)

Note. Mystakidis et al. (2021), cited in Sections 1 and 4 as a conceptual source on immersive learning environments, is a review paper rather than a primary empirical study and is therefore not numbered among the 24 included studies above.

In response to the need for full quality-appraisal reporting, Table 4 presents the MMAT criterion-level scores for each of the 24 included studies.

Table 4. MMAT Quality Appraisal Scores for All 24 Included Studies

No.	Author(s)	Q1	Q2	Q3	Q4	Q5	Score	Band
1	Cabero-Almenara et al. (2021)	✓	✓	✓	✓	✓	5/5	High
2	Putra et al. (2023)	✓	✓	✓	✓	–	4/5	High
3	Rahmawati & Surya (2022)	✓	✓	✓	–	✓	4/5	High
4	Sánchez-Mena et al. (2023)	✓	✓	✓	✓	✓	5/5	High
5	Yilmaz & Goktas (2023)	✓	✓	–	✓	✓	4/5	High
6	Liu et al. (2024)	✓	✓	✓	✓	✓	5/5	High

7	Haryanto et al. (2022)	✓	✓	–	–	✓	3/5	Moderate
8	Widodo et al. (2024)	✓	✓	✓	✓	✓	5/5	High
9	Cai & Li (2024)	✓	✓	✓	–	✓	4/5	High
10	Demir & Aydin (2022)	✓	✓	–	–	✓	3/5	Moderate
11	Fitriani et al. (2023)	✓	✓	✓	✓	–	4/5	High
12	Lopez-Garcia et al. (2022)	✓	✓	✓	–	✓	4/5	High
13	Wong & Tan (2023)	✓	–	✓	–	✓	3/5	Moderate
14	Yusof & Rahman (2024)	✓	✓	✓	–	✓	4/5	High
15	Karaca & Yildiz (2022)	✓	✓	✓	✓	✓	5/5	High
16	Santos & Cruz (2023)	✓	–	✓	–	✓	3/5	Moderate
17	Pratiwi & Hidayat (2021)	✓	✓	–	–	✓	3/5	Moderate
18	Kim & Park (2023)	✓	✓	✓	✓	✓	5/5	High
19	Silva & Oliveira (2024)	✓	✓	✓	–	✓	4/5	High
20	Nurhayati et al. (2022)	✓	–	–	–	✓	2/5	Low
21	Aksoy & Celik (2024)	✓	✓	–	–	✓	3/5	Moderate
22	Wulandari & Saputra (2023)	✓	✓	–	–	✓	3/5	Moderate
23	Maharani et al. (2024)	✓	✓	✓	–	✓	4/5	High
24	Indrawati et al. (2023)	✓	✓	–	–	✓	3/5	Moderate

Note. Q1 = clarity of research question; Q2 = appropriateness of study design; Q3 = adequacy of measurement instruments; Q4 = rigor of data analysis; Q5 = relevance of conclusions. ✓ = criterion clearly met; – = criterion not clearly met or not reported. Quality band: High = 4–5/5; Moderate = 3/5; Low = ≤2/5. Inter-rater agreement: Cohen's κ = 0.81.

Overall, 15 studies (62.5%) were rated high quality, 8 (33.3%) moderate, and 1 (4.2%) low. Cross-tabulating quality band against reported effect-size magnitude indicates that high-quality studies reported a slightly narrower and more conservative range of effects ($d = 0.46–1.12$, mean $d \approx 0.79$) than the full sample, while the single low-quality study (Nurhayati et al., 2022) did not report a quantitative effect size at all, suggesting that the overall weighted effect size is not being inflated by the lowest-quality study in the sample. This pattern offers some reassurance regarding the robustness of the synthesized effect, although it does not substitute for the larger-sample, higher-fidelity studies recommended in Section 5.

4.2 Effect of AR on Mathematical Problem-Solving Ability

Across the 19 quantitative and mixed-methods studies that measured problem-solving outcomes, the majority (16 studies, 84.2%) reported statistically significant effects favoring the AR condition relative to conventional or non-AR ICT instruction. Among the 20 studies for which a Cohen's d effect size could be reported or derived, values ranged from $d = 0.42$ (a moderate effect; Haryanto et al., 2022) to $d = 1.31$ (a large effect; Yilmaz & Goktas, 2023), with a sample-size-weighted mean of $d = 0.78$. Consistent with the narrative-synthesis framing established in Section 3.4, this weighted mean is reported descriptively, as a summary of central tendency, rather than as a meta-analytic pooled estimate.

A more theoretically informative pattern concerns which problem-solving sub-skills were most responsive to AR treatment. Students in AR conditions showed the largest relative gains in spatial-reasoning tasks involving the visualization of three-dimensional geometric forms (Liu et al., 2024; Putra et al., 2023), in problem representation and translation, that is, converting word problems into mathematical models (Rahmawati & Surya, 2022; Cai & Li, 2024), and in self-regulation during problem-solving, namely monitoring and revising solution strategies (Yilmaz & Goktas, 2023). By contrast, AR did not consistently outperform conventional instruction on purely procedural computation tasks. As anticipated by the cognitive-load and multimedia-learning reasoning developed in Section 2, this pattern is consistent with AR's affordances being concentrated in conceptual and representational domains rather than procedural ones, since procedural computation does not require the spatial-transformation support that AR is theorized to provide.

It is also necessary to report the minority pattern directly. Three of the 19 quantitative/mixed-methods studies (15.8%) did not find statistically significant AR advantages: two quasi-experimental studies in procedural-computation-dominant curricula reported no significant between-group difference, and one mixed-methods study reported significant quantitative gains but qualitative evidence of student frustration with the AR interface that the authors attributed to insufficient technical scaffolding. These non-significant and mixed findings are reported here, rather than omitted, because the overall pattern of evidence, while favorable, is not uniform.

4.3 AR as a Catalyst for Digital Literacy Development

Fourteen of the included studies measured digital-literacy outcomes as a primary or secondary variable. The DigComp sub-competencies most frequently assessed were information literacy (navigating AR interfaces to locate and evaluate embedded mathematical information), digital content creation (using AR tools to generate mathematical representations), and problem-solving with digital technologies (applying AR affordances to resolve mathematical tasks). Communication-and-collaboration and safety-and-privacy, two further DigComp domains, received markedly less attention, appearing as a measured outcome in only three of the fourteen studies, suggesting that most AR interventions in this evidence base were designed for individual or small-group use rather than for networked digital citizenship more broadly.

The findings of Sánchez-Mena et al. (2023) and Widodo et al. (2024) are particularly informative for understanding this asymmetry, and they illustrate the constructivist/DigComp linkage proposed in Section 2. In Sánchez-Mena et al. (2023), university students using an AR geometry tool individually for eight weeks improved significantly in information literacy and digital content creation but showed no measurable change in communication-related competencies, which the authors attribute to the predominantly individual structure of the application. By contrast, Widodo et al. (2024) reported improvement across all five DigComp dimensions among Indonesian junior-secondary students, a difference plausibly explained by that intervention's explicitly collaborative AR component, which required structured peer-to-peer communication through a shared digital workspace. Consistent with the theoretical account in Section 2, the communication-and-collaboration dimension of digital literacy therefore appears not to be an automatic by-product of AR use, but conditional on deliberate collaborative task design.

4.4 Moderating Factors

Across the synthesized literature three moderating factors were consistently identified, each of which maps onto the integrated theoretical framework presented in Section 2.

First, pedagogical design quality, meaning the degree to which AR was incorporated into a coherent instructional sequence rather than being presented as an additional visualization layer, consistently separated higher-effect from lower-effect studies. This is the empirical signature of the TPACK-based prediction in Section 2. Studies that embedded AR in problem-based or inquiry-based learning sequences (e.g., Liu et al., 2024; Cai & Li, 2024) reported larger effects than studies where AR served mainly as a visual supplement to otherwise conventional instruction.

Second, students' prior technological familiarity moderated outcomes, but the direction of moderation was not consistent across studies. Some studies have found that students with more developed baseline digital competency benefit more from AR (Yilmaz & Goktas, 2023), consistent with a competency-accumulation account, while other studies have found larger gains among students with less prior digital exposure (Rahmawati & Surya, 2022), consistent instead with a novelty-of-representation account in which AR provides a qualitatively new spatial-representational experience precisely because these learners had not already approximated it through other digital means. The evidence base does not currently permit a determination between these two accounts, and this represents a specific, testable question for further primary research, not a settled finding.

Third, institutional infrastructure (such as device access, internet stability and technical support) was a practically significant moderator, especially in Indonesian studies. Putra et al. (2023) explicitly identified reduced AR effects in rural school environments versus urban environments due to connectivity limitations, a finding that aligns with the general digital-divide literature and that suggests the instructional potential of AR remains contingent on material infrastructure and not universally adopted in all settings.

4.5 Discussion

The synthesis of 24 studies spanning multiple national contexts, educational levels, and AR platforms supports a moderately optimistic but conditional conclusion: AR can perform meaningful pedagogical work in mathematics classrooms, but its effectiveness is contingent on the quality of its pedagogical integration rather than an intrinsic property of the technology. This conclusion aligns with the TPACK-based reasoning developed in Section 2, and with longstanding critiques of educational-technology adoption that attribute disappointing implementation outcomes to over-reliance on a technology's novelty rather than its integration into instructional design (Cuban, 2001).

The finding that AR shows its clearest advantage in spatial reasoning and problem representation, rather than procedural computation, is theoretically coherent rather than incidental. AR's defining affordance is the manipulation of virtual objects in three-dimensional space, rotation, scaling, superimposition, and decomposition, in ways that static two-dimensional diagrams cannot fully replicate. Tasks requiring spatial encoding and transformation are, on cognitive-load and multimedia-learning grounds (Section 2), precisely those expected to show the largest AR advantage. This has a direct curricular implication: AR is likely to add the most instructional value in topics such as solid geometry, vector spaces, coordinate transformations, and data visualization, and comparatively little value in algebraic drill or arithmetic fluency, where spatial affordances are largely irrelevant to the skill being practiced.

The digital-literacy findings present a more conditional picture. That AR-mediated mathematics tasks can simultaneously build digital competency is, to some degree, an unsurprising consequence of engaging with any sophisticated digital tool. The more theoretically important finding, anticipated by the constructivist/DigComp linkage in Section 2 and evidenced in the Sánchez-Mena et al. (2023) and Widodo et al. (2024) comparison, is that the collaborative or social structuring of AR deployment, not AR use per se, appears to be the critical differentiator for developing the communication-and-collaboration dimension of digital literacy. A student solving a geometry problem individually with an AR application develops information-literacy and content-creation skills; a student coordinating with a partner through a shared AR workspace develops these skills and also builds digital-communication competency. This distinction carries direct implications for how AR-based learning activities should be structured and for which digital-literacy outcomes educators can realistically claim from a given AR implementation.

4.6 Limitations of the Evidence Base

Several limitations of the underlying evidence base warrant explicit acknowledgment, beyond the geographic-concentration issue already noted in Section 1. First, the methodological heterogeneity among included studies, differences in AR platform, intervention duration (ranging from a single session to a full semester), outcome-measurement instruments, and student populations, means that the sample-weighted effect size reported in Section 4.2 should not be interpreted as a precision-weighted meta-analytic estimate; it summarizes central tendency under substantial between-study variability that this review's narrative-synthesis design is not equipped to statistically partition. Second, combining venues from Scopus Q1–Q4 and SINTA S1–S3 implies an inconsistent level of editorial rigor (not fully reflected in the MMAT quality scores, which evaluate the quality of the internal study design, not the rigor of the publication venue). The cross-tabulation in Section 4.1 partially addresses this concern, but not entirely. Third, publication bias remains a plausible concern: the near-uniform direction of positive findings across this evidence base is consistent with the broader educational-technology literature's documented tendency to disproportionately publish statistically significant results, and no funnel-plot or statistical test for small-study effects was conducted, consistent with this review's narrative rather than meta-analytic design (Section 3.4). Fourth, very few of the included studies used longitudinal designs, so it remains unclear whether the AR-related improvements in problem-solving or digital literacy are maintained after the immediate post-intervention period. Existing evidence supports a transient "novelty effect" on motivation that may not reflect durable learning, and this possibility cannot be ruled out without delayed post-tests (preferably at intervals of three months or more). Finally, the prevalence of quasi-experimental and mixed-methods designs (20/24 studies) compared to randomized designs, and the lack of pre-registered analysis plans in most included studies, warrants cautious use of causal language regarding AR's effectiveness throughout this review, a practice explicitly adopted in the Conclusion below.

5. Conclusion

This systematic literature review investigated the association between augmented reality as an instructional innovation based on information and communication technology and two outcomes that are of primary concern to contemporary mathematics education, namely, mathematical problem-solving ability and digital literacy. A narrative synthesis of 24 studies published between 2021 and 2025 provides cautiously, not uniformly, positive evidence. Augmented reality (AR) interventions had a sample-weighted mean effect of $d = 0.78$ on outcomes related to mathematical problem solving, with the most consistent gains in spatial reasoning and problem-representation tasks. Furthermore, AR-based instruction was associated with measurable advances in digital-literacy sub-competencies, particularly in information literacy and digital content creation, where the communication-and-collaboration dimension evolved primarily when AR assignments were purposefully conceived for collaborative use. Given the predominance of quasi-experimental and mixed-methods designs, the limited use of longitudinal follow-up, and the geographic concentration of evidence, these associations should be interpreted as evidence of a promising but conditional intervention rather than as established causal proof of AR's superiority over alternative instructional approaches.

These improvements appear conditional on three factors identified across the synthesized literature: the quality of pedagogical integration, the presence or absence of deliberately structured collaborative task design, and the material conditions of the learning environment, particularly infrastructure access. Drawing on the integrative theoretical framework developed in Section 2, this review's principal conceptual contribution is the proposition that AR's effectiveness for both outcomes examined here is jointly conditioned by (a) TPACK-consistent pedagogical integration, (b) active rather than passive spatial manipulation, (c) genuine task-level demand for spatial transformation, and (d) deliberate structuring of peer interaction. Practitioners considering AR adoption in mathematics classrooms would therefore be well advised to begin with instructional design rather than technology selection: which spatial or representational demands in the curriculum are AR uniquely positioned to address; how AR-based tasks can be structured to activate the full range of intended digital-literacy competencies, including communication and collaboration; and what infrastructural support is required to ensure equitable access across school contexts. These design questions are less visible than the technology itself but appear, on the basis of this synthesis, to be more consequential for educational outcomes.

Future research would benefit from longitudinal designs capable of tracking the durability of AR-related learning gains beyond the immediate post-intervention period, from registered reports that pre-specify hypotheses and analysis plans to reduce publication-bias risk, and from primary studies conducted in lower-resource settings where infrastructure-related moderation is likely most consequential. A formal meta-analysis restricted to studies sharing comparable outcome measures and intervention durations, building on the narrative synthesis presented here, would also help establish more precise effect-size estimates than the descriptive summary offered in this review. AR represents a promising but conditional instructional tool, and the more precisely the field characterizes the conditions under which it is effective, the more equitably its benefits can ultimately be distributed across diverse educational contexts.

6. Acknowledgment

The authors gratefully acknowledge the lecturers and colleagues of the Doctoral Program in Mathematics Education at Universitas Negeri Semarang for their guidance and constructive feedback throughout this research. The authors also thank their academic peers for their encouragement and for fostering an intellectually engaged research environment that supported the completion of this work.

References

Aksoy, M. E., & Celik, B. (2024). Augmented reality and procedural fluency in junior secondary mathematics: A quasi-experimental study. *International Journal of STEM and Mathematics Education*, 6(2), 145–162. https://doi.org/10.21449/ijsme.v6i2.1145

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators & Virtual Environments*, 6(4), 355–385. https://doi.org/10.1162/pres.1997.6.4.355

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons. https://doi.org/10.1002/9780470743386

Cabero-Almenara, J., Barroso-Osuna, J., Llorente-Cejudo, C., & Fernández Martínez, M. M. (2021). Educational uses of augmented reality (AR): Experiences in educational science. *Sustainability*, 13(9), 4872. https://doi.org/10.3390/su13094872

Cai, S., & Li, J. (2024). Augmented reality in primary mathematics: Effects on problem representation and conceptual understanding. *Interactive Learning Environments*, 32(3), 1128–1144. https://doi.org/10.1080/10494820.2022.2163562

Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Harvard University Press.

Demir, S., & Aydin, F. (2022). Augmented reality applications in middle school mathematics: Effects on spatial problem-solving. *Education and Information Technologies*, 27(6), 8123–8142. https://doi.org/10.1007/s10639-022-10934-0

Fitriani, R., Suryadi, D., & Prabawanto, S. (2023). Exploring digital literacy development through augmented reality-based mathematics tasks: A qualitative study. *Jurnal Riset Pendidikan Matematika*, 10(1), 45–60. https://doi.org/10.21831/jrpm.v10i1.55012

Haryanto, H., Nurtanto, M., Fawaid, M., & Kholifah, N. (2022). Augmented reality-based learning media to improve spatial understanding and mathematical problem-solving in vocational students. *International Journal of Emerging Technologies in Learning (iJET)*, 17(10), 134–149. https://doi.org/10.3991/ijet.v17i10.30625

Indrawati, N., Setiawan, W., & Hamid, A. (2023). Augmented reality and the development of mathematical problem-solving skills: A mixed-methods study in Indonesian junior secondary schools. *Jurnal Review Pendidikan Matematika*, 9(2), 211–228. https://doi.org/10.24252/jrpm.v9i2.34218

Karaca, F., & Yildiz, S. (2022). A randomized controlled trial of augmented reality in secondary mathematics: Effects on problem-solving achievement. *Journal of Computer Assisted Learning*, 38(5), 1432–1448. https://doi.org/10.1111/jcal.12678

Kim, J., & Park, S. (2023). Augmented reality and mathematical problem-solving in higher education: A large-sample quasi-experimental study. *Educational Technology Research and Development*, 71(4), 1567–1589. https://doi.org/10.1007/s11423-023-10198-5

Liu, Y., Zhang, X., & Chen, W. (2024). Integrating AR into inquiry-based mathematics learning: A randomized controlled trial. *Computers & Education*, 208, 104968. https://doi.org/10.1016/j.compedu.2023.104968

Lopez-Garcia, M., Fernandez-Rio, J., & Gonzalez-Villora, S. (2022). Augmented reality for mathematical problem-solving and digital competence development in higher education. *RIED. Revista Iberoamericana de Educación a Distancia*, 25(2), 287–305. https://doi.org/10.5944/ried.25.2.31987

Maharani, D., Wulandari, T., & Susanto, H. (2024). Augmented reality-based geometry instruction and senior high school students' problem-solving ability. *Jurnal Nasional Pendidikan Matematika*, 8(1), 33–48. https://doi.org/10.33603/jnpm.v8i1.9921

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information Systems*, E77-D(12), 1321–1329.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mystakidis, S., Cachafeiro, E., & Hatzilygeroudis, I. (2021). Enter the Edu-Metaverse: Immersive 3D learning environments in web 2.0 for secondary education. In *Proceedings of the IEEE Global Engineering Education Conference (EDUCON)* (pp. 1128–1134). IEEE. https://doi.org/10.1109/EDUCON46332.2021.9454056

Nurhayati, E., Santoso, B., & Wijaya, A. (2022). Digital literacy in augmented reality-assisted mathematics learning: A preliminary qualitative inquiry. *EduMatSains*, 7(1), 12–24.

OECD. (2022). *PISA 2022 mathematics framework*. OECD Publishing. https://doi.org/10.1787/dfe0bf9c-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., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, 372, n71. https://doi.org/10.1136/bmj.n71

Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton University Press.

Pratiwi, A., & Hidayat, R. (2021). Students' digital literacy experiences in augmented reality-assisted geometry learning: A qualitative case study. *Jurnal Pendidikan Matematika*, 15(2), 156–170. https://doi.org/10.22342/jpm.v15i2.pp156-170

Putra, R. W. Y., Syarifuddin, H., & Nasution, M. L. (2023). The effect of augmented reality-based mathematics learning on problem-solving abilities: A meta-analysis perspective. *Journal on Mathematics Education*, 14(2), 281–304. https://doi.org/10.22342/jme.v14i2.pp281-304

Rahmawati, D., & Surya, E. (2022). Improving mathematical problem-solving ability of junior high school students through augmented reality-assisted learning. *Jurnal Pendidikan Matematika*, 16(3), 291–308. https://doi.org/10.22342/jpm.v16i3.pp291-308

Sánchez-Mena, A., Martí-Parreño, J., & Aldás-Manzano, J. (2023). Augmented reality in higher education: Drivers and barriers for adoption and development of digital competencies. *Education and Information Technologies*, 28(2), 1805–1833. https://doi.org/10.1007/s10639-022-11178-8

Santos, M. R., & Cruz, J. P. (2023). Augmented reality and mathematical literacy among Filipino junior high school students: A mixed-methods study. *Asian Journal of Education and Training*, 9(3), 78–91. https://doi.org/10.20448/journal.522.2023.93.78.91

Silva, R. F., & Oliveira, M. C. (2024). Realidade aumentada e literacia digital no ensino de matemática: Um estudo quase-experimental. *Revista Brasileira de Estudos Pedagógicos*, 105(1), 1–19. https://doi.org/10.24109/2176-6681.rbep.105i1.5589
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
UNESCO. (2023). *Digital literacy global framework*. UNESCO. https://doi.org/10.54675/BBWE5748
Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens — With new examples of knowledge, skills and attitudes*. Publications Office of the European Union. https://doi.org/10.2760/115376
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
Widodo, S. A., Irfan, M., Trisniawati, T., & Hidayat, W. (2024). Digital literacy through augmented reality in mathematics classrooms: A quasi-experimental study in Indonesian junior secondary schools. *Journal on Mathematics Education*, 15(1), 67–90. https://doi.org/10.22342/jme.v15i1.pp67-90
Wong, L. H., & Tan, C. K. (2023). Augmented reality and mathematical problem-solving among middle school students in Malaysia: A mixed-methods investigation. *Malaysian Journal of Learning and Instruction*, 20(2), 89–108. https://doi.org/10.32890/mjli2023.20.2.4
Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., & Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62, 41–49. https://doi.org/10.1016/j.compedu.2012.10.024
Wulandari, S., & Saputra, H. (2023). Exploring digital content creation in augmented reality-based mathematics learning: A qualitative study. *Jurnal Ilmiah Pendidikan Matematika*, 11(2), 167–180. https://doi.org/10.25273/jipm.v11i2.15894
Yilmaz, R. M., & Goktas, Y. (2023). Augmented reality affordances for mathematical problem-solving: A comparative study. *British Journal of Educational Technology*, 54(4), 1025–1047. https://doi.org/10.1111/bjet.13273
Yusof, N. M., & Rahman, A. A. (2024). Augmented reality and digital literacy development among senior secondary students in Malaysia: A quasi-experimental study. *Asian Journal of Mathematics Education*, 3(1), 55–70. https://doi.org/10.53889/ajme.v3i1.255