M A T H L I N E
JURNAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA
ISSN 2502-5872 (Print)
ISSN 2622-3627 (Elektronik)

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

Volume 11 Number 4, November 2026, 1089-1108

AUGMENTED REALITY AND CRITICAL THINKING IN SCHOOLS: A SYSTEMATIC REVIEW AND THE ARITHMETIC GAP

Nafsia Lestarya Purba[1], Yunus Abidin[2*], Dede Trie Kurniawan[3], Mario Capestrano Gesiradja[4]
[1,2,3]Elementary School Teacher Education Study Program, Universitas Pendidikan Indonesia, Indonesia
[4]Community Education Study Program, Faculty of Education, Universitas Pendidikan Indonesia, Indonesia

*Correspondence: yunusabidin@upi.edu

Received 18 October 2026; Revised 4 November 2026; Accepted 12 November 2026

ABSTRACT

Augmented reality (AR) is promoted as a way to develop critical thinking in schools, but the evidence is scattered across subjects, designs, and measures. This review aims to map and appraise the evidence on how AR affects critical thinking among primary and secondary students, and to examine whether primary-school arithmetic, a core topic of mathematics education, has been studied. Following PRISMA 2020, we searched Scopus and ERIC (2016–2026) for open-access articles and included 27 studies: primary (n = 6), lower secondary (n = 4), upper secondary (n = 11), and vocational schools (n = 6); 17 were in science and four in mathematics. We appraised quality with the Mixed Methods Appraisal Tool (MMAT) 2018 and synthesised findings narratively. Most studies reported higher critical thinking after AR-supported lessons, yet no study clearly reported random assignment to conditions, six had no comparison group, the effect of AR was rarely separable from its accompanying pedagogy, and 24 of 27 studies met two or fewer of the five MMAT criteria. The mathematics studies addressed geometry, spatial reasoning, or trigonometry. None addressed arithmetic, so the gap concerns critical thinking in AR-supported arithmetic within the reviewed corpus. The evidence was encouraging but weak, and it clustered in visually intuitive topics. Future studies would need controlled designs with pretest adjustment and indicator-level measurement.

Keywords: Augmented Reality, Critical Thinking, Systematic Literature Review, Primary School Mathematics, Arithmetic.

How to Cite: Purba, N. L., Abidin, Y., Kurniawan, D. T., & Gesiradja, M. C. (2026). Augmented Reality and Critical Thinking in Schools: A Systematic Review and the Arithmetic Gap. *Mathline: Jurnal Matematika dan Pendidikan Matematika, 11*(4), 1089-1108. http://doi.org/10.31943/mathline.v11i4.1235

PRELIMINARY

Critical thinking is purposeful, reasoned, goal-directed thinking (Halpern & Dunn, 2021). Augmented reality (AR), which overlays virtual content on the real environment and responds in real time, is one way to make abstract content something students can question and reason about. Reviews report gains in achievement and motivation, along with problems of usability and teacher preparedness (Akçayır & Akçayır, 2017; Garzón et al., 2019).

Critical thinking is, however, hard to define and harder to measure: frameworks such as FRISCO (Ennis, 1996) organise similar ideas differently, and some instruments capture self-reported disposition rather than performance. Mixing these constructs treats different things as one outcome, so we recorded for each study what was measured and how.

Here, critical thinking is purposeful, reasoned judgement (analysing, evaluating, inferring; Halpern & Dunn, 2021; see also Ennis, 2018), and a study counted only if it measured it directly. It differs from problem solving, reasoning, spatial literacy, and the broader category of higher-order thinking skills (HOTS), which were counted only when the full text showed that components of critical thinking were measured.

Evidence on AR and higher-order thinking has begun to be pooled. Lu et al. (2025) ran a random-effects meta-analysis of 17 studies (21 effect sizes) that pooled critical thinking with other higher-order outcomes. The overall effect was large ($g = 1.109$), larger for skills than for dispositions. Narrower syntheses cover critical thinking in science (Kusumaningrum et al., 2025; eight articles, no quality appraisal) and in physics (Silitonga & Emiliannur, 2026; 24 studies, pooled effect 0.765), and reviews of AR in mathematics describe the field by topic and outcome (Ahmad & Junaini, 2020; Assidiqi & Sugiman, 2026; Islim et al., 2024; Syarifah et al., 2026). Supplementary Table S6 compares five of them with this review; none covered critical thinking across school subjects with criterion-level quality appraisal.

Three questions remain open. The first is quality: pooled estimates depend on primary studies that are often small quasi-experiments, single-group trials, or development studies in which the effect of AR cannot be separated from the accompanying pedagogy. Among the syntheses we identified, three appraised quality: Lu et al. (2025) used a 14-criterion checklist, judged all 17 studies acceptable (scores above 55%), and did not test quality as a moderator; Assidiqi and Sugiman (2026) classified evidence strength on a three-level scheme of their own, which they acknowledge was not a full risk-of-bias assessment; and Syarifah et al. (2026) applied the CASP checklist to 15 primary-school studies and judged quality adequate. The second is construct: studies differ in what they measure and in the indicator framework, so it is unclear what a pooled estimate represents. The third is coverage: existing syntheses are limited to one subject or pool outcomes across subjects without a map of topics, and Lu et al. had only two effect sizes from elementary education and no separate category for mathematics.

The third question matters for mathematics. Arithmetic is the foundation of primary mathematics, yet reviews of AR in mathematics report a concentration on geometry and

spatial topics (Assidiqi & Sugiman, 2026; Islim et al., 2024; Syarifah et al., 2026); numbers and operations accounted for only 5 of 96 studies in Islim et al. (2024). Assidiqi and Sugiman (2026), who reviewed 42 studies of AR and virtual reality, found the most consistent outcomes for spatial ability and conceptual understanding and more variable evidence for problem solving. AR has nevertheless been used for arithmetic topics, such as multiplication tables (Rebollo et al., 2022), fractions (Özdemir & Özçakır, 2019), and arithmetic games for sixth graders (Volioti et al., 2023), but these studies reported achievement, engagement, attitudes, or usability, not critical thinking. Whether any study has examined critical thinking in AR-supported arithmetic is therefore unclear, and we use that combination as a test case. Resource-limited and rural settings are a secondary, descriptive concern: AR needs devices, markers or applications, and sometimes a connection (Ismail et al., 2025).

This review maps and appraises the empirical evidence on AR and critical thinking from primary through upper secondary and vocational schooling, with primary-school arithmetic as a test case, a specific objective addressed by RQ5. Whereas Assidiqi & Sugiman (2026) reviewed cognitive outcomes of AR and virtual reality within mathematics, this review focuses on critical thinking across school subjects and appraises the methodological quality of each study with a standardized tool. It addresses five research questions (RQ):

1. How are the studies distributed by school level, subject, topic, country, and, where reported, rural or urban setting (RQ1)?
2. How is critical thinking measured, and what effects of AR are reported (RQ2)?
3. Which forms of AR and pedagogical frameworks were used, and could their effects be separated (RQ3)?
4. How sound is the methodology of the studies according to the Mixed Methods Appraisal Tool (MMAT) 2018 (RQ4)?
5. To what extent do studies of AR address critical thinking in primary-school arithmetic (RQ5)?

METHODS

We conducted a systematic review reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021) and synthesised findings narratively, because studies used different instruments, indicators, and effect measures. The protocol was not registered in advance; we deposited it, with Supplementary Files S1–S6, on the Open Science Framework after the searches

(anonymised link for review: https://osf.io/qt8x3/?view_only=113eca949f6d43fc84e60292ae296023). We searched Scopus and the Education Resources Information Center (ERIC) on 16 October 2026. The Scopus string combined three concept blocks (augmented reality; critical thinking and related higher-order terms; school level) in titles, abstracts, and keywords for 2016–2026, with article-type and open-access limits applied afterwards (54 of 130 records retained). The ERIC string was broader, pairing the AR concept with critical, higher-order, and analytical thinking, problem solving, and reasoning, without school-level terms, and was limited to peer-reviewed full-text sources from the last ten years (48 of 165 records retained). The ERIC string therefore favoured sensitivity over precision, and school level, the other eligibility criteria, and relevance to critical thinking were judged at screening, not by the string. Open access (Scopus) and full-text availability (ERIC) were applied as filters so that eligibility could be judged from full texts; the 115 records they removed (12 Scopus, 103 ERIC) were not screened (Figure 1), and we did not seek their full texts elsewhere, which may have introduced selection bias. We did not search national databases such as Garuda or SINTA, so Indonesian studies indexed only there may be missing. Full strings are in Supplementary File S2.

Studies were eligible if they were empirical journal articles in English or Indonesian, involved students in primary (SD), lower secondary (SMP), upper secondary (SMA/MA), or vocational secondary (SMK) schools, used AR as the intervention or main topic, and measured critical thinking. Higher-order thinking skills (HOTS) were counted only when the full text showed that components of critical thinking were operationalised and measured. Studies of preschool children, university students, pre-service teachers, or teachers were excluded, as were studies in which AR was only a context, in which critical thinking was not measured or a different construct was measured, and non-empirical papers (exclusion codes E1–E5, defined in the Figure 1 caption). Design was not an exclusion criterion, and self-report measures were retained but coded as such. Duplicates were removed by matching normalised titles. The first author screened titles and abstracts and assessed full texts (full-text exclusions are in Supplementary Table S3). A second reviewer (the fourth author) independently screened a random sample of 19 of the 94 records (7 of the 34 advanced to full text and 12 of the 60 excluded) and agreed with the first author on 18 (95%; Cohen's $\kappa = 0.89$).

For each study we extracted country, rural or urban setting where reported, school level, subject and topic, design and sample size, type of AR, pedagogical framework, critical

thinking instrument and indicators, main results with their statistics as reported, and stated limitations. We did not recompute statistics or convert effect sizes, because designs and instruments were not comparable; information a report did not provide was recorded as not reported (Supplementary Table S5). Quality was appraised with the MMAT 2018 (Hong et al., 2018), using the category matching each study's design (non-randomised, 22 studies, including five of six single-group studies; mixed methods, three; quantitative descriptive, one; randomised, one). We chose it because one instrument covers all designs in the corpus. Each of the five criteria was scored 1 for "yes" and 0 for "no" or "can't tell". The MMAT discourages summary scores; we report them only to compare studies at a glance (high, 4–5; moderate, 3; low, 0–2). These bands are our own convention, have no validated meaning, and do not weight or exclude studies; interpretation rests on the criterion-level ratings. The first author rated all studies. The second reviewer independently rated six randomly drawn studies (22%), stratified by school level. They agreed on the design category for five of six studies and on 14 of 25 criterion ratings (56%; $\kappa = 0.36$). The two reviewers then discussed every disagreement against the full texts and agreed consensus ratings, which are not an independent check. The first author re-checked the other 21 studies with the same decision rules (Supplementary Table S1 notes). Studies were grouped by school level, subject and topic, design, AR form, pedagogical framework, and instrument, and effects are reported as stated by the original authors, with the direction of results reported by design (Supplementary Table S4), since a gain in a single group cannot be attributed to the intervention. Results were coded positive when authors reported higher critical thinking in the AR condition or at posttest, qualified when positive only for some indicators or with an unexplained anomaly, and null when no difference was reported. These codes record what authors reported, not causal effectiveness.

Because the search was not built around arithmetic, we also ran a targeted, non-systematic web search in October 2026 for studies of AR in primary-school arithmetic without a critical thinking term and for Indonesian studies of AR and critical reasoning in primary mathematics. Its purpose was to check whether the gap reflected our critical thinking terms. Its results are context only (not screened, appraised, or part of the PRISMA flow) and cannot show that no further studies exist, so the arithmetic gap is conditional on the indexed, open-access records that our critical thinking terms retrieved. Supplementary Tables S1, S3, S4, S5, and S6 and File S2 are provided as a separate file.

RESULT AND DISCUSSION

Study Selection and Characteristics

Database filters removed 193 of 295 records before screening (115 by access restrictions, 78 by other filters; Figure 1), leaving 102; after eight duplicates were removed, 94 were screened by title and abstract and 60 were excluded, most often because critical thinking was not measured (E2, 32 records) or the record was not an empirical journal article (E4, 14). Thirty-four full texts were assessed and seven excluded, all because critical thinking was not measured or a different construct was measured (E2; Supplementary Table S3), including four studies of HOTS or spatial literacy (e.g., Pujiastuti & Haryadi, 2023). Twenty-seven studies were included (Table 1; Supplementary Table S5).

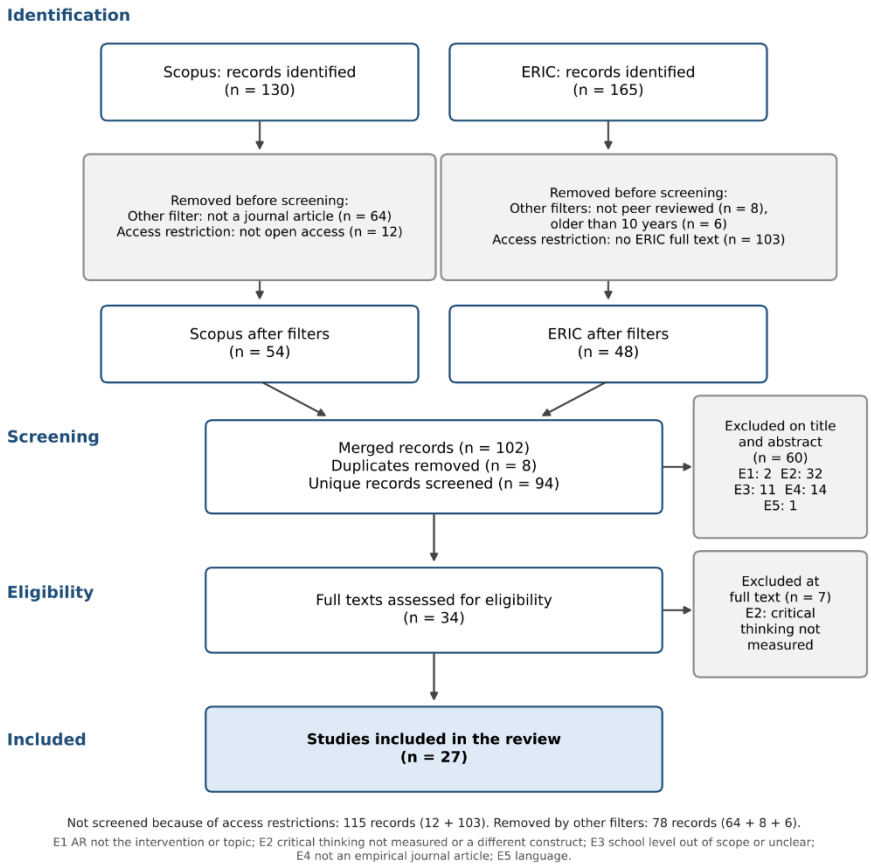


Figure 1. PRISMA 2020 Flow Diagram. E1 = AR not the intervention or topic; E2 = critical thinking not measured or a different construct; E3 = school level out of scope or unclear; E4 = not an empirical journal article; E5 = language.

The studies were published between 2021 and 2026 (nine of them in 2026, up to the search date). Only 10 of the 94 records screened dated from 2017–2020 and none was eligible, consistent with the recent growth of the field, although the open-access filter may also have removed earlier studies. Twenty-three studies (85%) came from Indonesia, so the findings describe above all the Indonesian context. Only Damopolii et al. (2022) explicitly

described its school as rural (Manokwari), but location was extracted only where reported, so this shows how often rural settings are described, not how often AR was studied there. Six studies involved primary pupils, four lower secondary, eleven upper secondary, and six vocational students; 17 concerned science, four mathematics, one Indonesian language, and five vocational or general topics. Six used a single group (pre–posttest only), one was a survey, and 20 compared AR with another condition; in two of these (Efendi et al., 2025; Rahmawati et al., 2026) the comparison was not clear from the report. No study clearly reported random assignment to conditions, so results are reported by design.

Measurement of Critical Thinking and Reported Effects

Twenty-two studies measured critical thinking with tests, rubrics, or performance tasks; five used self-report scales (Dewi et al., 2025; Dilmen & Atalay, 2021; Efendi et al., 2025; Irwanto et al., 2026, who measured disposition rather than skill; Wen et al., 2023). Indicator frameworks varied: Facione (Akbar et al., 2025; Sukatiman et al., 2026; Sukmawati et al., 2024), Ennis or FRISCO (Atmojo et al., 2025; Darmawan et al., 2026; Hayat et al., 2026), the Cornell test (Alkhabra et al., 2023), and Reynders' indicators (Rizki et al., 2024); some did not specify indicators (Badriyah et al., 2023; Dafit & Ansari, 2026).

Where a comparison group existed, the AR group generally scored higher (see "Findings by school level"), but effects were expressed in different metrics (mean differences, normalised gains, partial η^2, Cohen's *d*) that cannot be compared directly. Among single-group studies, effect sizes ranged from $d = 0.53$ with a low normalised gain of 0.21 (Usmeldi et al., 2026) to $d = 1.22$ (Damopolii et al., 2022) and $d = 1.24$ (Hasyim & Haris, 2026). Syntheses report large effects (Lu et al., 2025, g = 1.109; Silitonga & Emiliannur, 2026, 0.765 for physics); we agree on direction but did not pool effects.

Results were not uniformly positive. Alkhabra et al. (2023) found no difference in critical thinking by type of AR (image or marker) or gender; only students' mental capacity mattered. Wen et al. (2023) found that only the class combining AR with an inquiry framework differed significantly from the control ($p = .018$). In Darmawan et al. (2026), three of six FRISCO indicators improved significantly (focus, reason, and inference), situation was borderline ($p = .050$), and clarity did not improve ($p = .481$). Efendi et al. (2025) reported significant effects on critical thinking and creativity, but the path coefficients were negative (−0.414 and −0.410), which the authors described as a negative relationship yet interpreted as support for AR.

The construct measured may also matter. Lu et al. (2025) found larger effects for skills ($g = 1.243$) than for dispositions ($g = 0.371$), yet the one study in our corpus that

measured disposition (Irwanto et al., 2026) still found a significant adjusted difference (partial $\eta^2 = .233$). Framework-aligned tests allow indicator-level analysis, as in Darmawan et al. (2026), whereas studies reporting only a total score or no indicators (Badriyah et al., 2023; Dafit & Ansari, 2026) leave it unclear what changed, and self-report scales may reflect enthusiasm for a new medium as much as thinking.

Findings by School Level

Primary schools (six studies). All six compared AR with a control or alternative condition, but the analyses were uneven. Three clearly favoured AR: (Akbar et al., 2025; means of 78.90 versus 71.94, $t(59) = -2.503$, $p = .015$, although without a pretest and with random assignment described for a quasi-experimental design), Ismail et al. (2025; normalised gains of 0.62–0.69 versus 0.34–0.44), and Dafit & Ansari (2026; partial $\eta^2 = .487$). Three gave qualified results. In Dilmen & Atalay (2021), the between-group difference on the overall 21st-century skills scale was not significant and critical thinking was tested only within the AR group; in Wen et al. (2023), only AR combined with an inquiry framework differed from the control; and Atmojo et al. (2025) reported a significant but negative path coefficient for the AR class (−0.768), interpreted as favouring AR, despite unequal pretest means.

Lower secondary schools (four studies). Two studies were in science and two in mathematics. Hanggara & Qohar (2024) reported pretest equivalence and a large posttest difference (23.84 points), although the normalised gain of the AR class (0.394) was medium, and Hayat et al. (2026) reported normalised gains of 0.56 versus 0.35. Darmawan et al. (2026), a single group of 30 students, found uneven gains across indicators, which a single-group design cannot attribute to the intervention.

Upper secondary schools (eleven studies). Six of the eleven studies were in physics, two in chemistry, two in biology, and one in mathematics. Four used a single group (Damopolii et al., 2022; Hasyim & Haris, 2026; Nirfayanti et al., 2026; Saidin et al., 2024), and three of them reported large changes ($d = 1.22$, $d = 1.24$, and $t = 12.56$, respectively). Controlled studies varied: Irwanto et al. (2026) found a significant difference in critical thinking disposition after pretest adjustment (partial $\eta^2 = .233$), Rahmayani et al. (2024) reported large partial η^2 values (.531–.787), and Asrizal et al. (2024) reported a small difference (means of 77 versus 72).

Vocational schools (six studies). The studies spanned physics, civil engineering, accounting, informatics, building design, and a survey of AR acceptance. Sukatiman et al. (2026) provided the strongest experimental evidence in the review (216 students in six

schools, covariate adjustment, adjusted means of 78.42 versus 68.42). Badriyah et al. (2023) reported much larger gains in AR classes at two schools but were rated low (MMAT 2/5) because several criteria could not be judged. Usmeldi et al. (2026) found a modest change (d = 0.53) in a single group, and Dewi et al. (2025) found a significant indirect effect (β = 0.234) of AR acceptance on self-rated critical thinking, which, as a survey of perceptions, is not evidence that AR instruction improves critical thinking.

AR Technologies and Pedagogical Frameworks

The studies used AR in several forms: marker- or image-based AR on smartphones (Akbar et al., 2025; Alkhabra et al., 2023; Damopolii et al., 2022); AR embedded in learning materials (P. S. Dewi & Kuswanto, 2023; Hasyim & Haris, 2026; Rahmayani et al., 2024); game-based forms (Hanggara & Qohar, 2024; Irwanto et al., 2026; Kurnianto et al., 2024); and AR combined with software or platforms (Darmawan et al., 2026; Efendi et al., 2025; Nirfayanti et al., 2026; Sukmawati et al., 2024). AR was rarely used alone: inquiry-based frameworks accompanied it in Ismail et al. (2025), Usmeldi et al. (2026), and Wen et al. (2023), project-based learning in Atmojo et al. (2025), STEM or ethno-STEM approaches in Asrizal et al. (2024) and Rahmayani et al. (2024), and a multi-step instructional model in Darmawan et al. (2026). The reported effect is therefore usually that of a package combining AR, a pedagogical approach, and the novelty of a new medium.

Only a few designs allowed any separation. Wen et al. (2023) compared AR with an inquiry framework, the framework alone, and a control; only the combined condition differed from the control, and it did not outperform the framework alone. Alkhabra et al. (2023) compared two kinds of AR and found no difference, and Rahmayani et al. (2024) contrasted an AR e-book with a PDF and a printed textbook, which compares formats, not instructional approaches. The evidence therefore speaks to whether lessons that include AR are associated with higher critical thinking, not to which feature is responsible.

Methodological Quality

No study was rated high quality, three were rated moderate (Dewi et al., 2025; Irwanto et al., 2026; Sukatiman et al., 2026), and 24 were rated low (Table 1). Recurring weaknesses were the absence of random assignment and of reported baseline equivalence, the lack of pretest adjustment, no report of whether the intervention was delivered as planned, and small samples (16–35 students in five studies); six studies had no control group. Among the 22 non-randomised studies, 17 used appropriate measurements, and 16 had complete outcome data, but none showed that participants were representative, only three accounted for confounders, and none reported that the intervention was delivered as intended

(Table 1). The weaknesses are therefore mainly of design and reporting, not of measurement as the MMAT defines it. Because most studies could not show representative samples, control of confounders, or delivery as intended, we read their reported directions with low confidence.

Table 1. Characteristics, MMAT criterion-level ratings, and reported direction of the critical thinking result of the 27 studies (sample size in parentheses)

Study	Level	Design (sample)	C1	C2	C3	C4	C5	Score	Direction
Dafit & Ansari (2026)	Primary	Mixed methods, control (147)	Y	?	?	?	N	1/5	Positive
Akbar et al. (2025)	Primary	Posttest only, control (61)	?	Y	Y	N	?	2/5	Positive
Ismail et al. (2025)	Primary	Pre–post, control (53)	?	Y	Y	?	?	2/5	Positive
Atmojo et al. (2025)	Primary	Pre–post, control (n.r.)	?	Y	?	N	?	1/5	Qualified
Wen et al. (2023)	Primary	Three conditions (117)	?	?	?	Y	?	1/5	Qualified
Dilmen & Atalay (2021)	Primary	Mixed methods, control (62)	Y	?	N	N	N	1/5	Qualified
Hayat et al. (2026)	Lower sec.	Quasi-exp., control (60)	?	Y	?	?	?	1/5	Positive
Darmawan et al. (2026)	Lower sec.	Single group (30)	N	Y	Y	N	?	2/5	Qualified
Hanggara & Qohar (2024)	Lower sec.	Non-equivalent control (77)	?	Y	Y	?	?	2/5	Positive
Alkhabra et al. (2023)	Lower sec.	Quasi-exp., two types of AR (120)	?	Y	Y	?	?	2/5	Null
Hasyim & Haris (2026)	Upper sec.	Single group (35)	?	Y	Y	N	?	2/5	Positive
Irwanto et al. (2026)	Upper sec.	Pre–post, control (72)	?	Y	Y	Y	?	3/5	Positive
Nirfayanti et al. (2026)	Upper sec.	Single group (30)	?	?	Y	N	?	1/5	Positive
Rahmawati et al. (2026)	Upper sec.	Field trial, comparison unclear (60)	?	?	?	?	?	0/5	Positive
Saidin et al. (2024)	Upper sec.	Single group, mixed methods (16)	Y	Y	?	?	N	2/5	Positive
Rizki et al. (2024)	Upper sec.	Quasi-exp., control (54)	?	Y	Y	?	?	2/5	Positive
Dewi & Kuswanto (2023)	Upper sec.	Control group (72)	?	Y	Y	?	?	2/5	Positive
Damopolii et al. (2022)	Upper sec.	Single group (20)	?	Y	Y	N	?	2/5	Positive

Study	Level	Design (sample)	C1	C2	C3	C4	C5	Score	Direction
Rahmayani et al. (2024)	Upper sec.	Three classes (93)	?	Y	Y	?	?	2/5	Positive
Kurnianto et al. (2024)	Upper sec.	Quasi-exp., control (62)	?	Y	Y	?	?	2/5	Positive
Asrizal et al. (2024)	Upper sec.	Quasi-exp., control (68)	?	?	?	?	?	0/5	Positive
Usmeldi et al. (2026)	Vocational	Single group (100)	?	Y	Y	N	?	2/5	Positive
Sukatiman et al. (2026)	Vocational	Pre–post, control (216)	?	Y	Y	Y	?	3/5	Positive
Sukmawati et al. (2024)	Vocational	Quasi-exp., control (120)	?	Y	Y	?	?	2/5	Positive
Efendi et al. (2025)	Vocational	Experiment, comparison unclear (n.r.)	?	?	?	N	?	0/5	Qualified
Badriyah et al. (2023)	Vocational	Cluster-sampled control (120)	?	Y	Y	?	?	2/5	Positive
Dewi et al. (2025)	Vocational	Survey (141)	Y	?	Y	?	Y	3/5	Positive

Note. C1–C5 are the five MMAT 2018 criteria for the study's design category (wording in Supplementary Table S1); Y = yes, N = no, ? = can't tell. Score = number of criteria met, not a validated total. Direction = direction reported by the authors (Positive; Qualified = positive only for some indicators or conditions, or with an unexplained anomaly; Null = no difference), not evidence of a causal effect. Quasi-exp. = quasi-experimental; n.r. = not reported.

Our corpus and that of Lu et al. (2025) are largely distinct (only Damopolii et al., 2022, appears in both); 24 of our 27 studies were rated low, whereas they judged all 17 of theirs acceptable at a 55% cut-off; the instruments and thresholds differ, so the two appraisals are not directly comparable. The studies with the strongest comparisons or most detailed indicator-level analysis reported effects bounded in size or scope: a 10-point adjusted difference in Sukatiman et al. (2026), a difference only for the AR class that also used an inquiry framework in Wen et al. (2023), and gains on some FRISCO indicators only in Darmawan et al. (2026). Direction was positive in 15 of the 20 studies with a comparison condition, in 5 of the 6 single-group studies, and in the one survey (21 of 27 overall; Supplementary Table S4). Single-group gains are pre–post improvements, comparisons with a control show an association between AR-supported lessons and higher scores, and no study clearly reported random assignment to conditions (Badriyah et al., 2023, the only study rated as randomised, described cluster random sampling of classes; Akbar et al., 2025, described random assignment without a pretest); the evidence therefore does not establish that AR itself improves critical thinking. One study reported no difference (Alkhabra et al., 2023,

between types of AR), and five reported qualified results (Atmojo et al., 2025; Darmawan et al., 2026; Dilmen & Atalay, 2021; Efendi et al., 2025; Wen et al., 2023); in Atmojo et al. and Efendi et al., significant but negative path coefficients were nonetheless interpreted as favouring AR. This predominance of positive findings may reflect publication bias in the primary studies and selection bias in this review (two databases, filters that removed 115 records, and a second reviewer for subsets only), neither of which we could assess, so the safest reading is that AR-supported lessons are compatible with gains whose size is uncertain.

Mathematics Topics and Arithmetic

Four studies focused on mathematics: volume of solids in primary school (Akbar et al., 2025), polyhedra (Hanggara & Qohar, 2024) and nets of solids (Darmawan et al., 2026) in lower secondary school, and trigonometric equations in upper secondary school (Nirfayanti et al., 2026). The first three concern geometry or spatial reasoning. Of the six primary-school studies, four were in science, one in Indonesian language, and one in mathematics (volume of solids, a geometry topic), so none of the 27 addressed number or arithmetic operations. Physics (seven studies), general science (six), and vocational or general topics (five) accounted for most of the corpus (Figure 2); mathematics accounted for four studies (15%). Within this corpus, no study examined AR and critical thinking in primary arithmetic.

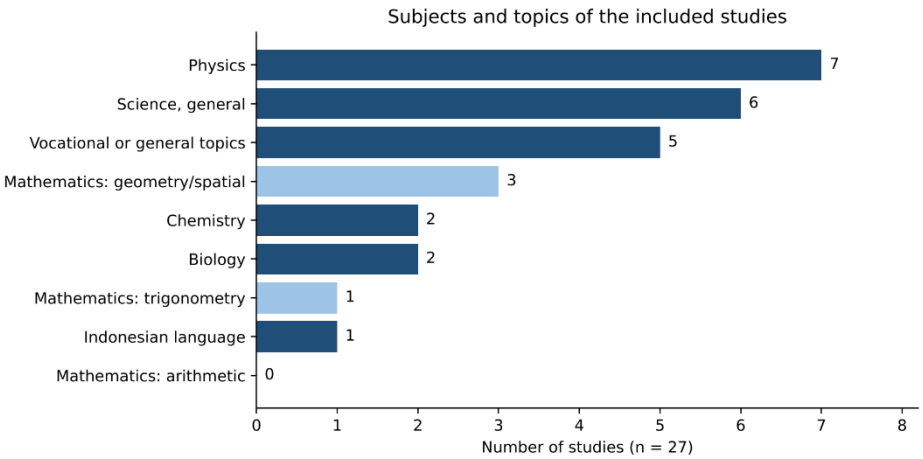


Figure 2. Subjects and Topics of the Included Studies (*n* = 27). Light bars denote mathematics topics.

The topical pattern agrees with earlier work: Lu et al. (2025) found the largest effects for three-dimensional visualisation (g = 1.216) and immersion (g = 1.198), which fits our corpus's concentration on physics, geometry, and spatial topics, and the mathematics syntheses (Assidiqi & Sugiman, 2026; Islim et al., 2024; Syarifah et al., 2026) show the

same concentration. We read AR as a representational support that gives learners something concrete to interpret; critical thinking probably develops only when tasks require them to justify and evaluate what they see, which may be why the inquiry framework mattered in Wen et al. (2023) and why gains were uneven across indicators (Darmawan et al., 2026). Multimedia learning theory (Mayer, 2021) offers a complementary explanation: well-designed visual representations help learners build mental models, and AR gives access to three-dimensional models that are hard to provide otherwise. Cognitive load theory (Sweller, 2020) suggests a competing account: AR can split attention between real and virtual views and add extraneous load, although a systematic review found AR generally less cognitively demanding, with evidence resting on media comparisons (Buchner et al., 2022). These are our interpretations, untested by the reviewed studies; selection is equally plausible, since AR may be chosen for topics where its visual appeal is obvious.

The supplementary check shows that the gap needs careful wording. AR has been used for arithmetic in primary schools: multiplication tables in a pilot with 37 children (Rebollo et al., 2022), fractions with 60 fifth-grade students (Özdemir & Özçakır, 2019), and arithmetic games evaluated for usability with sixth-grade pupils (Volioti et al., 2023). None measured critical thinking. One Indonesian primary-school study reported an N-gain of 0.79 for critical reasoning with an AR mathematical storybook (Ermawati et al., 2024), but its topic was geometry and it was a development study. The gap is thus the absence of studies combining AR, arithmetic, and a measure of critical thinking, not the absence of AR in arithmetic.

Arithmetic differs from geometry in what must be represented. Quantity and operations can be modelled with objects, so it is not self-evident that AR adds more than concrete materials do (effects of Cuisenaire rods in arithmetic varied with fidelity to the accompanying teaching; Benson et al., 2022). Critical thinking in primary arithmetic usually means explaining why a procedure works, judging whether an answer is reasonable, comparing strategies, and spotting patterns; Assidiqi and Sugiman (2026) similarly conclude that stronger claims about reasoning require tasks and instruments that capture explanation and justification. None of the 27 studies measured these skills in number work, so findings for geometry should not be extended to arithmetic.

Implications and Limitations

For practice, teachers can treat the evidence as suggestive, not conclusive, for geometry, science, and vocational content, but should not assume it transfers to arithmetic; AR should serve a question that students must answer, assessment should ask students to

justify and evaluate, and devices, markers or applications, and connectivity should be checked before adoption (Ismail et al., 2025). For research, four priorities follow: (1) controlled studies of AR in number and operations, comparing AR with concrete materials and measuring critical thinking by indicator; (2) factorial or multi-arm designs, as in Wen et al. (2023); (3) validated instruments aligned with a stated framework, distinguishing skill from disposition; and (4) routine reporting of pretest adjustment, baseline equivalence, effect sizes with confidence intervals, implementation fidelity, and school location.

These conclusions have limits. We searched only Scopus and ERIC and included only open-access articles, so 115 records were not assessed (we did not seek their full texts elsewhere) and Indonesian-language studies indexed elsewhere, including arithmetic studies in primary schools, may be missing; Ermawati et al. (2024), openly available and reporting critical reasoning in AR-supported primary mathematics, was not retrieved. A systematic search of national databases such as Garuda with the same protocol would be a useful extension. A second reviewer checked only subsets, and agreement on MMAT criteria was low ($\kappa = 0.36$), so appraisal errors cannot be excluded; we could not consult every earlier synthesis in full; and many MMAT judgements were "can't tell" because reports omitted information. Excluding HOTS studies that did not operationalise critical thinking (four) may understate mathematics coverage, and with six primary-school studies, conclusions about primary education are provisional. Our claims, including those about arithmetic, concern the indexed, open-access corpus ("no included study"), not the whole literature ("no study exists"). With 23 of 27 studies from Indonesia, the findings describe mainly Indonesian classrooms and their transfer to other education systems is uncertain. Strengths are adherence to PRISMA 2020 and open reporting of search strings, exclusions, and item-level appraisal.

CONCLUSION

In the 27 studies reviewed, AR-supported lessons were usually reported alongside higher critical thinking, but the evidence is methodologically limited and inconsistently measured, so the findings are compatible with gains whose size is uncertain. The research questions are answered as follows.

1. Distribution (RQ1). The evidence is concentrated in Indonesia (23 of 27 studies, so the conclusions mainly describe Indonesia), in upper secondary and vocational schools (17 of 27), and in science (17 of 27); only four studies concern mathematics, and only one explicitly described a rural school.

2. Measurement and effects (RQ2). Critical thinking was measured with tests (22 studies) or self-report scales (five) under different frameworks. Authors reported higher critical thinking in 21 of 27 studies, a reported direction and not demonstrated causal effectiveness (no study clearly reported random assignment to conditions), and effects are not comparable across metrics.
3. AR forms and pedagogy (RQ3). AR took several forms and was usually combined with inquiry, project-based, or STEM frameworks, so its effect was rarely separable from the pedagogy.
4. Quality (RQ4). No study met four or five MMAT criteria and 24 met two or fewer. Among the 22 non-randomised studies, 17 used appropriate measurements and 16 had complete outcome data, but none showed representative participants, three accounted for confounders, and none reported delivery as intended (Table 1); the weaknesses are mainly of design and reporting.
5. Arithmetic (RQ5). The mathematics studies addressed geometry, spatial reasoning, or trigonometry; none of the 27 studies, all of which measured critical thinking, addressed arithmetic (AR has been used in arithmetic without such a measure, so this concerns the reviewed corpus).

The combination of AR, primary-school arithmetic, and indicator-level measurement of critical thinking is unstudied in this corpus, but whether AR adds value over concrete materials there is itself open. The next step is a well-controlled study in primary schools that compares a structured arithmetic approach with and without AR, and with concrete materials alone, and measures critical thinking by indicator. Results from geometry or science should not be assumed to carry over to arithmetic.

DECLARATIONS

Author Contribution : Nafsia Lestarya Purba: Conceptualization, Data Curation, Formal Analysis, Methodology, Writing – Original Draft, and Visualization.
Yunus Abidin: Conceptualization, Supervision, and Writing – Review & Editing.
Dede Trie Kurniawan: Methodology, Validation, and Writing – Review & Editing.

Mario Capestrano Gesiradja: Validation and Writing – Review & Editing.

Funding Statement : The authors received no specific funding for this study.

Conflict of Interest : The authors declare no conflict of interest.

Additional Information : **Ethics approval and consent.** Not applicable. This study is a systematic review of published literature and involved no human participants or new data collection. **Data availability.** The review protocol and Supplementary Files S1–S6 are available on the Open Science Framework (https://osf.io/qt8x3/?view_only=113eca949f6d43fc84e60292ae296023). **Use of generative AI.** The authors declare that Claude AI (Anthropic) was used solely for English language improvement, including grammar, clarity, and academic style refinement. This tool did not contribute to the conceptualization, research design, data analysis, or theoretical interpretation of the manuscript.

REFERENCES

Ahmad, N. I. N., & Junaini, S. N. (2020). Augmented reality for learning mathematics: A systematic literature review. *International Journal of Emerging Technologies in Learning*, *15*(16), 106–122. https://doi.org/10.3991/ijet.v15i16.14961

Akbar, A., Herman, T., Suryadi, D., Putra, E. D., & Blegur, J. (2025). Integrating augmented reality in mathematics learning to improve critical thinking skills of elementary school students. *Emerging Science Journal*, *9*(2), 764–779. https://doi.org/10.28991/ESJ-2025-09-02-014

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, *20*, 1–11. https://doi.org/10.1016/j.edurev.2016.11.002

Alkhabra, Y. A., Ibrahem, U. M., & Alkhabra, S. A. (2023). Augmented reality technology in enhancing learning retention and critical thinking according to STEAM program. *Humanities and Social Sciences Communications*, *10*(1). https://doi.org/10.1057/s41599-023-01650-w

Asrizal, A., Amnah, R., Zan, A. M., Hidayati, H., Helma, H., & Annisa, N. (2024). Ethno-STEM integrated digital teaching material with augmented reality to promote students' learning skills and innovation. *International Journal of Education in Mathematics, Science, and Technology*, *12*(6), 1638–1656. https://doi.org/10.46328/ijemst.4640

Assidiqi, H. & Sugiman. (2026). Augmented and virtual reality for abstract mathematical visualization: A systematic review of cognitive outcomes. *Mathline: Jurnal Matematika Dan Pendidikan Matematika*, *11*(3), 651–666. https://doi.org/10.31943/mathline.v11i3.1154

Atmojo, I. R. W., Saputri, D. Y., Dewi, R. K., Salimi, M., Roslan, R. M., & Halim, L. (2025). The effect of the implementation of a project-based learning model assisted by

augmented reality on sixth graders' critical thinking skills on solar system materials. *Educational Process: International Journal*, *15*. https://doi.org/10.22521/edupij.2025.15.131

Badriyah, N. L., Yusuf, M., & Efendi, A. (2023). Augmented reality as a media for distance learning in the digital era: Contribution in improving critical thinking skills. *International Journal of Information and Education Technology*, *13*(11), 1769–1775. https://doi.org/10.18178/ijiet.2023.13.11.1988

Benson, I., Marriott, N., & McCandliss, B. D. (2022). Equational reasoning: A systematic review of the Cuisenaire–Gattegno approach. *Frontiers in Education*, *7*, 902899. https://doi.org/10.3389/feduc.2022.902899

Buchner, J., Buntins, K., & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, *38*(1), 285–303. https://doi.org/10.1111/jcal.12617

Dafit, F., & Ansari, K. (2026). Augmented reality-supported reading instruction: Fostering reading literacy and critical thinking in primary schools. *Online Learning in Educational Research*, *6*(2), 55–67. https://doi.org/10.58524/oler.v6i2.1111

Damopolii, I., Paiki, F. F., & Nunaki, J. H. (2022). The development of comic book as marker of augmented reality to raise students' critical thinking. *TEM Journal*, *11*(1), 348–355. https://doi.org/10.18421/TEM111-44

Darmawan, P., Rofiki, I., Sulaiman, T., Susandi, A. D., Kandaga, T., Fatimah, S., & Puspita, E. (2026). Development of ARGEONET integrated with metaverse and M6 learning model to enhance students' mathematical critical thinking skills. *Infinity Journal*, *15*(3), 717–756. https://doi.org/10.22460/infinity.v15i3.p717-756

Dewi, F. K., Hidayat, H., Ernawati, & Effendi, H. (2025). Bridging technology and cognition: Investigating augmented reality acceptance and thinking skill development using SEM-IPMA analysis. *Salud, Ciencia y Tecnología*, *5*. https://doi.org/10.56294/saludcyt20251943

Dewi, P. S., & Kuswanto, H. (2023). The effectiveness of the use of augmented reality-assisted physics e-module based on pedicab to improve mathematical communication and critical thinking abilities. *Journal of Technology and Science Education*, *13*(1), 53–64. https://doi.org/10.3926/jotse.1714

Dilmen, İ., & Atalay, N. (2021). The effect of the augmented reality applications in science class on students' 21st century skills and basic skills. *Journal of Science Learning*, *4*(4), 337–346. https://doi.org/10.17509/jsl.v4i4.32900

Efendi, R., Ambiyar, Estuhono, & Wulandari, R. A. (2025). Bridging the Industry 4.0 skills gap: An immersive augmented reality mobile learning approach for vocational education. *International Journal of Interactive Mobile Technologies*, *19*(6), 60–74. https://doi.org/10.3991/ijim.v19i06.53825

Ennis, R. H. (1996). *Critical thinking*. Prentice Hall.

Ennis, R. H. (2018). Critical thinking across the curriculum: A vision. *Topoi*, *37*, 165–184. https://doi.org/10.1007/s11245-016-9401-4

Ermawati, D., Riswari, L. A., Hilyana, F. S., & Wijayanti, E. (2024). Pengembangan buku cerita matematis berbasis augmented reality untuk kemampuan bernalar kritis siswa SD [Development of an augmented reality-based mathematical storybook for elementary students' critical reasoning]. *Jurnal Pendidikan Matematika: Judika Education*, *7*(2). https://doi.org/10.31539/judika.v7i2.12454

Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality*, *23*(4), 447–459. https://doi.org/10.1007/s10055-019-00379-9

Halpern, D. F., & Dunn, D. S. (2021). Critical thinking: A model of intelligence for solving real-world problems. *Journal of Intelligence*, *9*(2), 22. https://doi.org/10.3390/jintelligence9020022

Hanggara, Y., & Qohar, A. (2024). The impact of augmented reality-based mathematics learning games on students' critical thinking skills. *International Journal of Interactive Mobile Technologies*, *18*(7), 173–187. https://doi.org/10.3991/ijim.v18i07.48067

Hasyim, M., & Haris, A. (2026). The effectiveness of an augmented reality-integrated physics textbook on students' critical and computational thinking skills. *International Journal of Learning, Teaching and Educational Research*, *25*(7), 846–863. https://doi.org/10.26803/ijlter.25.7.38

Hayat, M. S., Sumarno, S., Nada, N. Q., Ismail, A., Gumilar, S., Saprudin, S., Ihtiar, A., Bariroh, G., & Hayati, S. A. (2026). Promoting AR-scaffolded inquiry model in science learning to improve students' critical thinking and communication skills. *International Journal of Information and Education Technology*, *16*(5), 1219–1228. https://doi.org/10.18178/ijiet.2026.16.5.2590

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, *34*(4), 285–291. https://doi.org/10.3233/EFI-180221

Irwanto, I., Zeral, L., & Saputra, I. (2026). The effect of augmented reality game-based learning on students' critical thinking disposition and academic achievement. *Journal of Technology and Science Education*, *16*(1), 68–79. https://doi.org/10.3926/jotse.3607

Islim, O. F., Namlı, Ş., Sevim-Cirak, N., Özçakır, B., & Lavicza, Z. (2024). Augmented reality in mathematics education: A systematic review. *Participatory Educational Research*, *11*(4), 115–139. https://doi.org/10.17275/per.24.52.11.4

Ismail, A., Isrokatun, I., Bahkti, D. D., Gumilar, S., Saprudin, S., & Hayat, M. S. (2025). AR-enhanced inquiry learning in elementary science: Developing students' 4C skills. *International Journal of Learning, Teaching and Educational Research*, *24*(11), 826–851. https://doi.org/10.26803/ijlter.24.11.39

Kurnianto, F., Rahmat, A. D., & Kuswanto, H. (2024). The effect of augmented reality integrated traditional games Nglarak Blarak to improve critical thinking and graphical representation skills. *Pegem Journal of Education and Instruction*, *14*(2), 340–346. https://doi.org/10.47750/pegegog.14.02.41

Kusumaningrum, B. D., Hayat, M. S., & Sumarno. (2025). The future of science learning: Meta-analysis of the effectiveness of augmented reality in enhancing critical thinking. *Jurnal Penelitian Pendidikan IPA*, *11*(5), 1055–1062. https://doi.org/10.29303/jppipa.v11i5.11071

Lu, Z., Chiu, M. M., Wang, S., Mao, W., & Lei, H. (2025). Effects of augmented reality on students' higher-order thinking: A meta-analysis. *Journal of Educational Computing Research*, *63*(3), 728–747. https://doi.org/10.1177/07356331241309644

Mayer, R. E. (2021). *Multimedia learning (3rd ed.)*. Cambridge University Press. https://doi.org/10.1017/9781316941355

Nirfayanti, Setyawan, D., & Djafar, S. (2026). Development of AR-integrated inductive thinking-based teaching modules to improve critical thinking and mathematical generalisation skills. *Multidisciplinary Science Journal*, *8*(6). https://doi.org/10.31893/multiscience.2026405

Özdemir, D., & Özçakır, B. (2019). An analysis of the effects of augmented reality activities in teaching fractions on 5th grade students' math achievement and attitudes. *Adıyaman Üniversitesi Eğitim Bilimleri Dergisi*, *9*(1), 21–41. https://doi.org/10.17984/adyuebd.495731

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. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pujiastuti, H., & Haryadi, R. (2023). Hybrid learning impact with augmented reality to improve higher order thinking skills of students. *International Journal of Advanced and Applied Sciences*, *10*(12), 7–18. https://doi.org/10.21833/ijaas.2023.12.002

Rahmawati, T. F., Subiantoro, A. W., & Olur, B. (2026). Development of digestive system learning application based on discovery learning assisted by augmented reality. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, *12*(1), 136–150. https://doi.org/10.22219/jpbi.v12i1.42398

Rahmayani, F., Kuswanto, H., & Rahmat, A. D. (2024). Development of e-book integrated augmented reality based on STEM approaches to improve critical thinking and multiple representation skills in learning physics. *International Journal of Information and Education Technology*, *14*(4), 632–641. https://doi.org/10.18178/ijiet.2024.14.4.2087

Rebollo, C., Remolar, I., Rossano, V., & Lanzilotti, R. (2022). Multimedia augmented reality game for learning math. *Multimedia Tools and Applications*, *81*(11), 14851–14868. https://doi.org/10.1007/s11042-021-10821-3

Rizki, I. A., Suprapto, N., Saphira, H. V., Alfarizy, Y., Ramadani, R., Saputri, A. D., & Suryani, D. (2024). Cooperative model, digital game, and augmented reality-based learning to enhance students' critical thinking skills and learning motivation. *Journal of Pedagogical Research*, *8*(1), 339–355. https://doi.org/10.33902/JPR.202423825

Saidin, N. F., Halim, N. D. A., Yahaya, N., & Zulkifli, N. N. (2024). Enhancing students' critical thinking and visualisation skills through mobile augmented reality. *Knowledge Management and E-Learning*, *16*(1), 1–41. https://doi.org/10.34105/j.kmel.2024.16.001

Silitonga, M. A., & Emiliannur, E. (2026). Effectiveness of augmented reality in enhancing critical thinking in physics learning: A meta-analysis. *Jurnal Penelitian Pembelajaran Fisika*, *12*(1), 108–117. https://doi.org/10.24036/jppf.v12i1.50

Sukatiman, Saputro, I. N., & Budiarto, M. K. (2026). Improving critical thinking skills through a flipped project-based learning model integrated with mockup media and augmented reality. *Electronic Journal of E-Learning*, *24*(1), 61–74. https://doi.org/10.34190/ejel.24.1.4176

Sukmawati, F., Suparmi, Prihatin, R., & Santosa, E. B. (2024). Design and evaluation a mobile augmented reality to enhance critical thinking skills for vocational high schools. *Salud, Ciencia y Tecnología*, *4*, 1000. https://doi.org/10.56294/saludcyt2024.1000

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Syarifah, M., Atmoko, A., & Ekawati, R. (2026). The effectiveness of using augmented reality media in mathematics learning in primary school: A systematic literature

review (SLR). *Jurnal PAJAR (Pendidikan Dan Pengajaran)*, *10*(1), 1–13. https://doi.org/10.33578/pjr.v10i1.343

Usmeldi, U., Amini, R., Fadhilah, F., & Fadillah, M. A. (2026). Augmented reality and guided inquiry to support 21st-century competencies: A pre- and post-study in a vocational physics module. *International Journal of Online and Biomedical Engineering*, *22*(1), 4–20. https://doi.org/10.3991/ijoe.v22i01.57997

Volioti, C., Orovas, C., Sapounidis, T., Trachanas, G., & Keramopoulos, E. (2023). Augmented reality in primary education: An active learning approach in mathematics. *Computers*, *12*(10), 207. https://doi.org/10.3390/computers12100207

Wen, Y., Wu, L., He, S., Ng, N. H.-E., Teo, B. C., Looi, C. K., & Cai, Y. (2023). Integrating augmented reality into inquiry-based learning approach in primary science classrooms. *Educational Technology Research and Development*, *71*(4), 1631–1651. https://doi.org/10.1007/s11423-023-10235-y