EDUMATSAINS
JURNAL PENDIDIKAN MATEMATIKA DAN SAINS
E-ISSN:2527-7235
Volume 11 Issue. 2, October 2026, pp 293-310
http://ejournal.uki.ac.id/index.php/edumatsains

Design Characteristics of Board Games for Learning Number and Operations: A Systematic Literature Review

Nabila Zahida[1*], Yuli Bangun Nursanti[2], Imam Sujadi[3]

[1,2,3] Pendidikan Matematika FKIP Universitas Sebelas Maret

e-mail: nabilazahida02@student.uns.ac.id

Article Info

Article history:
Received: October 1st, 2026
Revised: October 26th, 2026
Accepted: October 27th, 2026
Available online: October 31st, 2026

https://doi.org/10.33541/edumatsains.v11i2.8787

Abstract

Board games are increasingly used in mathematics classrooms, yet little is known about how their design embeds number content beyond early childhood. This systematic literature review maps the design characteristics of board games for teaching number and operations in primary, secondary, and higher education, synthesises their reported outcomes, and identifies research gaps. Following PRISMA 2020, 669 records published between 2016 and 2026 were retrieved from Scopus, Semantic Scholar, and Google Scholar. After screening and quality appraisal with Mullet et al.'s rubric, 16 studies were synthesised narratively. Studies were concentrated in primary education (n = 9) and Indonesia (n = 9), and most used research-and-development designs (n = 9); only 5 included a comparison group. Race-track games with question or task cards were the most common design (n = 9), whereas intrinsic integration, in which moving or building on the board enacts a number operation, appeared in 6 studies that tended to have higher quality scores, an association rather than a causal effect. Procedural knowledge was targeted more often than conceptual knowledge. All 11 achievement studies reported improvement, but only 4 controlled studies showed an advantage over a comparison condition. Board games were associated with motivation, communication, and collaboration, whereas creativity was not measured and retention was examined in only 1 study. Integers were addressed by only 2 studies, and none examined integer subtraction, multiplication, or division in lower secondary school. The review proposes design principles that favour intrinsic integration and calls for controlled studies of board games for integer operations.

Keywords: educational board game; number and operations; intrinsic integration; game-based learning; mathematics education

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

1. Introduction

Number and operations form the foundation of school mathematics. Concepts such as fractions, integers, powers, and the four arithmetic operations underpin later work in algebra, measurement, and data, yet they remain persistent sources of difficulty. Students often transfer the properties of whole numbers to fractions, treating numerators and denominators as separate integers (Navarrete-Ulloa & Munoz-Rubke, 2022), and the introduction of negative numbers requires them to build new mental representations of order and magnitude (Bofferding & Hoffman, 2019; Morais et al., 2026). In Indonesia, the problem is visible in national assessment data: on number items of the 2018/2019 junior high school national examination, only 38.51% of students answered correctly (A. M. Putri & Setyadi, 2022). Learning media that make abstract number relations tangible and engaging are therefore needed.

Board games are one such medium. They are inexpensive, require no digital infrastructure, and invite face-to-face interactions that engage players in teamwork and computational thinking (Bayeck, 2020). Evidence from early childhood shows that playing linear number board games improves children's numerical magnitude knowledge because the spatial layout of the board mirrors the number line (Siegler & Ramani, 2009). However, the effect of a board game depends on its design rather than on the label "game" itself. The Mechanics–Dynamics–Aesthetics (MDA) framework distinguishes the rules and components of a game (mechanics) from the behaviour they produce during play (dynamics) and the experiences players have (aesthetics) (Hunicke et al., 2004). Habgood and Ainsworth (2011) further showed that learning is stronger when content is intrinsically integrated into the core mechanics, so that playing well requires using the target knowledge, rather than being attached to the game as quiz questions. What students learn also matters: mathematical knowledge can be factual, conceptual, or procedural, and these types develop in different ways (Rittle-Johnson & Schneider, 2014).

Previous reviews have addressed related but different questions. Bayeck (2020) reviewed board gameplay and learning across disciplines and identified mathematics as one of its main domains, but did not analyse how number content is embedded in game design. Arum et al. (2026), in this journal, synthesised the pedagogical design of digital educational games for conceptual understanding in mathematics. Sariningtias et al. (2026) reviewed how digital and non-digital technologies are integrated into innovative learning environments to support mathematical knowledge, skills, and STEM learning, and their procedure informed the method of the present review. Research on board games for numeracy, in turn, has focused largely on preschool children (e.g., Siegler & Ramani, 2009). To our knowledge, no review has synthesised how physical board games are designed to teach number and operations to school and university students, how number content is integrated into their mechanics, and what outcomes these designs produce.

This review addresses that gap. Its aim is to map the design characteristics of board games for learning number and operations in primary, secondary, and higher education, to synthesise the outcomes they report, and to identify gaps for future design. Six research questions (RQs) guided the review:

(RQ1) How are the studies distributed across years, countries, education levels, number topics, and research designs?

(RQ2) What formats, components, and mechanics characterise the board games?

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

(RQ3) How is number content integrated into the games?
(RQ4) What types of mathematical knowledge are targeted, and what cognitive outcomes are reported?
(RQ5) What affective outcomes and 21st-century skills are reported?
(RQ6) What gaps remain, and what are the implications for designing board games for number and operations?
The novelty of the review lies in combining the MDA and intrinsic-integration lenses with a knowledge-type analysis to link design decisions with learning outcomes.

2. Methods

2.1. Review design and search strategy

This study is a systematic literature review reported according to the PRISMA 2020 statement (Page et al., 2021). The selection and appraisal procedure was adapted from Sariningtias et al. (2026). Searches were run on 29 September 2026 in Publish or Perish (version 8.17) across Scopus, Semantic Scholar, and Google Scholar. For every source, the same Boolean string was entered in the Keywords field: ("board game" OR "board games" OR "boardgame" OR "tabletop game" OR "tabletop games") AND ("mathematics" OR "math" OR "mathematical") AND ("number" OR "numbers" OR "number sense" OR "arithmetic" OR "number concept" OR "number operation" OR "number operations"). Publication years were limited to 2016–2026. The maximum number of results was set to 200 for Scopus, 1,000 for Semantic Scholar, and 500 for Google Scholar. Scopus and Semantic Scholar returned fewer records than these limits, so all hits were retrieved; Google Scholar was capped at the first 500 records in the relevance order returned by the database, without any manual selection. All retrieved records were exported from Publish or Perish as RIS files and are available from the authors on request. Because the Semantic Scholar interface supports only simple keyword searches with AND/OR operators, phrase matching may not have been applied exactly in that source. The search settings and the number of records retrieved are presented in Table 1.

Table 1.
Search settings and results

Database	Search settings	Records retrieved
Scopus	Keywords field; 2016–2026; maximum 200 results	83
Semantic Scholar	Keywords field; 2016–2026; maximum 1,000 results	86
Google Scholar	Keywords field; 2016–2026; maximum 500 results (first 500 by relevance)	500
Total		669

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

2.2. Deduplication and eligibility criteria

Records were merged and deduplicated with a staged procedure adapted from Bramer et al. (2016). Records matching on first author, year, title, and journal or pages were removed automatically. Pairs with highly similar titles (similarity ≥ 0.95) were removed only when an identical DOI confirmed the match; the remaining 26 possible pairs were checked manually, and 24 were confirmed as duplicates. In total, 78 duplicates were removed, leaving 591 records. The eligibility criteria are presented in Table 2. Publication status was verified for each report in the Scopus Sources list and the SINTA database for the year of publication. Conference papers were eligible only when the proceedings were indexed in Scopus; this was verified for all four included conference papers (Calderon et al., 2018; Espinoza-Espinosa et al., 2022; Jafarkhani et al., 2025; Nurhasanah et al., 2023). The indexing status of each included study is reported in Table 3. A mixed-age sample was retained only when results for school-age participants were reported separately.

Table 2.
Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Publication type	Articles in journals indexed in Scopus or accredited SINTA 1–4 in the year of publication; papers in international conference proceedings indexed in Scopus	Non-Scopus proceedings, national seminars, SINTA 5–6 or unaccredited journals, non-Scopus international journals, books, theses, repositories
Year	2016–2026	Outside 2016–2026
Medium	Physical, technology-enhanced, or digital games that retain a board format	Card, dice, or tile games without a board; purely digital games; gamification
Domain	Number and operations (whole numbers, integers, fractions, decimals, powers and roots, factors and primes, arithmetic operations)	Algebra, geometry, measurement, probability, general reasoning
Study type	Primary studies reporting design and/or implementation with data (validation, trial, or effectiveness)	Reviews, essays, concept-identification-only or perception-only studies
Level	Primary, secondary, and higher education	Early childhood and kindergarten
Language	English or Indonesian	Other languages
Accessible	Full text available	Full text not accessible

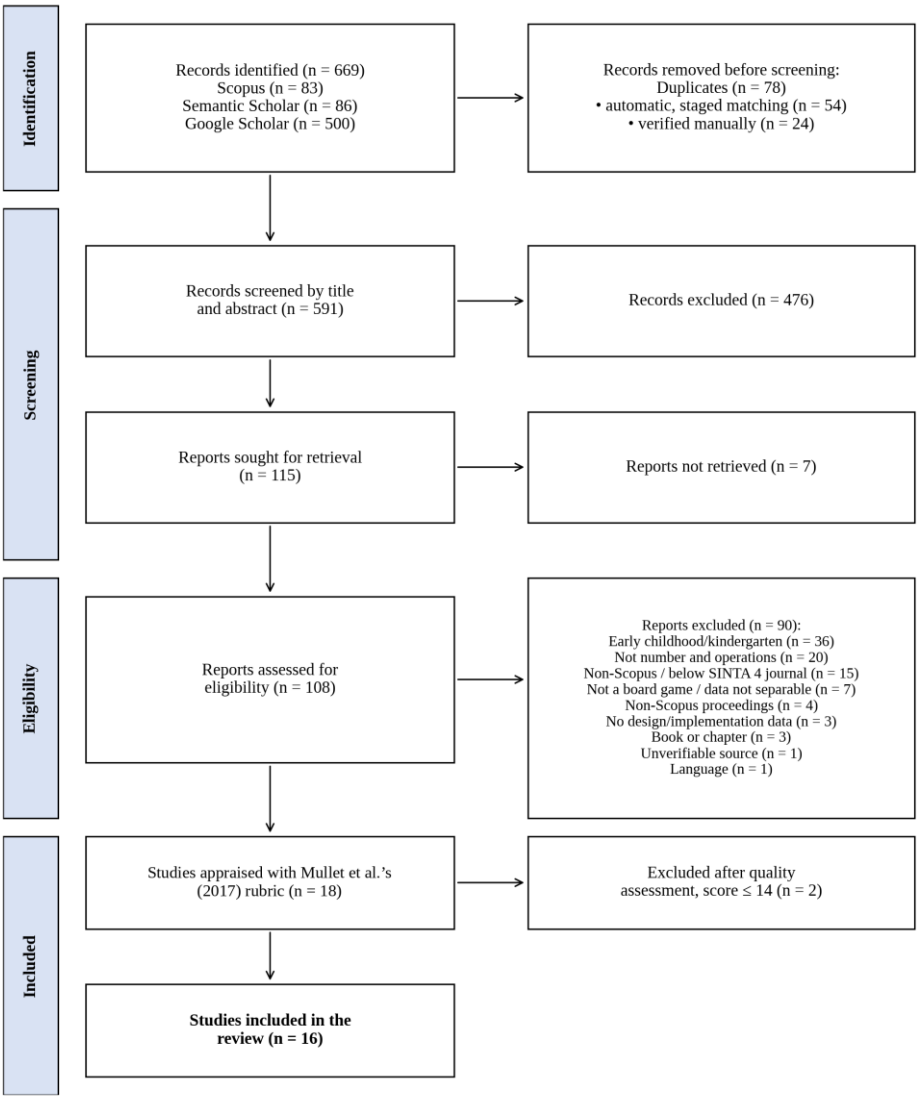
2.3. Selection process

Titles and abstracts were screened by the first author against the criteria in Table 2, using predefined exclusion codes (E0–E8; Supplementary Material). Records that were clearly irrelevant were first flagged with keyword-based rules, and every flagged record was then checked manually before exclusion. Records whose eligibility was uncertain at this stage were retained for full-text assessment. All decisions were recorded with their codes. At the screening

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

stage, 476 records were excluded, mainly because board games were not the object of study or the content was not mathematical. Of the 115 reports sought for retrieval, seven could not be obtained through open access, institutional subscription, or requests to the authors (Supplementary Table S4). The remaining 108 reports were assessed in full text, and 90 were excluded with recorded reasons (Supplementary Table S2). Of these, seven reports were excluded because the game consisted only of cards, dice, or tiles without a board, or because data on the board game could not be separated from other activities. The complete process is shown in Figure 1.

Figure 1
PRISMA 2020 flow diagram of the study selection process



This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

2.4. Quality appraisal

The methodological quality of the 18 eligible studies was appraised with Mullet et al.'s (2017) rubric, following Sariningtias et al. (2026). This rubric was chosen because the eligible studies were methodologically heterogeneous, including research-and-development, design-based, quasi-experimental, experimental, and case studies, so design-specific risk-of-bias tools could not be applied uniformly. The rubric instead evaluates reporting components that are common to educational research across designs, and it has recently been applied in a review of mathematics learning environments (Sariningtias et al., 2026). The rubric rates seven criteria (objectives, literature review, theoretical framework, participants, methods, results and conclusions, and significance) on a four-point scale (1 = does not meet the standard, 2 = nearly meets the standard, 3 = meets the standard, 4 = exceeds the standard), giving totals from 7 to 28. Studies scoring 14 or lower were excluded (Sariningtias et al., 2026); this threshold corresponds to half of the maximum score, that is, an average of 2 ("nearly meets the standard") or lower across the seven criteria. All studies were scored by the first author, and all scores were checked by the second author, who agreed with all 18 scores; therefore, no disagreements required resolution. Two studies scored 14 and were excluded (Supplementary Table S3). Garry et al. (2019) focused on the technical development of a digital board-format game, did not articulate a theoretical framework, and supported its claim of effectiveness only with a player-perception questionnaire. Chasanah et al. (2025) likewise lacked a theoretical framework and reported a pre–post increase without an adequate statistical test. Because neither study provided statistically tested evidence of learning achievement, their exclusion did not remove achievement evidence from the synthesis. This left 16 studies with scores from 15 to 26 (M = 19.81, Mdn = 20; Supplementary Table S1). Because the rubric appraises reporting quality rather than risk of bias, the scores were interpreted as indicators of methodological transparency and were used to weigh the evidence in the synthesis.

2.5. Data extraction and synthesis

A structured form was used to extract bibliographic data, country, education level, sample, research design, number topic, game format, components, core mechanics, type of content integration, targeted knowledge type, instruments, and cognitive and affective outcomes. Integration was coded as intrinsic when the board actions themselves enacted the number operation, partially intrinsic when number cards or stations were combined with movement in a way that required constructing number relations, and extrinsic when the game served mainly as a vehicle for answering question cards (Habgood & Ainsworth, 2011). Knowledge types were coded following Rittle-Johnson and Schneider (2014). Frequencies were tabulated descriptively, and findings were synthesised thematically for each RQ. When studies reported different results, they were compared by education level, design, integration type, and quality score rather than merged into a single conclusion. Because the studies differed widely in design, outcome, and instrument, and only five used a comparison group, often without the statistics needed to compute comparable effect sizes, a meta-analysis was not appropriate. A narrative synthesis was therefore conducted following the Synthesis Without Meta-analysis (SWiM) guideline (Campbell et al., 2020). Outcomes were grouped into learning achievement, validity

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

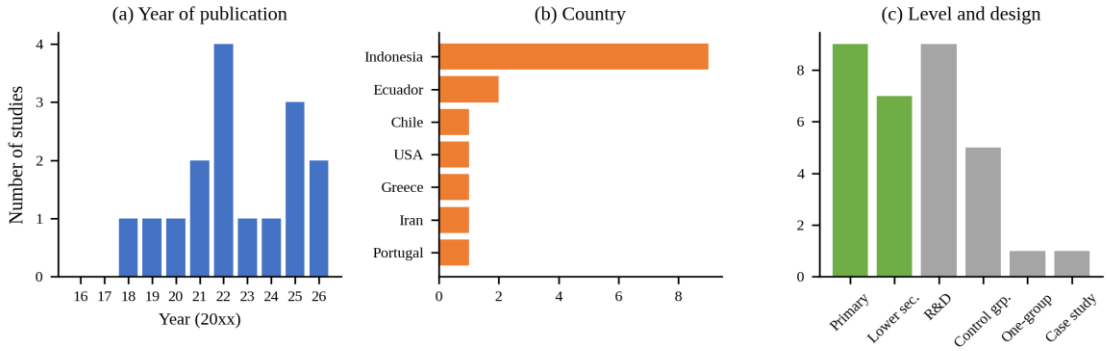
and practicality of the game, and affective outcomes and perceptions. Achievement results were further classified by the strength of the comparison: a statistically significant advantage over a comparison group, a statistically significant within-group pre–post gain, or a descriptive increase without a statistical test. Effect sizes were reported as given by the original authors. Cohen's *d* expresses the difference between group means in standard-deviation units, whereas the normalised gain (Hake, 1998) is the proportion of the possible improvement that was achieved, calculated as (post − pre)/ (maximum − pre). Because the two indices use different scales, they were not compared directly. Partial η^2 was interpreted as the proportion of outcome variance attributable to the condition.

3. Result and Discussion

3.1. Research trends (RQ1)

The 16 included studies were published between 2018 and 2026, with a peak in 2022 (four studies) and a renewed increase in 2025–2026 (five studies) (Figure 2a). Indonesia contributed more than half of the studies ($n = 9$), followed by Ecuador ($n = 2$) and single studies from Chile, Greece, Iran, Portugal, and the United States (Figure 2b). Nine studies involved primary school students and seven involved lower secondary students; no study involving upper secondary or university students met the criteria. Research-and-development (R&D) or design-based studies dominated ($n = 9$), whereas only five studies used a comparison group (Bofferding & Hoffman, 2019; Malliakas et al., 2021; Morais et al., 2026; Navarrete-Ulloa & Munoz-Rubke, 2022; Suwandayani et al., 2026) (Figure 2c). The topics were whole numbers and arithmetic operations ($n = 9$, including three that covered several number topics), fractions or rational numbers ($n = 4$), integers ($n = 2$), and powers and roots ($n = 1$). The characteristics of each study are summarised in Table 3.

Figure 2
Distribution of the included studies by (a) year of publication, (b) country, and (c) education level and research design



The geographical pattern mirrors that reported by Sariningtias et al. (2026) for technology-integrated learning environments, in which a small number of countries accounted for most studies. In this review, however, the concentration is in Indonesia, where seven of the nine

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

included studies used research-and-development or design-based approaches. This productivity is valuable for local practice, but the predominance of small, uncontrolled R&D trials limits how far the effectiveness claims can be generalised.

Table 3.
Characteristics and quality scores of the included studies

Study	Ctry	Level (n)	Design	Topic	Core mechanic	Integration	Index	Q
Prihandoko and Yunianta (2021)	ID	Grade 7 (7)	R&D	Number (mixed)	Race track + question cards	Extrinsic	SINTA 2	15
A. M. Putri and Setyadi (2022)	ID	Grade 7 (11)	R&D	Number (mixed)	Race track + question cards	Extrinsic	SINTA 2	15
Wahyuningsih and Setyadi (2020)	ID	Grade 9 (30)	R&D	Powers and roots	Race track; steps by item difficulty	Extrinsic	SINTA 2	15
Navarrete-Ulloa and Munoz-Rubke (2022)	CL	Grade 5 (39)	RCT, active control	Fractions	Stacking blocks on number-line tracks	Intrinsic	Scopus	24
Bofferding and Hoffman (2019)	US	Grade 1 (48)	Experiment	Negative integers	Linear track −10 to 10 moved by ± cards	Intrinsic	Scopus	26
Malliakas et al. (2021)	GR	Grade 7, dyslexia (124)	Quasi-experiment	Fractions	Area progression with solved-example cards	Extrinsic	Scopus	18
Calderon et al. (2018)	EC	Ages 8-13, blind (3)	Develop-ment	Multi-digit operations	Placing Braille tabs in	Intrinsic	Scopus (proc.)	19

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

					place-value rows; camera feedback			
Suwandayani et al. (2026)	ID	Grade 2 (60)	Quasi-experiment	Early numeracy	Sowing seeds (congklak/mancala)	Intrinsic	SINTA 2	23
Fathurrohman et al. (2022)	ID	Secondary (70)	Design and development	Arithmetic statements	Maze race; building statements from number and operation cards	Partially intrinsic	Scopus	23
Espinoza-Espinosa et al. (2022)	EC	Grade 4 (37)	Development, pre-post	Addition and subtraction	Race track with superhero narrative	Extrinsic	Scopus (proc.)	17
Nurhasanh et al. (2023)	ID	Lower secondary (target)	R&D (design, validation)	Fractions	Family game built on multiple representations	Intrinsic	Scopus (proc.)	20
Putra et al. (2024)	ID	Grade 8 (76)	One-group pre-post	Whole numbers and fractions	Maze race; building statements from cards and dice	Partially intrinsic	Scopus	20
Y. M. Putri and Andaryani (2025)	ID	Grade 5 (6 + 22)	R&D	Fractions	Snakes and ladders + question cards	Extrinsic	SINTA 2	20

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Rahayu and Subekti (2025)	ID	Grade 3	R&D	Four operations	Race track + bilingual inquiry cards	Extrinsic	Scopus	18
Jafarkhani et al. (2025)	IR	Grade 1, dyscalculia (2)	Case study	Operations and magnitude	Ludo with number stations	Partially intrinsic	Scopus (proc.)	22
Morais et al. (2026)	PT	Grade 6 (29)	Quasi-experiment	Integer addition	Placing ± booster cards on rocket boards	Intrinsic	Scopus	22

Note. ID = Indonesia; CL = Chile; US = United States; GR = Greece; EC = Ecuador; IR = Iran; PT = Portugal; RCT = randomised controlled trial; Q = Mullet quality score (7–28). Index = indexing status in the year of publication; Scopus (proc.) = Scopus-indexed conference proceedings; *n* = number of participants enrolled.

3.2. Characteristics of the board games (RQ2)

Thirteen games were physical, two were physical games accompanied by a digital version (Espinoza-Espinosa et al., 2022; Fathurrohman et al., 2022), and one was a technology-enhanced hybrid in which a camera and voice feedback monitored a physical board (Calderon et al., 2018). The typical components were a printed board, pawns, dice, and decks of question, reward, or punishment cards. Several games added thematic narratives, such as a jungle labyrinth, an outer-space race, superheroes in the city of Quito, or a bakery, to frame the mathematical tasks (Espinoza-Espinosa et al., 2022; Nurhasanah et al., 2023; Prihandoko & Yunianta, 2021; Wahyuningsih & Setyadi, 2020).

Six mechanic families were identified. Race tracks combined with question or task cards were the most frequent ($n = 9$): players rolled dice or answered items to advance, and correct answers were rewarded with extra steps or tokens (e.g., A. M. Putri & Setyadi, 2022; Y. M. Putri & Andaryani, 2025). The other families were placement or construction on the board ($n = 3$), a linear number track in which movement itself represented addition and subtraction ($n = 1$), sowing in the traditional game congklak ($n = 1$), area-based progression through scaffolded problems ($n = 1$), and a family game structured around multiple representations of fractions ($n = 1$). In MDA terms, the race-track mechanics mostly produced competitive dynamics and an aesthetic of challenge, while placement mechanics produced dynamics of comparison and strategic decision-making (Hunicke et al., 2004).

3.3. Integration of number content (RQ3)

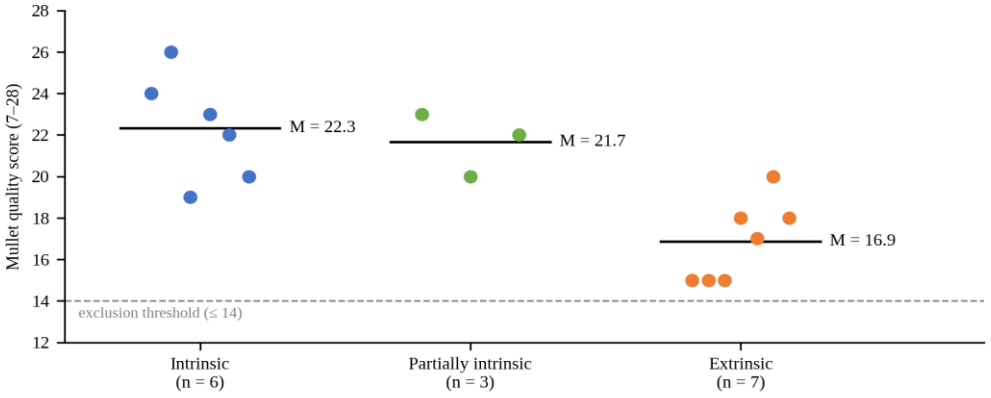
Integration type was the clearest point of divergence among the studies (Figure 3). In six studies, the number operation was the game action. In the Building Fractions game, stacking

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

blocks on two parallel number lines meant adding fractions, and comparing the lengths of the stacks determined the winner (Navarrete-Ulloa & Munoz-Rubke, 2022). In Absolute Blast!, launching a rocket required adding positive and negative boosters, and victory depended on the largest absolute value (Morais et al., 2026). Moving along a board numbered from −10 to 10 made the order of integers visible (Bofferding & Hoffman, 2019), sowing congklak seeds required continuous counting (Suwandayani et al., 2026), and placing Braille tabs in place-value rows reproduced the structure of column arithmetic (Calderon et al., 2018). In the sixth, a family game was designed around abstraction and multiple representations of fractions (Nurhasanah et al., 2023). Three games were partially intrinsic: players built mathematical statements from number and operation cards within a maze race (Fathurrohman et al., 2022; Putra et al., 2024) or solved number challenges at Ludo stations (Jafarkhani et al., 2025). The remaining seven games were extrinsic: the board determined movement, while the mathematics resided in question cards that could be replaced without changing the game.

Figure 3

Mullet quality scores of the included studies by the type of content integration



The integration type was associated with methodological quality. Studies of intrinsically and partially intrinsically integrated games had mean quality scores of 22.3 and 21.7, compared with 16.9 for extrinsic games (Figure 3). Intrinsic designs also accounted for four of the five controlled studies. In the included studies, therefore, theory-driven designs grounded in analogical reasoning, number-line representations, or cultural practice tended to be evaluated more rigorously and scored higher on the theoretical-framework criterion (means of 3.3, 2.7, and 1.7 for intrinsic, partially intrinsic, and extrinsic designs; Supplementary Table S1). This should be read as an association rather than as evidence that intrinsic integration causes better learning, because intrinsic designs were also more often published in Scopus-indexed outlets (five of six, compared with three of seven extrinsic designs), so publication venue and study quality are confounded with integration type. The pattern is nevertheless consistent with the intrinsic integration account, which holds that embedding content in the core mechanic makes learning depend on the game actions themselves (Habgood & Ainsworth, 2011). From a TPACK-like perspective, as used by Sariningtias et al. (2026), the strongest designs aligned content (number relations), pedagogy (play-based exploration), and the material affordances of the board.

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

3.4. Knowledge types and cognitive outcomes (RQ4)

Procedural knowledge was the most frequent target (n = 7), mainly fluency in arithmetic operations through repeated question cards. Three studies targeted conceptual knowledge, namely fraction magnitude and equivalence (Navarrete-Ulloa & Munoz-Rubke, 2022; Nurhasanah et al., 2023) and the order and value of negative integers (Bofferding & Hoffman, 2019), and six targeted both types (Table 4). Factual knowledge, such as number facts, was rarely an explicit goal but was implicitly practised in procedural designs.

Eleven studies measured learning achievement, and the strength of their evidence differed considerably (Table 5). Four of the five controlled studies reported an advantage of the game condition over a comparison condition on at least one measure. Students who played congklak outperformed a textbook control group (d = 1.27) (Suwandayani et al., 2026). In a randomised study of Building Fractions, the game group achieved a higher normalised gain than an active control group playing an identical game without numbers (13.26% vs. 0.73%; partial η^2 = .13) (Navarrete-Ulloa & Munoz-Rubke, 2022). Among first graders, an integer board game produced a significant multivariate effect on integer order and value tasks but not on a number placement task (Bofferding & Hoffman, 2019), showing that the choice of measure shapes what learning can be detected. In Absolute Blast! the game group, which started significantly lower, reached the level of the control group at post-test and corrected more of its errors (Morais et al., 2026). In the fifth controlled study, dyslexic students improved significantly from pre-test to post-test (6.02 to 11.70), but the control group also improved (6.33 to 8.62), and the between-group difference was reported only descriptively (Malliakas et al., 2021). The remaining six studies used one-group or case designs. Four one-group studies reported statistically significant pre–post gains, for example from 61.43 to 82.86 (Prihandoko & Yunianta, 2021), from 47.55 to 69.27 (A. M. Putri & Setyadi, 2022), and from 10.18 to 18.97 (Putra et al., 2024), together with normalised gains of .71 and .76 (Y. M. Putri & Andaryani, 2025). With samples as small as seven students and no comparison group, these gains cannot be attributed to the game rather than to additional practice or testing effects. One study reported a positive post-test without statistical analysis (Espinoza-Espinosa et al., 2022), and a case study of two students with dyscalculia reported gains of 42% and 58% on KeyMath-3 that were maintained at a two-week follow-up (Jafarkhani et al., 2025). Five studies reported only validation, practicality, usability, or perception data, so they provide evidence of feasibility rather than effectiveness (Table 5).

Table 4.
Targeted knowledge types and reported cognitive outcomes

Knowledge type	Studies	Typical design feature	Reported outcomes
Conceptual (n = 3)	Bofferding and Hoffman (2019); Navarrete-Ulloa and Munoz-Rubke (2022); Nurhasanah et al. (2023)	Board layout represents the number structure (number line, block length, multiple representations)	Gains in integer order and value; normalised gain in fraction knowledge; valid design

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Procedural (*n* = 7)	Calderon et al. (2018); Espinoza-Espinosa et al. (2022); Fathurrohman et al. (2022); Prihandoko and Yunianta (2021); A. M. Putri and Setyadi (2022); Rahayu and Subekti (2025); Wahyuningsih and Setyadi (2020)	Question or task cards attached to movement; placing digits in place-value rows	Pre–post gains in R&D trials; usability and practicality; motivation
Conceptual and procedural (*n* = 6)	Jafarkhani et al. (2025); Malliakas et al. (2021); Morais et al. (2026); Putra et al. (2024); Y. M. Putri and Andaryani (2025); Suwandayani et al. (2026)	Operations embedded in actions (sowing, adding boosters) or scaffolded solved examples	Gains over control groups; error correction and stability of correct answers; gains maintained at follow-up

Table 5.
Evidence on learning achievement in the included studies

Study	Design	*n*	Comparison	Measure	Result	Evidence type	Follow-up
Suwandayani et al. (2026)	Quasi-exp.	60	Textbook instruction	Numeracy test, 30 items (α = .82)	Post-test 78.47 vs. 66.13; $t(58) = 4.96$, $p < .001$; $d = 1.27$	Sig. vs. comparison	—
Navarrete-Ulloa and Munoz-Rubke (2022)	Randomised exp.	38	Same game without numbers	Fraction test, 31 items (α = .88/.89)	NG 13.26% vs. 0.73%; $F(1, 35) = 5.25$, $p = .028$, partial $\eta^2 = .13$	Sig. vs. comparison	—
Bofferding and Hoffman (2019)	Experiment	48 (grade 1)	Control activity	Integer order and value; number placement	Sig. multivariate effect; no effect on number placement	Sig. vs. comparison (partial)	—
Morais et al. (2026)	Quasi-exp.	29	Puzzle tasks	Researcher-made test, 60 items; error correction	E lower at pre-test ($p = .032$), equal at post-test ($p = .342$); more errors corrected ($p < .001$)	Sig. on change measures	—
Malliakas et al. (2021)	Quasi-exp.	124	Regular teaching	Fraction test, 8 items	E 6.02 → 11.70 (sig.); C 6.33 → 8.62	Sig. within E; between groups descriptive	—
Prihandoko and Yunianta (2021)	One group	7	—	Multiple-choice test, 10 items	61.43 → 82.86; paired *t* test, $p < .001$	Sig. pre–post	—
A. M. Putri and Setyadi (2022)	One group	11	—	Multiple-choice test, 15 items	47.55 → 69.27; paired *t* test, $p < .001$	Sig. pre–post	—

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" CC BY 4.0 *license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Putra et al. (2024)	One group	76	—	Numeracy test (validated)	10.18 → 18.97; Wilcoxon signed-rank test, $p < .001$	Sig. pre–post	—
Y. M. Putri and Andaryani (2025)	One group	6 + 22	—	Pre–post test	$p < .001$; N-gain .71 and .76	Sig. pre–post	—
Espinoza-Espinosa et al. (2022)	One group	37	—	Questionnaire (pre–post)	Positive post-test results reported without statistics	Descriptive	—
Jafarkhani et al. (2025)	Case study	2	—	KeyMath-3	+42% and +58%	Descriptive	2 weeks; maintained

Note. NG = normalised gain (Hake, 1998); *d* = Cohen's *d*; E = experimental group; C = comparison group; Sig. = statistically significant. The five studies without achievement measures (Calderon et al., 2018; Fathurrohman et al., 2022; Nurhasanah et al., 2023; Rahayu & Subekti, 2025; Wahyuningsih & Setyadi, 2020) reported validation, practicality, usability, or motivation data only. *n* = number of participants analysed.

3.5. Affective outcomes and 21st-century skills (RQ5)

Affective outcomes were reported in most studies, although usually with non-standardised questionnaires. Motivation and interest increased, for example from 68% to 81% on a motivation questionnaire (Wahyuningsih & Setyadi, 2020), and student response rates above 90% were common in the Indonesian R&D studies (Prihandoko & Yunianta, 2021; A. M. Putri & Setyadi, 2022). For students with special needs, Math Ludo reduced math-avoidance behaviour by 51% and increased peer interaction by 60% (Jafarkhani et al., 2025). Regarding the 4C skills, communication and collaboration were the most visible: Absolute Blast! stimulated students to verbalise, negotiate, and validate their reasoning (Morais et al., 2026), and the superhero game fostered teamwork and autonomy (Espinoza-Espinosa et al., 2022). Critical thinking was inferred from strategic decisions in placement games, whereas creativity was mentioned but not measured in any study. This profile resembles that of technology-based learning environments, in which collaboration and communication were also the most frequently supported skills (Sariningtias et al., 2026), but it shows that board games achieve these effects through face-to-face rather than mediated interaction.

3.6. Research gaps and design implications (RQ6)

Four gaps emerged. First, integers were addressed by only two studies, one with first graders and one limited to addition in grade 6 (Bofferding & Hoffman, 2019; Morais et al., 2026). No study examined integer subtraction, multiplication, or division, even though these operations are central to the lower secondary curriculum. This gap is consistent with a recent needs analysis in an Indonesian lower secondary school, which found that integer-operation instruction relied on written practice without game-based media, that limited digital infrastructure favoured non-digital media, and that seventh-grade students' accuracy and efficiency in integer operations, particularly addition and subtraction, fell short of the fluency standard (Zahida et al., in press). Second, upper secondary and higher education were absent. Third, rigorous evidence was scarce: only five studies used a comparison group, samples were often very small, and no study examined retention beyond a short follow-up. Fourth, the Indonesian studies, which constitute

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

the largest share of the evidence, mostly used extrinsic designs, so their gains may reflect additional practice rather than the game mechanics themselves.

Three design principles follow from the synthesis. First, the number structure should be built into the board and its actions, for example by representing magnitude as distance or length and operations as movement, combination, or cancellation, rather than placing problems on question cards (Habgood & Ainsworth, 2011; Navarrete-Ulloa & Munoz-Rubke, 2022). Second, games should create situations that require players to explain and negotiate their moves, because verbalisation appeared to be the route through which students corrected misconceptions about integers (Morais et al., 2026). Third, games should be evaluated with comparison groups, validated instruments, and delayed post-tests so that the contribution of the mechanics can be distinguished from that of extra practice. A board game for integer operations in lower secondary school that follows these principles would address the clearest gap identified in this review.

This review has limitations. The Google Scholar search was limited to 500 records, and the search string required number-related terms, so studies describing number content only in other words may have been missed. In addition, the Semantic Scholar interface supports only simple Boolean searches, so phrase matching in that source may have been less precise. Seven reports could not be retrieved, and the review was limited to English and Indonesian publications. Screening, data extraction, and quality appraisal were conducted by the first author, and all quality scores were checked by the second author; the third author guided the conceptualisation and direction of the review and critically reviewed the substance of the manuscript. Although decision rules were defined in advance and every decision was documented (Supplementary Tables S1–S4), the absence of fully independent double screening may have affected the reliability of the selection. Finally, the Mullet rubric appraises reporting quality rather than risk of bias, so the quality scores should not be read as estimates of bias in the reported effects.

4. Conclusion

This review synthesised 16 studies on board games for learning number and operations published between 2018 and 2026. Regarding RQ1, the studies were concentrated in primary (n = 9) and lower secondary education (n = 7), were mostly conducted in Indonesia (n = 9), and were dominated by research-and-development designs, with only five using a comparison group. Regarding RQ2, most games were physical and combined race tracks with question or task cards (n = 9), followed by placement or construction mechanics (n = 3). Regarding RQ3, number content was intrinsically integrated in six games, partially integrated in three, and extrinsically integrated in seven; intrinsically integrated games were more often grounded in theory and evaluated with stronger designs, although this association does not show that intrinsic integration causes better outcomes. Regarding RQ4, procedural knowledge was targeted more often than conceptual knowledge; all eleven studies that measured achievement reported improvement, but only four controlled studies showed an advantage over a comparison condition, and most other gains came from small one-group trials. Regarding RQ5, board games were associated with motivation, communication, and collaboration, whereas creativity was not measured and retention was examined in only one study. Regarding RQ6, the main gaps were the scarcity of studies on integers, with none addressing integer subtraction, multiplication, or

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

division in lower secondary school, the absence of studies in upper secondary and higher education, and the limited use of controlled designs. Future research should therefore develop intrinsically integrated board games for integer operations in lower secondary school and evaluate them using comparison groups, validated instruments, and delayed post-tests.

5. Acknowledgments

The authors express their gratitude to God Almighty for the strength and health granted throughout the completion of this research. The authors also thank the researchers whose work is examined in this review, particularly those who made their publications accessible. Sincere appreciation is extended to the lecturers, colleagues, and all parties who served as discussion partners, encouraged the authors to begin and complete this study, and offered valuable insights during the challenging stages of writing this manuscript. The authors also thank the editor and the anonymous reviewers for their constructive comments, which substantially improved this manuscript.

6. References

Arum, N. K., Suweken, I. G., & Ardana, I. M. (2026). Pedagogical design of digital educational games for improving students' mathematical concepts understanding: A systematic literature review. *EduMatSains: Jurnal Pendidikan, Matematika dan Sains, 10*(3), 162–174. https://doi.org/10.33541/edumatsains.v10i3.7609

Bayeck, R. Y. (2020). Examining board gameplay and learning: A multidisciplinary review of recent research. *Simulation & Gaming, 51*(4), 411–431. https://doi.org/10.1177/1046878119901286

Bofferding, L., & Hoffman, A. (2019). Children's integer understanding and the effects of linear board games: A look at two measures. *The Journal of Mathematical Behavior, 56*, Article 100721. https://doi.org/10.1016/j.jmathb.2019.100721

Bramer, W. M., Giustini, D., de Jonge, G. B., Holland, L., & Bekhuis, T. (2016). De-duplication of database search results for systematic reviews in EndNote. *Journal of the Medical Library Association, 104*(3), 240–243. https://doi.org/10.3163/1536-5050.104.3.014

Calderon, C. A., Guajala, M., Lanchi, J., Barba-Guaman, L., Bermeo, C., & Rivas-Echeverria, F. (2018). A machine vision system applied to the teaching of mathematics for blind or visually impaired children. In *2018 IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA)* (pp. 1–7). IEEE. https://doi.org/10.1109/ICA-ACCA.2018.8609818

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., Thomas, J., Welch, V., & Thomson, H. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ, 368*, Article l6890. https://doi.org/10.1136/bmj.l6890

Chasanah, N., Sulianto, J., & Siswanto, J. (2025). Design and implementation of flashcard-based game learning media for elementary school mathematics. *Sosioedukasi: Jurnal Ilmiah Ilmu Pendidikan dan Sosial, 14*(4), 2718–2727. https://doi.org/10.36526/sosioedukasi.v14i4.6460

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Espinoza-Espinosa, D., Jadán-Guerrero, J., Santórum, M., & Nunes, I. L. (2022). Learning based on board game to inspire mathematical thinking. In *Human Factors and Systems Interaction* (Vol. 52, pp. 330–336). AHFE International. https://doi.org/10.54941/ahfe1002173
Fathurrohman, M., Nindiasari, H., & Rahayu, I. (2022). A conventional and digital mathematical board game design and development for use by students in learning arithmetic. *Journal on Mathematics Education, 13*(4), 631–660. https://doi.org/10.22342/jme.v13i4.pp631-660
Garry, M. H., Yamasari, Y., Nugroho, S. M. S., & Purnomo, M. H. (2019). Design and implementation serious game "Tic Tac Toe Math." In *2019 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)* (pp. 1–6). IEEE. https://doi.org/10.1109/CENIM48368.2019.8973336
Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences, 20*(2), 169–206. https://doi.org/10.1080/10508406.2010.508029
Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809
Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI*. AAAI Press.
Jafarkhani, F., Hasanvand, M. B., Hosseini, S. A., & Shirkarami, A. (2025). Impact of Math Ludo on skills in students with dyscalculia. *Proceedings of the European Conference on Games Based Learning, 19*(1), 438–445. https://doi.org/10.34190/ecgbl.19.1.4141
Malliakas, E., Jiménez-Fanjul, N., & Marín-Díaz, V. (2021). Educational intervention through a board game for the teaching of mathematics to dyslexic Greek students. *Social Sciences, 10*(10), Article 370. https://doi.org/10.3390/socsci10100370
Morais, A., Sousa, H., Domínguez, C., Ribeiro, J., & Aires, A. P. (2026). Enhancing understanding of integer addition with the Absolute Blast! game: A didactic experiment with middle school students. *Millenium: Journal of Education, Technologies, and Health, 2*(29), Article e44367. https://doi.org/10.29352/mill0229.44367
Mullet, D. R., Rinn, A. N., & Kettler, T. (2017). Catalysts of women's talent development in STEM: A systematic review. *Journal of Advanced Academics, 28*(4), 253–289. https://doi.org/10.1177/1932202X17735305
Navarrete-Ulloa, J. A., & Munoz-Rubke, F. (2022). Playing board games to learn rational numbers: A proof-of-concept. *Mind, Brain, and Education, 16*(4), 293–299. https://doi.org/10.1111/mbe.12335
Nurhasanah, F., Usodo, B., Sutopo, Chrisnawati, H. E., Kuswardi, Y., Setiawan, L., & Febrianti, T. S. (2023). Designing a family board games for teaching fractions based on theory of abstraction and multiple representation. *AIP Conference Proceedings, 2540*, Article 070013. https://doi.org/10.1063/5.0105869
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*, Article n71. https://doi.org/10.1136/bmj.n71

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia

Prihandoko, G. K., & Yunianta, T. N. H. (2021). Pengembangan board game "Labyrinth in the Forest" untuk siswa sekolah menengah pertama materi bilangan. *Jurnal Cendekia: Jurnal Pendidikan Matematika, 5*(1), 578–590. https://doi.org/10.31004/cendekia.v5i1.548

Putra, D. D., Pamungkas, A. S., Nindiasari, H., Fathurrohman, M., & Porter, A. (2024). Impact the Labirin: The board game on the student's numeracy ability. *JRAMathEdu (Journal of Research and Advances in Mathematics Education), 9*(1), 32–42. https://doi.org/10.23917/jramathedu.v9i1.3299

Putri, A. M., & Setyadi, D. (2022). Pengembangan media board game Jumanji Matematika pada materi bilangan kelas VII SMP. *Jurnal Cendekia: Jurnal Pendidikan Matematika, 6*(2), 2086–2098. https://doi.org/10.31004/cendekia.v6i2.840

Putri, Y. M., & Andaryani, E. T. (2025). Development of snake and ladder board game media to improve fifth-grade students learning outcomes on fraction. *Jurnal Pendidikan Progresif, 15*(2), 908–926. https://doi.org/10.23960/jpp.v15i2.pp908-926

Rahayu, W. A., & Subekti, P. (2025). Implementation of "E-Lens" inquiry-based learning game to improve literacy and numeracy skills. *Jurnal Cakrawala Pendidikan, 44*(3), 510–522. https://doi.org/10.21831/cp.v44i3.76930

Rittle-Johnson, B., & Schneider, M. (2014). Developing conceptual and procedural knowledge of mathematics. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 1118–1134). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.014

Sariningtias, R., Nurhasanah, F., & Pramesti, G. (2026). Integrating technology into innovative learning environments to support mathematical knowledge, skills, and STEM learning. *International Journal on Education Insight, 7*(2), 253–270. https://doi.org/10.12928/ijei.v7i2.14956

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology, 101*(3), 545–560. https://doi.org/10.1037/a0014239

Suwandayani, B. I., Ananda, L. J., Elvierayani, R. R., Yayuk, E., Ekowati, D. W., Trisniawati, & Mohamed, Z. B. (2026). The effect of traditional congklak game-based learning on the numeracy achievement of second-grade elementary students. *Primary: Jurnal Pendidikan Guru Sekolah Dasar, 15*(1), 1–14. https://doi.org/10.33578/jpfkip.v15i1.1-14

Wahyuningsih, D., & Setyadi, D. (2020). Pengembangan board game "Zathura Mathematics" pada materi bilangan berpangkat dan bentuk akar. *Kreano: Jurnal Matematika Kreatif-Inovatif, 11*(1), 46–55. https://doi.org/10.15294/kreano.v11i1.22493

Zahida, N., Nursanti, Y. B., & Sujadi, I. (in press). Needs analysis for developing a board game for integer-operation procedural fluency in junior high school. *Unnes Journal of Mathematics Education*.

This is an open access article under the HYPERLINK "https://creativecommons.org/licenses/by/4.0/" *CC BY 4.0 license.*
Copyright ©2025 by Author. Published by Universitas Kristen Indonesia