Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan
di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index

Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

Age–Mechanic–Function (AMF) Framework: A Developmental Framework for Analyzing Cognitive and Social Learning Opportunities in Analog Board Games for Children Aged 3–8

Prisyafandiafif Charifa[1]*, Syaiful Lokan[2], Ima Apriliani[3]
[1]*Anomali Lab, Indonesia.
[2]Sekolah.mu, Indonesia.
[3]Universitas Indonesia, Indonesia.
*Corresponding Author. Email: fandi@anomalilab.com

Abstract: This study aims to construct and preliminarily appraise the Age–Mechanic–Function (AMF) Framework as a systematic, evidence-informed framework for analyzing cognitive and social learning opportunities in analog board games designed for children aged 3–8 years. The study combined an integrative review, structured content analysis of a purposive sample of 20 commercially available games, and a preliminary expert appraisal involving three experts in game facilitation, game design, and early childhood education. The 20 games were purposively selected based on variation in game mechanics, reported availability, and market visibility, and satisfied the predefined inclusion and exclusion criteria by spanning all three proposed developmental age ranges and representing the full spectrum of proposed mechanic families. The framework distinguishes among proposed design range, mechanic demands, potential practice opportunities, demonstrated outcome evidence, evidence directness, and facilitation conditions. For the experts' ratings of the original combined game-level mappings, ordinal Krippendorff's alpha was .860, while pairwise unweighted kappa values ranged from .798 to .805. These estimates indicate substantial agreement in category assignments within the small expert panel but do not establish developmental validity, causal effects, or validated age cutoffs. Accordingly, the AMF Framework is proposed as a transparent decision aid for generating and comparing design hypotheses rather than as a diagnostic tool or an alternative safety-rating system. Broader expert consensus and prospective studies involving children are needed before claims regarding learning effectiveness can be made. The AMF Framework may support game developers in designing learning mechanics that are informed by available evidence and provide early childhood educators with a structured basis for evaluating and integrating educational board games into classroom practice.

Article History
Received: 26-06-2026
Revised: 30-07-2026
Accepted: 27-08-2026
Published: 20-09-2026

Key Words:
Analog Games;
Developmental Alignment;
Early Childhood;
Executive Function;
Guided Play.

How to Cite: Prisyafandiafif Charifa, Syaiful Lokan, & Ima Apriliani. (2026). Age–Mechanic–Function (AMF) Framework: A Developmental Framework for Analyzing Cognitive and Social Learning Opportunities in Analog Board Games for Children Aged 3–8. *Jurnal Kependidikan : Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran, Dan Pembelajaran*, *12*(3), 1343-1355. https://doi.org/10.33394/jk.v12i3.22425

https://doi.org/10.33394/jk.v12i3.22425

Introduction

Many analog board games organize play through tangible components and explicit rules, often in face-to-face interaction. These features can require children to remember instructions, wait, compare alternatives, coordinate movement, communicate with others, and respond to winning or losing (Arifa & Tyas, 2025; Bawawa et al., 2025). Reviews and intervention studies indicate that some analog-game procedures can support closely aligned outcomes, particularly early numerical knowledge (Laski & Siegler, 2014; Ramani et al., 2012; Ramani & Siegler, 2011). However, findings vary across ages, mechanics, facilitation, comparators, and measures. A later active-control trial found no added numerical benefit from

Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

a linear number-board game beyond usual classroom teaching, illustrating why positive findings should not be generalized to all board games or settings (James-Brabham et al., 2024).

Developmental change adds a second source of complexity. Working memory, inhibitory control, cognitive flexibility, spatial representation, language, and social understanding change substantially across early childhood, but trajectories overlap and vary among children (Best & Miller, 2010; Cowan, 2016; Diamond, 2013; Garon et al., 2008; Zelazo & Carlson, 2012). A game that requires a child to hold several rules, interpret hidden information, or manage a delayed goal may therefore be feasible for one child and inaccessible for another of the same age. Facilitation can also change task access through demonstration, visual cues, questioning, pacing, or cooperative support (Fisher et al., 2013; Skene et al., 2022; Weisberg et al., 2013).

Existing guidance often compresses three different propositions into a single educational claim: a mechanic may require a capacity, repeated play may provide opportunities to practice that capacity, and an independent measure may show developmental change. The first is a description of task demand, the second is a practice hypothesis, and the third is an empirical outcome. Treating them as equivalent can make a plausible design interpretation appear to be evidence of effectiveness. Reviews of analog game-based learning similarly note variation in study quality and limited explanation of how particular mechanics relate to specific outcomes (O'Neill & Holmes, 2022; Sousa et al., 2023). Recent review work on reveal-and-react analog games illustrates this evidentiary problem. It maps visual-search, timed-response, and memory-match mechanics to executive-function domains. However, because the studies varied in design, measures, and participants aged 3-18, the findings were synthesized narratively rather than statistically (Charifa & Apriliani, 2025). The 3–18 age range spans vast developmental shifts in cognitive, emotional, and motor skills

Two design traditions help describe this problem without resolving it empirically. The Mechanics-Dynamics-Aesthetics framework distinguishes formal rules from the interactions and experiences that emerge during play (Hunicke et al., 2004). Cognitive-load theory provides a general account of working-memory limits and element interactivity (Sweller et al., 2019). As a design inference, AMF applies this account to game features such as rule count, symbol density, hidden information, reading, time pressure, and simultaneous goals. Learning-mechanic mapping likewise encourages designers to make the connection between gameplay and pedagogical intent explicit (Arnab et al., 2015). These approaches identify analyzable structures, but none demonstrates that a child improves simply because a game contains a theoretically relevant mechanic. A developmental framework therefore needs both a design vocabulary and an evidence boundary.

The Age-Mechanic-Function (AMF) Framework was developed to make this reasoning chain explicit. It links provisional developmental design ranges with observable mechanic demands, possible practice opportunities, evidence status, and facilitation or contextual conditions. The framework is not intended to replace publisher safety or minimum-age guidance, diagnose readiness, or promise learning outcomes. Instead, it asks users to document why a mechanic might be accessible, what a child might practice while using it, and what evidence is still absent.

This study aims to establish a literature-informed theoretical framework that conceptually separates mechanic demands, practice opportunities, and demonstrated outcomes in analog games, evaluate its analytical utility across 20 analog games, and assess expert consistency regarding age-mechanic-function mappings. In doing so, it contributes a nuanced structural lens for game design and developmental research, an illustrative content analysis mapping demand-opportunity profiles, and three testable hypotheses—focusing on mechanic-

Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

demand fit, facilitation moderation, and optimal challenge—to guide prospective empirical research on player feasibility and sustained engagement.

Research Method

An integrative review was used. This design allowed developmental theory, empirical studies, reviews, and game-design literature to serve distinct purposes (Snyder, 2019; Whittemore & Knafl, 2005). The retained review record states that Scopus, ERIC, PsycINFO, and Web of Science were searched between January and March 2025. Publications in English or Bahasa Indonesia from 2000 to 2025 were described as eligible. The record reports 1,143 identified records, 421 duplicates removed, 722 records screened, 137 full texts assessed, and 33 included sources. This article therefore describes an integrative review and reports the selection totals retained from the earlier review record; it does not present a fully reproducible systematic review. No meta-analysis or numerical risk-of-bias score was produced. Gashaj et al. (2021), James-Brabham et al. (2024), and Charifa and Apriliani (2025) were added later to update the evidence base and are not presented as verified members of the originally reported 33-source set.

Evidence was classified by population and age fit, modality and mechanic similarity, study design and comparator, measured endpoint, and facilitation or contextual conditions. Direct analog interventions with age-overlapping samples were eligible to support narrow outcome statements. Observational studies were treated as associational. Studies of blocks, puzzles, physical play, or circle-time activities were classified as adjacent evidence, while digital games and older populations provided indirect design context. Developmental theory was used to describe constructs and possible task demands. The reconstructed evidence table includes the positive, null, and conflicting findings available in the retained evidence set. Because the original included-source register and extraction log were unavailable, it was not possible to confirm that every source from the reported earlier search had been retained.

A purposive sample of 20 commercially available analog games was selected to provide variation in matching, dexterity, route planning, deduction, hidden information, cooperation, and resource exchange. Reported selection criteria were a publisher minimum age of eight years or below, at least 100 BoardGameGeek user ratings at the recorded retrieval date, reported availability in Indonesia, accessible rules in English or Bahasa Indonesia, and rule-governed interaction. These variables supported pragmatic sampling and did not establish representativeness, cultural validity, educational value, or developmental fit.

Each game was coded for its mechanic family, information structure, inferred cognitive, motor, language, and social demands, possible practice opportunities, evidence directness, and facilitation or boundary conditions. The literature synthesis and game-content coding were integrated into proposed design ranges of 3-4, 5-6, and 7-8 years. These ranges are adjacent heuristic categories, not developmental stages or validated thresholds. Seventeen general mechanic-family mappings were generated and then applied to the 20-game sample.

The retained appraisal record identifies three domain experts. This combined assessment was structured as a Preliminary Plausibility Appraisal to evaluate the holistic consistency of the framework, deliberately prioritizing an overall conceptual examination over separate construct validations across individual dimensions. One was a senior board-game community facilitator with more than ten years of experience in family and school settings. The second was a professional analog-game designer, and the third was an early childhood educator with more than fifteen years of classroom experience. The retained manuscript states that they worked independently and received game-rule summaries, selected developmental background materials, and a description of AMF. However, the

Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index
Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

preserved materials do not include the complete instrument, administration record, or qualitative comments. Each expert provided one combined plausibility rating for each game. The rating covered the proposed age range, mechanic, and function together and used three categories: High (3), Moderate (2), or Low/No (1). Because the components were not rated separately, a rating does not reveal which component shaped the judgment.

Agreement was recalculated from the 20 × 3 matrix using ordinal Krippendorff's alpha and pairwise unweighted Cohen's kappa. An interval-distance alpha was retained as a sensitivity analysis. Percentile 95% confidence intervals were estimated from 10,000 nonparametric resamples of games while preserving each three-rater vector. These statistics describe agreement in category assignments within the panel; they do not measure mapping accuracy, construct validity, or child outcomes (Krippendorff, 2019; Landis & Koch, 1977).

For exact reproduction, each of the 10,000 bootstrap replicates drew 20 row indices with replacement. The procedure used Python 3.13's random.Random(20260729).randrange(20), preserved each three-rater vector, recalculated every statistic, and obtained the 95% confidence interval from the type-7 2.5th and 97.5th percentiles. Ratings were coded Low/No = 1, Moderate = 2, and High = 3. No ratings were missing, and all 10,000 replicates produced finite estimates.

Results and Discussion

Evidence Boundaries in the AMF Framework

The first result is an evidence-classification scheme that separates five sources of inference. According to literature synthesis, Table 1 shows that a direct analog intervention can support only the age, procedure, comparator, and outcome it tested. Observational exposure can support an association, whereas adjacent activities, digital games, and developmental theory can inform hypotheses without demonstrating an analog-game effect.

Table 1. Evidence Classes and Permissible Inference in AMF

Evidence class	Typical source	Permissible use	Prohibited inference
Direct analog intervention	Age-overlapping analog game with comparator and measured outcome	Support a narrowly matched outcome statement	General benefit across mechanics, ages, or outcomes
Analog observational evidence	Board-game exposure or behavior observed without causal control	Describe associations or implementation patterns	Causal developmental change
Adjacent structured-play evidence	Blocks, puzzles, physical games, or circle-time activities	Inform construct or facilitation hypotheses	Board-game or title-specific effect
Indirect digital/older-population evidence	Digital games or samples outside the target ages	Provide design context and testable propositions	Direct transfer to analog games for ages 3-8
Theory or expert judgment	Developmental theory, design framework, or professional appraisal	Define constructs, plausible demands, and questions for testing	Efficacy, validated age cutoff, or clinical utility

This hierarchy was necessary because the strongest direct evidence was construct-specific. Guided linear number-board procedures improved closely aligned numerical measures in some studies (Laski & Siegler, 2014; Ramani et al., 2012; Ramani & Siegler, 2011). However, (James-Brabham et al., 2024), found no added benefit beyond teaching as

Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

usual. Broader commercial-game interventions pooled games, mechanics, or wider age groups, limiting mechanic-by-age interpretation (Vita-Barrull et al., 2023). Longitudinal exposure evidence was observational (Gashaj et al., 2021), while serious-game meta-analyses largely concerned digital modalities and heterogeneous populations (Girard et al., 2013; Wouters et al., 2013).

Adjacent evidence remains useful when its role is explicit. Block and puzzle studies can inform spatial-language and spatial-transformation hypotheses, but they do not establish board-game effects (Ferrara et al., 2011; Levine et al., 2011). Likewise, developmental reviews identify changes in working memory, inhibition, and flexibility that may constrain participation (Best & Miller, 2010; Garon et al., 2008). They cannot assign a precise age to a mechanic or prove that practicing the task changes the underlying capacity.

The Condensed Age-Mechanic-Function Matrix

The AMF matrix organizes analysis around proposed design range, mechanic family, participation demand, possible practice opportunity, evidence status, and conditions. Table 2 condenses the 17 detailed mappings retained proposed AMF design matrix from two panel. Panel A states the proposed range, mechanic family, participation demand, and possible practice opportunity. Panel B states demonstrated outcome evidence, evidence directness, and boundary conditions. An absent outcome is shown explicitly rather than inferred from structural similarity. The ranges summarize questions to ask, not abilities every child possesses.

Table 2. Condensed AMF Framework by Proposed Design Range

Proposed range	Illustrative mechanic families	Inferred participation demands	Possible practice opportunities	Key conditions and evidence boundary
3-4	Perceptual matching, one-rule sorting, simple turn exchange, tactile placement	Discriminate salient features; imitate a demonstrated rule; wait briefly; coordinate grasp and placement	Feature labeling, rule-following rehearsal, turn exchange, controlled action	Demonstration, few categories, short turns, component safety, motor and sensory access; no general outcome is inferred
5-6	Multi-feature matching, simple sequencing, visible clue deduction, cooperative route choice	Maintain a short sequence; compare two or more features; inhibit an immediate guess; coordinate a shared choice	Updating visible information, categorization talk, elementary planning, collaborative explanation	Visual supports, limited hidden information, language scaffolds, manageable pace; evidence is mechanic- and outcome-specific
7-8	Hidden information, multi-step route planning, resource exchange, perspective comparison	Track delayed goals; integrate several constraints; interpret another player's information; revise a plan	Planning discussion, flexible strategy use, perspective comparison, delayed-goal management	Reading/numeracy load, deception norms, prior experience, emotional safety, accessibility, and facilitation; proposed range is not an optimal-age claim

Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index
Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

The matrix distinguishes demand from practice and outcome. A matching mechanic may require feature discrimination and offer repeated comparison, yet no improvement should be claimed unless an independent, aligned measure demonstrates change. A hidden-information game may require perspective comparison but also impose language, memory, and cultural demands. AMF therefore records both the hypothesized opportunity and the conditions that might make participation feasible or problematic.

This distinction also changes how age guidance is interpreted. AMF did not evaluate product safety or publisher minimum-age requirements. It asks a separate question about the cognitive and social demands of a specific implementation. A difference between a publisher label and a proposed AMF range is therefore a prompt to examine rules, support, and the individual child; it is not evidence that the publisher label is incorrect.

Operationally, a proposed mapping is constructed through four questions: What must the player perceive, remember, inhibit, coordinate, or communicate? What might repeated participation allow the player to practice? Has a sufficiently similar age-overlapping analog study measured change in that outcome? What facilitation, accessibility, cultural, or safety conditions could alter the mapping? Keeping the answers in separate fields makes a claim traceable. When the third question cannot be answered, the outcome field remains explicitly unsupported rather than being filled by theoretical similarity.

The framework also separates implementation from a mechanic label. "Deduction," for example, can involve visible or hidden clues, individual or shared decisions, language-heavy or pictorial evidence, and different memory loads. "Route building" can involve one-step spatial choice or several delayed goals. MDA and mechanic-mapping approaches help identify the formal and runtime structure, while cognitive-load theory helps locate avoidable (Arnab et al., 2015; Hunicke et al., 2004; Sweller et al., 2019). AMF adds the developmental and evidentiary questions, but it does not presume an optimal challenge level or causal benefit.

This structure makes the framework's propositions testable and refutable. A proposed range can be rejected if children within that range consistently cannot access the implementation under the specified supports. A practice hypothesis can be rejected if aligned observation does not show the expected behavior. An outcome claim can be rejected if a controlled study finds no transfer or an adverse pattern. Null findings therefore refine the framework rather than being treated as exceptions to an assumed positive effect.

Illustrative Application to 20 Games

The 20 games were selected as a purposive, illustrative sample chosen for their mechanic variation, reported availability, and market visibility to demonstrate the application of the AMF framework. While specific inclusion criteria prioritized commercial analog board and card games targeting early childhood to elementary ages with recorded publisher age recommendations, BoardGameGeek user ratings, and distinct core mechanics, exclusion criteria ruled out digital serious games, non-game laboratory tasks, and titles lacking standard descriptive data. Across the sample, the selected titles successfully represent all three targeted developmental age ranges—3–4, 5–6, and 7–8 years—as well as the full spectrum of proposed mechanic families, ranging from basic dexterity, color matching, and turn-taking to tile placement, route planning, deduction, bluffing, and spatial optimization.

For the proposed 3-4 range, First Orchard combines color matching, token movement, and a shared goal. Its observable demands include matching a visible feature, following a short turn sequence, and waiting. The structure may provide opportunities for turn exchange and rule-following rehearsal when an adult demonstrates the sequence. Animal Upon Animal adds controlled grasp and placement; any use must consider component safety, motor access,

Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index
Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

and the emotional effects of collapse. These observations do not show that either title improves inhibition, cooperation, or motor development.

For the proposed 5-6 range, Outfoxed! requires players to retain and exclude visible clues and delay a conclusion. A shared clue record and adult prompts may reduce memory and language demands. Dragomino requires tile-adjacency comparison, visual matching, and simple counting and may prompt spatial or categorization talk. Block-play evidence can inform this hypothesis, but it is adjacent rather than title-specific (Ferrara et al., 2011). SOS Dino requires path forecasting and shared move choices; the collaborative opportunity depends on language, planning support, and turn structure.

For the proposed 7-8 range, My First Stone Age combines memory, routes, and resources, requiring sequence tracking and delayed-goal management. Eat Me If You Can! combines hidden information and role selection, requiring perspective comparison and acceptance of deception as a play convention. Evidence about children's strategic lying helps define the construct but does not validate this game's age placement or effect (Evans et al., 2011). Ticket to Ride: First Journey requires set tracking, map reading, route choice, and delayed goals. These examples make the coding logic inspectable while leaving developmental change unclaimed.

Across the sample, mechanic families did not form a simple ladder from easy to difficult. The same title combined multiple demands, and the same label produced different demands when information visibility, time pressure, conflict, or scoring changed. Several publisher ages fell outside the proposed design ranges, but those differences were not tested against child behavior and should not be interpreted as corrections. The crosswalk instead identifies cases for later observation: which support is needed, which demand causes difficulty, whether children use the hypothesized process during play, and whether any change transfers beyond the task.

Preliminary Expert Appraisal

The three experts rated the plausibility of the original combined game-level mappings, not the separate cells in the reconstructed matrix. Each combined rating covered a proposed age range, mechanic, and function together. Across 60 judgments, 34 were High, 25 Moderate, and one Low/No. Ten games were unanimously High, seven unanimously Moderate, two had mixed High/Moderate ratings, and one had a Low/Moderate pattern.

Table 3. Summary of the Preliminary Three-Expert Appraisal

Indicator	Result	Interpretation boundary
Ordinal Krippendorff's alpha	.860, 95% bootstrap CI [.653, 1.000]	Consistency in use of ordered plausibility categories within this panel
Interval-distance alpha, sensitivity analysis	.825, 95% bootstrap CI [.631, 1.000]	Sensitivity estimate only; category spacing is not assumed in the primary result
Pairwise unweighted Cohen's kappa	.798-.805	Pairwise agreement in category assignments
Pairwise observed agreement	.90 for each pair	Descriptive agreement, not validity
Judgment distribution	34 High, 25 Moderate, 1 Low/No	Plausibility ratings of the original combined game-level mappings
Main limitation	Three raters; one combined item per game	Does not identify whether range, mechanic, function, or facilitation caused disagreement

Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355

The point estimates indicate substantial agreement in category assignments, but the bootstrap intervals are wide and the panel was small. More importantly, reliability is not validity. Experts can apply a category consistently even when an item is poorly specified or when the mapping does not predict children's behavior. The appraisal therefore supports refinement of the coding language but cannot validate the proposed ranges, establish child outcomes, or confer cultural generalizability.

The combined rating format also limited diagnostic value. A Moderate judgment could reflect concern about the age range, the mechanic label, the inferred function, or the required scaffolding. A later validation study should evaluate these elements separately. The raw rating and calculation notes are important to the distribution rather than relying only on coefficients.

Practical Use and Claim Boundaries

AMF is best used as a structured conversation and documentation aid. An educator can identify what a game version appears to require, plan visual or verbal supports, and decide what behavior to observe. A designer can compare rule variants and record which change reduces memory, language, time-pressure, or coordination demands. A researcher can turn a mechanic-function claim into a testable hypothesis with an aligned comparator and outcome. In each case, the framework exposes assumptions; it does not assure an outcome.

Facilitation should be treated as a contextual variable rather than an automatic amplifier. Guided-play research suggests that adult structuring and scaffolding can affect task performance and selected learning outcomes, although effects vary by implementation, outcome, and comparator (Fisher et al., 2013; Skene et al., 2022; Weisberg et al., 2013). AMF therefore treats demonstration, prompting, pacing, visual supports, and cooperative support as implementation variables to record, not as proven moderators.

Cultural and equity conditions are similarly consequential. Competition, bluffing, adult authority, verbal participation, and cooperation may carry different meanings across communities. Observations of competitive play show context-dependent emotional and behavioral responses, while cross-cultural analysis shows variation in the organization of cooperative goals in games; neither establishes developmental outcomes for the present sample (Emen Parlatan & Sığırtmaç, 2022; Leisterer-Peoples et al., 2021). Access to commercial products, language support, time, quiet space, adult facilitation, and sensory or motor accommodations also affects what participation demands. The Indonesia-based panel provides contextual professional judgment, not national validity.

Clinical boundaries must remain explicit. AMF is not a developmental assessment, clinical instrument, treatment protocol, or substitute for professional evaluation. A listed function is a hypothesis about task demand or practice opportunity, not a promised therapeutic benefit. Any future use in clinical settings should remain subordinate to established assessment, safety, and treatment procedures until validity and efficacy are studied. Table 4 translates these boundaries into stakeholder-specific uses. Its purpose is to make appropriate documentation visible at the point of use, not to certify a game or a child.

Table 4. Bounded Uses of the AMF Framework

User	Appropriate AMF use	Evidence or context to record	Interpretation to avoid
Educator or caregiver	Compare game versions, anticipate demands, plan supports, and observe participation	Exact rules, prompts, duration, group structure, child response, accessibility, and emotional safety	Readiness diagnosis or promised developmental benefit

Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

User	Appropriate AMF use	Evidence or context to record	Interpretation to avoid
Game designer	Identify which rule or interface feature changes memory, language, motor, time-pressure, or social demands	Prototype version, information visibility, rule exceptions, component design, and playtest observations	Treating design intent or engagement as learning evidence
Researcher or evaluator	Convert a mechanic-function proposition into a prespecified, testable question	Population, comparator, dosage, fidelity, primary outcome, missing data, adverse events, and transfer measure	Choosing favorable outcomes after observation or generalizing beyond the tested implementation
Clinician or specialist	Use the matrix only as an exploratory discussion aid within established professional procedures	Individual needs, safety, consent, accommodations, validated assessments, and scope of practice	Clinical assessment, treatment claim, or substitution for validated tools

The table also clarifies what AMF can contribute before efficacy evidence exists. The framework is designed to make the reasoning behind game selection or adaptation, its assumptions, and any disconfirming observations explicit. This documentation is useful even when the eventual result is null, because it distinguishes failure of access, failure of implementation, failure of the practice hypothesis, and failure of transfer.

Proposed Next-Stage Validation

The next stage should first refine the framework through multidisciplinary consensus and then test child-level feasibility, predictive relations, and outcomes in separate prospective studies. Table 5 specifies two proposed studies; these procedures are not methods or results from the present study.

Table 5. Proposed Next-Stage Validation Roadmap

Stage	Proposed design	Prespecified outputs	What the stage cannot establish alone
Multidisciplinary Delphi refinement	Recruit 7-12 experts across early childhood education, developmental psychology, psychometrics, occupational therapy, and educational game design; conduct 2-3 anonymous rounds with controlled feedback; rate demand, range, opportunity, evidence, and facilitation separately	Item-level CVI, scale-level CVI, Aiken's V, distribution of ratings, revision history, stability/consensus stopping rule	Child feasibility, learning effect, causal mechanism, safety, or optimal age
Prospective child feasibility and outcome study	Obtain ethics approval, parent/guardian consent, and child assent; preregister age-stratified sampling, active comparison, dosage, facilitation, fidelity, construct-aligned outcomes, acceptability, frustration, adverse events, and attrition	Recruitment/retention, fidelity, feasibility, outcome estimates with confidence intervals, and transparent null/adverse findings	General effectiveness beyond the sampled ages, settings, mechanics, and implementation

Jurnal Kependidikan:
Jurnal Hasil Penelitian dan Kajian Kepustakaan
di Bidang Pendidikan, Pengajaran dan Pembelajaran
https://e-journal.undikma.ac.id/index.php/jurnalkependidikan/index
Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

A Delphi study should use anonymous rounds so status differences do not dominate discussion and should provide controlled summaries of prior-round distributions and comments (Hsu & Sandford, 2007). Experts should meet explicit experience criteria and declare relevant design or commercial interests. A four-point relevance and clarity scale would permit calculation of item- and scale-level content validity indices, while Aiken's V can quantify ordered expert judgment (Aiken, 1985; Polit et al., 2007). Retention, revision, and deletion thresholds, as well as a stopping rule based on stability and consensus, must be defined before the first round. No such panel or statistics are reported in this study.

The later instrument should evaluate individual AMF fields rather than repeat the original combined judgment. Experts should rate three core fields separately: clarity of the mechanic definition, plausibility of the participation demand, and appropriateness of the age range as a design prompt. They should also rate the plausibility of the practice opportunity, accuracy of the evidence classification, and completeness of the facilitation and boundary conditions. Round-level distributions and minority reasons should be retained alongside summary indices. An item that reaches a numerical threshold may still require revision if disagreement identifies a safety, accessibility, or cultural problem. Conversely, consensus cannot repair an item whose supporting evidence is absent.

Child-level work should begin only after the items and implementation conditions are sufficiently stable. A feasibility phase should examine whether children can access the rules and whether the proposed demand codes predict observable difficulty, support needs, or successful participation (Eldridge et al., 2016). A later outcome study should use age-overlapping groups, an active comparison condition, prespecified primary outcomes, implementation-fidelity logs, and independent measures that distinguish performance during play from transfer. Preregistration, ethics review, parent or guardian consent, developmentally appropriate assent, and reporting of frustration, adverse events, and attrition are essential. No child data were collected for the present framework-development study.

Prospective measurement should keep access and feasibility, in-play behavior, proximal transfer, and wider outcomes separate; success at one layer cannot substitute for evidence at another. This depends on indicators, sampling, fidelity, comparator, and interpretation rules.

The two stages answer different validity questions. Delphi consensus can test whether the framework's language and content coverage are defensible across disciplines; it cannot show that the framework predicts child performance. A feasibility study can test recruitment, acceptability, access, and implementation; it is not a small efficacy trial. Only a later adequately powered comparison can estimate outcome differences, and those estimates remain conditional on the sampled ages, context, mechanic implementation, facilitation, comparator, and measure. Later replication should include locally developed or traditional games and low-resource adaptations across cultural and accessibility contexts. Preserving this sequence prevents professional agreement from being reported as child evidence and prevents pilot data from being overinterpreted as effectiveness.

Conclusion

The AMF Framework offers a transparent way to analyze early childhood board games without converting plausible mechanics into unsupported outcome claims. It separates mechanic demand, possible practice opportunity, demonstrated outcome evidence, evidence directness, and contextual conditions and illustrates this vocabulary across 20 purposively selected games. Point estimates suggested substantial agreement in category assignments among three experts who rated the original combined game-level mappings. However, the wide bootstrap intervals and small panel limit interpretation. This agreement does not establish developmental validity, learning effects, age cutoffs, cultural generalizability, or clinical utility.

AMF should therefore be used as a provisional decision aid alongside manufacturer information, accessibility and safety considerations, and knowledge of the individual child. Its value at this stage lies in making assumptions inspectable and testable. Multidisciplinary Delphi refinement and ethically approved prospective studies with children are required before the framework can support claims about validity or effectiveness.

Recommendation

To operationalize the AMF framework as a provisional decision aid, early childhood education (ECE) teachers should use its structured criteria to design guided-play activities and scaffold game mechanics according to children's individual developmental needs rather than relying solely on manufacturer age labels; parents are recommended to treat the framework as an inspectable guide for selecting age-appropriate tabletop games while supporting their child's engagement through active facilitation, accessibility adjustments, and emotional co-regulation; and future researchers must prioritize conducting multidisciplinary Delphi studies to refine the framework's coding rules alongside ethically approved, prospective empirical trials with children to rigorously validate its developmental accuracy, learning transfer, and cross-cultural generalizability.

References

Aiken, Lewis R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Arifa, D., & Tyas, D. N. (2025). Developing An Educational Board Game with Digital Question Cards : A TGT Model Approach to Enhance Natural and Social Sciences Learning Outcomes. *Jurnal Kependidikan : Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran, Dan Pembelajaran*, *11*(2), 862–873.

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., de Freitas, S., Louchart, S., Suttie, N., Berta, R., & De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, *46*(2), 391–411. https://doi.org/10.1111/bjet.12113

Bawawa, M., Purwanti, R., & Pertiwi, S. A. (2025). Development of A Basic English Board Game on Local Culture as An Interactive Learning Media for Elementary Students in Rural Papua. *Jurnal Kependidikan : Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran, Dan Pembelajaran*, *11*(1), 156–166.

Best, J. R., & Miller, P. H. (2010). A Developmental Perspective on Executive Function. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Charifa, P., & Apriliani, I. (2025). Playing to Focus: A Systematic Review of Reveal-and-React Board and Card Games for Executive Function Development in Children. *Bulletin of Counseling and Psychotherapy*, *7*(2). https://doi.org/10.51214/002025071524000

Cowan, Nelson. (2016). Working Memory Maturation: Can We Get at the Essence of Cognitive Growth? *Perspectives on Psychological Science*, *11*(2), 239–264. https://doi.org/10.1177/1745691615621279

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(Volume 64, 2013), 135–168. https://www.annualreviews.org/content/journals/10.1146/annurev-psych-113011-143750

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., & Lancaster, G. A. (2016). CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ*, *355*. https://doi.org/10.1136/bmj.i5239

Emen Parlatan, M., & Sığırtmaç, A. D. (2022). Preschool children's behaviours in competitive games: are these games really harmful to social-emotional development? *Early Child Development and Care*, *192*(14), 2200–2212. https://doi.org/10.1080/03004430.2021.1990910

Evans, A. D., Xu, F., & Lee, K. (2011). When All Signs Point to You: Lies Told in the Face of Evidence. *Developmental Psychology*, *47*, 39–49. https://doi.org/10.1037/a0020787

Ferrara, K., Hirsh-Pasek, K., Newcombe, N. S., Golinkoff, R. M., & Lam, W. S. (2011). Block Talk: Spatial Language During Block Play. *Mind, Brain, and Education*, *5*(3), 143–151. https://doi.org/10.1111/j.1751-228X.2011.01122.x

Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking Shape: Supporting Preschoolers' Acquisition of Geometric Knowledge Through Guided Play. *Child Development*, *84*(6), 1872–1878. https://doi.org/10.1111/cdev.12091

Garon, N., Bryson, S., & Smith, I. (2008). Executive Function in Preschoolers: A Review Using an Integrative Framework. *Psychological Bulletin*, *134*, 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gashaj, V., Dapp, L. C., Trninic, D., & Roebers, C. M. (2021). The effect of video games, exergames and board games on executive functions in kindergarten and 2nd grade: An explorative longitudinal study. *Trends in Neuroscience and Education*, *25*, 100162. https://doi.org/https://doi.org/10.1016/j.tine.2021.100162

Girard, C., Ecalle, J., & Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *Journal of Computer Assisted Learning*, *29*(3), 207–219. https://doi.org/10.1111/j.1365-2729.2012.00489.x

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi Technique: Making Sense of Consensus. *Practical Assessment, Research, and Evaluation*, *12*(1). https://doi.org/10.7275/pdz9-th90

Hunicke, R., Leblanc, M., & Zubek, R. (2004). MDA: A Formal Approach to Game Design and Game Research. *AAAI Workshop - Technical Report*, *1*.

James-Brabham, E., Jay, T., & Sella, F. (2024). No Evidence That Playing a Linear Number Board Game Improves Numerical Skills Beyond Teaching as Usual: A Randomized Controlled Trial in 4- to 5-Year-Old Primary School Children. *Journal of Educational Psychology*, *116*, 411–425. https://doi.org/10.1037/edu0000857

Krippendorff, K. (2019). *Content Analysis: An Introduction to Its Methodology* (Fourth Edition). SAGE Publications, Inc. https://doi.org/10.4135/9781071878781

Landis, J. R., & Koch, G. G. (1977). An Application of Hierarchical Kappa-type Statistics in the Assessment of Majority Agreement among Multiple Observers. *Biometrics*, *33*(2), 363–374. https://doi.org/10.2307/2529786

Laski, E., & Siegler, R. (2014). Learning From Number Board Games: You Learn What You Encode. *Developmental Psychology*, *50*(3), 853–864. https://doi.org/10.1037/a0034321

Leisterer-Peoples, S. M., Ross, C. T., Greenhill, S. J., Hardecker, S., & Haun, D. B. M. (2021). Games and enculturation: A cross-cultural analysis of cooperative goal

Vol. 12, No. 3 : September 2026
E-ISSN: 2442-7667
pp. 1343-1355
Email: jklppm@undikma.ac.id

structures in Austronesian games. *PLOS ONE*, *16*(11), e0259746-. https://doi.org/10.1371/journal.pone.0259746
Levine, S., Ratliff, K., Huttenlocher, J., & Cannon, J. (2011). Early Puzzle Play: A Predictor of Preschoolers' Spatial Transformation Skill. *Developmental Psychology*, *48*, 530–542. https://doi.org/10.1037/a0025913
O'Neill, D., & Holmes, P. (2022). The Power of Board Games for Multidomain Learning in Young Children. *American Journal of Play*, *14*, 58–98.
Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199
Ramani, G. B., Siegler, R., & Hitti, A. (2012). Taking It to the Classroom: Number Board Games as a Small Group Learning Activity. *Journal of Educational Psychology*, *104*, 661–672. https://doi.org/10.1037/a0028995
Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology*, *32*(3), 146–159. https://doi.org/10.1016/j.appdev.2011.02.005
Skene, K., O'Farrelly, C. M., Byrne, E. M., Kirby, N., Stevens, E. C., & Ramchandani, P. G. (2022). Can guidance during play enhance children's learning and development in educational contexts? A systematic review and meta-analysis. *Child Development*, *93*(4), 1162–1180. https://doi.org/https://doi.org/10.1111/cdev.13730
Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
Sousa, C., Rye, S., Sousa, M., Torres, P. J., Perim, C., Mansuklal, S. A., & Ennami, F. (2023). Playing at the school table: Systematic literature review of board, tabletop, and other analog game-based learning approaches. *Frontiers in Psychology*, *14*, 1160591. https://doi.org/10.3389/fpsyg.2023.1160591
Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5
Vita-Barrull, N., Estrada-Plana, V., March-Llanes, J., Guzmán, N., Fernández-Muñoz, C., Ayesa, R., & Moya-Higueras, J. (2023). Board game-based intervention to improve executive functions and academic skills in rural schools: A randomized controlled trial. *Trends in Neuroscience and Education*, *33*, 100216. https://doi.org/10.1016/j.tine.2023.100216
Weisberg, D. S., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Guided Play: Where Curricular Goals Meet a Playful Pedagogy. *Mind, Brain, and Education*, *7*(2), 104–112. https://doi.org/https://doi.org/10.1111/mbe.12015
Whittemore, R., & Knafl, K. (2005). The integrative review: updated methodology. *Journal of Advanced Nursing*, *52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x
Wouters, P., Nimwegen, C., Oostendorp, H., & Spek, E. (2013). A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. *Journal of Educational Psychology*, *105*, 249–265. https://doi.org/10.1037/a0031311
Zelazo, P. D., & Carlson, S. M. (2012). Hot and Cool Executive Function in Childhood and Adolescence: Development and Plasticity. *Child Development Perspectives*, *6*(4), 354–360. https://doi.org/10.1111/j.1750-8606.2012.00246.x

