International Journal of Arts and Social Science www.ijassjournal.com
ISSN: 2581-7922,
Volume 9 Issue 1, January 2026

Preschool Inductive Reasoning Training Via Digital Games: Design, Feasibility, and Iterative Insights

Yixuan Li[1], Rui Fang[2], Wei Sun[3]
[123](School of Digital Media and Design Arts, Beijing University of Posts and Telecommunications, China)

ABSTRACT: *This study develops a four-mechanism inductive reasoning game concept for 3–6-year-olds, targeting appearance-based, category-based, thematic, and causal induction. Using static interface materials, the concept was evaluated for feasibility and acceptability by 3 cross-disciplinary experts and 26 caregivers/educators. Results show moderate-to-high feasibility (expert grand mean=3.64) and acceptability (user adoption intention=4.27), with causal (G4) and appearance-based (G1) sub-games performing better than category (G2) and thematic (G3) ones. Five iteration priorities were identified, including reducing entry cost, redesigning G2's boundaries, and visualizing G3's abstract "stability" construct. This study provides a developmentally aligned game design framework and a pragmatic pre-implementation evaluation pathway for educational game developers.*

KEYWORDS *- Inductive reasoning, game-based learning, preschool education, cognitive development*

I. INTRODUCTION

With the rapid diffusion of AI-enabled learning technologies, game-based educational applications have become a common medium for supporting children's early cognitive development. Inductive reasoning is widely recognized as a foundational component of children's logical thinking, underpinning concept learning, generalization, and problem solving across domains [1, 2]. Prior studies suggest that inductive reasoning in preschool and early primary years remains highly dependent on concrete contexts, salient perceptual cues, and immediate action–feedback contingencies [3, 4]. Accordingly, a persistent design challenge is how to embed inductive reasoning practice into activities that children can readily understand and sustain, while remaining acceptable and manageable for families and educators [5].

Game-based learning offers a practical bridge between abstract cognitive training and children's everyday experiences. Compared with conventional instruction, digital games can leverage rules, feedback, and contextual variation to support iterative hypothesis testing and rule discovery [6]. In this context, visually grounded inductive reasoning games have gained attention; however, existing work often emphasizes effectiveness testing or technical implementation, while giving less emphasis to early-stage feasibility and acceptability of design proposals—especially when an interactive prototype is not yet available [7]. Under realistic development constraints, a clear and defensible evaluation pathway for design concepts remains necessary.

To address this gap, this study develops a four-mechanism inductive reasoning game concept for children aged 3–6 and evaluates it at the proposal stage using static interface materials. The app concept comprises four sub-games targeting appearance-based induction, category-based induction, thematic induction, and causal induction. The present paper focuses on research-grounded design rationale and early-stage feasibility/acceptability evidence, rather than claiming learning effectiveness.

Accordingly, this study is guided by three research questions: (RQ1) How do cross-disciplinary experts judge the goal alignment, mechanism rationality, and implementability of the four sub-games based on static design materials? (RQ2) How acceptable do caregivers and frontline educators perceive each sub-game to be in terms of comprehensibility, feedback clarity, replay suitability, and required adult support? (RQ3) What convergent iteration priorities emerge from triangulating expert comments and user questionnaire evidence?

II. RELATED WORK

Research in developmental and cognitive psychology indicates that children's inductive judgments draw on multiple sources of information. Younger children often rely on perceptual similarity (e.g., color, shape, spatial layout), while older children increasingly incorporate category knowledge, relational structure, and causal information when making generalizations [3, 4]. This developmental trajectory implies that inductive reasoning practice should be aligned with children's cognitive characteristics and supported by concrete operations and contextual cues rather than abstract rule statements.

From a pedagogical perspective, constructivist learning theory emphasizes knowledge construction through active engagement with tasks and environments, where feedback helps learners refine their mental models [8]. This theoretical stance provides a strong rationale for game-based learning designs that emphasize short iterative cycles, visible outcomes, and supportive prompts, enabling children to practice reasoning without heavy explicit instruction [6, 8].

In human–computer interaction and educational game design, early-stage validation methods have been increasingly discussed, including concept comprehensibility, interface intuitiveness, and user acceptance [7, 9]. When full functionality is unavailable, presenting scenario descriptions and representative screens and collecting structured subjective evaluations has been considered a pragmatic approach for pre-implementation validation [10]. These findings inform the methodological choice in this study to evaluate feasibility and acceptability using a static UI package.

Notably, the selection of the four inductive reasoning mechanisms (appearance-based, category-based, thematic, and causal) is grounded in developmental research on how preschoolers shift from perceptual cues toward more concept- and relation-based generalization. In this paper, G1 emphasizes perceptual similarity (typically salient at ages 3–4), G2 and G3 support the transition toward category- and theme-based generalization (roughly ages 4–6), and G4 introduces simple, highly observable cause-effect regularities suitable for older preschoolers (around ages 5–6).

III. METHOD

This study followed a development-oriented research approach, spanning needs analysis, concept design, static interface prototyping, feasibility evaluation, and iterative refinement. The evaluation was positioned as formative evaluation, aiming to verify feasibility and acceptability of the proposed design at the current stage rather than to test learning outcomes. A mixed-methods analysis strategy was adopted, combining quantitative Likert scale data with qualitative open-ended comments to triangulate findings.

3.1 Overview of Methods

Table 1 summarizes the methods employed across stages and their intended purposes.

Table 1. Overview of methods and purposes

Stage	Method	Participants / Data	Purpose
Needs analysis	Questionnaire (to identify typical users); Semi-structured interviews (online); Grounded-theory coding (open/axial/selective)	Parents of children aged 3–6; Typical parents (n=5); Interview transcripts	Capture usage patterns and major concerns to support interview sampling; Elicit pain points, expectations, and desired AI-support roles; Derive a demand

			structure to inform design decisions
Concept design	Mechanism design and interaction specification	Researcher	Define four sub-games and shared interaction logic
Prototyping	Static UI package + description cards	Researcher	Prepare representative screens and concise gameplay explanations
Feasibility review	Expert Likert-scale evaluation	Experts (n=3, cross-disciplinary)	Assess goal alignment, mechanism rationality, usability, implementability
Acceptability evaluation	User questionnaire (same items per sub-game)	Users close to target children (n=26)	Assess comprehensibility, interest, burden/frustration risk, accompaniment cost
Refinement	Synthesis of quantitative + qualitative feedback	Researcher	Revise rules, prompts, and information presentation

3.2 Needs Analysis (Questionnaire + Interviews)

Prior to concept development, a parent-facing questionnaire was used to understand children's usage contexts for logic-learning apps, caregivers' involvement, and perceived difficulties. Typical users were then selected based on two criteria: (1) the child had relatively rich experience using children's logic-learning apps, and (2) the caregiver had participated in the child's learning process. Semi-structured interviews were conducted with the selected parents to obtain deeper insights into pain points and expectations.

Because children aged 3–6 are difficult to interview reliably and may be distracted by the interview process, parents were used as proxy informants. Interviews were conducted online via WeChat voice calls, lasting approximately 20–30 minutes each. The interview guide focused on four themes: (a) major difficulties children encounter during use, (b) caregivers' concerns and support challenges, (c) expectations for an ideal logic-learning app, and (d) preferred roles of an AI agent (e.g., reading prompts aloud, explaining errors, answering questions, guiding thinking, and offering encouragement).

Interview transcripts were analyzed using grounded-theory procedures (open, axial, and selective coding) to consolidate user needs into design-relevant categories. The synthesis produced three minimal, design-oriented conclusions: (1) difficulty and task pacing are often poorly matched to children's abilities, leading to frustration; (2) interaction guidance and explanatory feedback are insufficient, reducing children's independence and increasing caregiver involvement; (3) sustaining attention and motivation in short home-learning sessions is challenging, and caregivers expect lower accompaniment cost and emotionally supportive assistance. These conclusions informed both the sub-game mechanisms and the evaluation dimensions used in this paper.

3.3 Materials

As an interactive prototype was not available at the current stage, the evaluation used a static UI package. Materials consisted of representative interface images for each sub-game and a brief gameplay description card (approximately 100–150 words) per sub-game. Each sub-game included a small set of key screens illustrating the core task state and feedback state(s). All materials were presented in a fixed order in either printed or digital format to ensure consistent exposure across participants.

3.4 Participants

Two participant groups were involved. The expert group included three cross-disciplinary professionals: Expert A (a preschool education researcher with 10 years of experience in early childhood logic education), Expert B (from the Xiaotiancai Education Team, specializing in educational game design with 6 years of practice), and Expert C (a senior interaction designer, focusing on children's digital products for 12 years). This sample size is consistent with common practices in early-stage educational technology concept validation, ensuring coverage of educational, design, and practical perspectives [11].

The user group consisted of 26 adults closely connected to the target children. Inclusion criteria were: (1) having frequent contact with children aged 3–6 (at least 3 times per week), (2) accompanying the child to use educational apps at least once a week in the past 6 months, and (3) being able to accurately describe the child's usage experience. Participants included parents (57.69%), kindergarten teachers (23.08%), after-school education practitioners (7.69%), and others (11.54%); the children they interacted with most frequently were aged 3–4 (46.15%) and 4–5 (38.46%).

3.5 Research Instruments

Two primary instruments were used and refined through pre-testing. The expert Likert-scale questionnaire (5-point) drew on established educational-game evaluation dimensions (e.g., usability/clarity and implementation suitability) and content-validity practices to assess four dimensions: goal alignment, mechanism rationality, usability (ease of understanding and clarity of prompts), and implementability (home/classroom applicability). A pre-test with 2 experts yielded a content validity index (CVI) = 0.89, indicating good content validity [12, 13].

The user questionnaire (5-point) was adapted from commonly used acceptance and app-evaluation dimensions for preschool learning apps and educational games (e.g., comprehensibility, interest, burden/frustration risk, and adult accompaniment cost) [5, 7, 9]. The same item set was used for each sub-game (6 items per sub-game). A pre-test with 8 users (consistent with the formal user group) showed Cronbach's α = 0.78 for the entire questionnaire; Cronbach's α for each sub-dimension ranged from 0.72 to 0.81, indicating acceptable internal consistency.

3.6 Research Procedure

Participants first reviewed the interface images and gameplay description cards (10–15 minutes). User participants then completed the questionnaire for all four sub-games (15–20 minutes). Experts completed the expert evaluation scale and provided open-ended suggestions (20–30 minutes). Quantitative scores and qualitative comments were synthesized to identify priority issues and to guide revisions to rules, prompting strategies, and information presentation.

3.7 Data Processing and Analysis

All items used a five-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). The frustration concern item was reverse-coded only when computing composite feasibility/acceptability scores, using: Recoded Score = 6 − Raw Score (so higher composite scores indicate lower frustration risk). For transparency, Table 5 reports the raw frustration concern item (higher = more concern). For each sub-game, dimension-specific mean scores were calculated using the following formulas:

(1) Comprehensibility Score = (Item 1 Score + Item 3 Score) / 2

(2) Interest/Willingness Score = Item 4 Score

(3) Low Burden/Low Frustration Score = (Item 2 Score + Recoded Item 5 Score) / 2

(4) Low Accompaniment Cost Score = Item 6 Score

The overall feasibility/acceptability score for each sub-game was computed as: Overall Score = (Item 1 + Item 2 + Item 3 + Item 4 + Recoded Item 5 + Item 6) / 6. Results were reported as mean ± standard deviation (M ± SD) and agreement rates (percentage of responses with scores $\geq$ 4, indicating positive evaluation).

Qualitative open-ended comments were analyzed using double coding. Two researchers independently coded the comments into pre-defined categories (e.g., interface clarity, feedback effectiveness, difficulty matching) and emerging categories (e.g., AI role expectations). The inter-coder reliability (Kappa coefficient) was 0.83, indicating high consistency; discrepancies were resolved through group discussion to ensure the reliability of qualitative findings.

IV. APP CONCEPT DESIGN AND STATIC PROTOTYPING

This study transforms inductive reasoning training into four game forms that are easy to operate, visualized, and open to repeated trial and error, integrated into a single application system through a unified "frame" design. Focusing on gameplay logic rather than interface presentation, this chapter aims to clarify how to convert children's inductive reasoning processes into game cycles featuring low reading load, short rounds, and sustainable participation. This enables children aged 3–6 to conduct multiple rounds of exploration within a closed loop of "identifying clues, making choices, understanding results, and continuing to try," gradually developing the ability to generalize and transfer rules through repeated comparisons.

To achieve this goal, the four sub-games share a set of "short-round inductive processes": each round first presents perceptible key differences as clues for children, guides them to make a single decision via tapping or dragging, then provides feedback through intuitive changes to help children establish a cognitive connection between "choice and consequence," and finally forms a continuous practice chain by progressing to the next round or allowing a retry. The core value of this design lies in transforming inductive reasoning from a "rule-explaining" model to a "self-discovery" model, enabling children to participate in reasoning practice without relying on complex language while reserving room for subsequent difficulty progression and adaptive support.

Within the unified process framework, the core differences between the four sub-games lie in two aspects: "sources of rules" and "methods of difficulty progression." The specific designs are as follows:

4.1 Sub-Game1 Narratives

Core Objective

Guide children to extract common features of similar things from perceptible attributes such as color, shape, and size, thereby completing matching or supplementation tasks.

Gameplay Design

Children first observe the implicit common features in the existing set of elements, then select the object that best fits these features from a limited number of candidates to complete the task. Results are presented in real-time visualization, allowing them to verify or revise their judgments before proceeding to the next round.

Difficulty Progression Logic

Difficulty is increased through controllable variables rather than adding complex explanations: first, the feature dimension expands from a single one (e.g., color only) to multi-dimensional combinations (e.g., color + shape); second, the similarity between distractors and target items is gradually increased to enhance children's ability of refined comparison and differentiation. Meanwhile, the scale of candidate options is controlled to balance exploration space and cognitive load. The core of the design is to enable children to stably master appearance-based rule-judging strategies through multiple rounds via "visual comparison" and "low-threshold trial and error," rather than relying on occasional correct single choices.

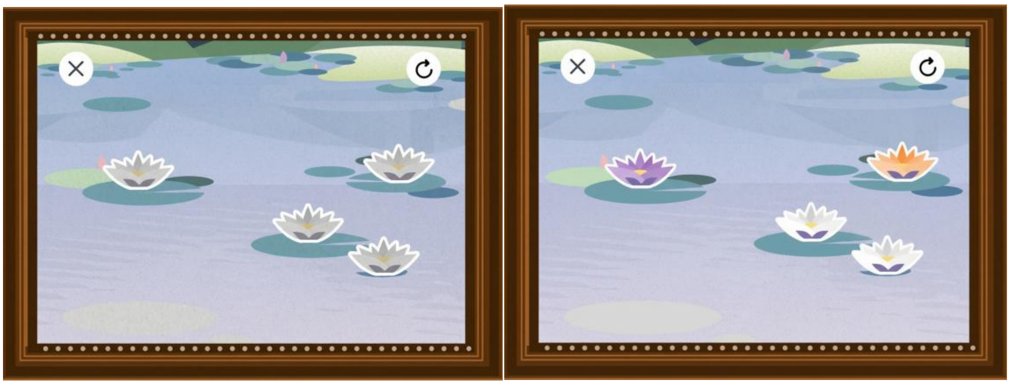

Fig. 1 Rendering of Sub-game 1 Interface

4.2 Sub-Game2 Narratives

Core Objective

Help children establish an understanding of category boundaries, form stable criteria for judging membership, and clarify the "basis for classifying similar things" through repeated placement and revision.

Gameplay Design

Children face multiple grouping targets (different sets or phased categorization tasks), select appropriate objects from candidate elements to place into corresponding category sets; real-time feedback is provided after each placement, allowing children to adjust subsequent choices until they complete the current phase of the categorization task and move to the next phase. Unlike G1, the core of G2 is a continuous membership judgment process, requiring children to repeatedly verify the "consistency of classification criteria and clarity of boundaries" through multiple decisions to form stable rules.

Difficulty Progression Logic

A difficulty gradient is achieved through variable control: initial category criteria start with intuitive and concrete clues (e.g., animals/plants, aquatic/terrestrial), gradually transitioning to comprehensive category judgments; the heterogeneity of candidate sets and the proportion of distractors are gradually increased, requiring children to accurately distinguish between "cross-category similarity" and "intra-category differences"; the scale of categorization gradually expands from a small number to a larger one, training children's ability to maintain rules and make continuous judgments. The core of the design is to make "category boundaries" a stable, verifiable standard, enabling children to establish classification rules through independent attempts rather than relying on external explanations from adults.

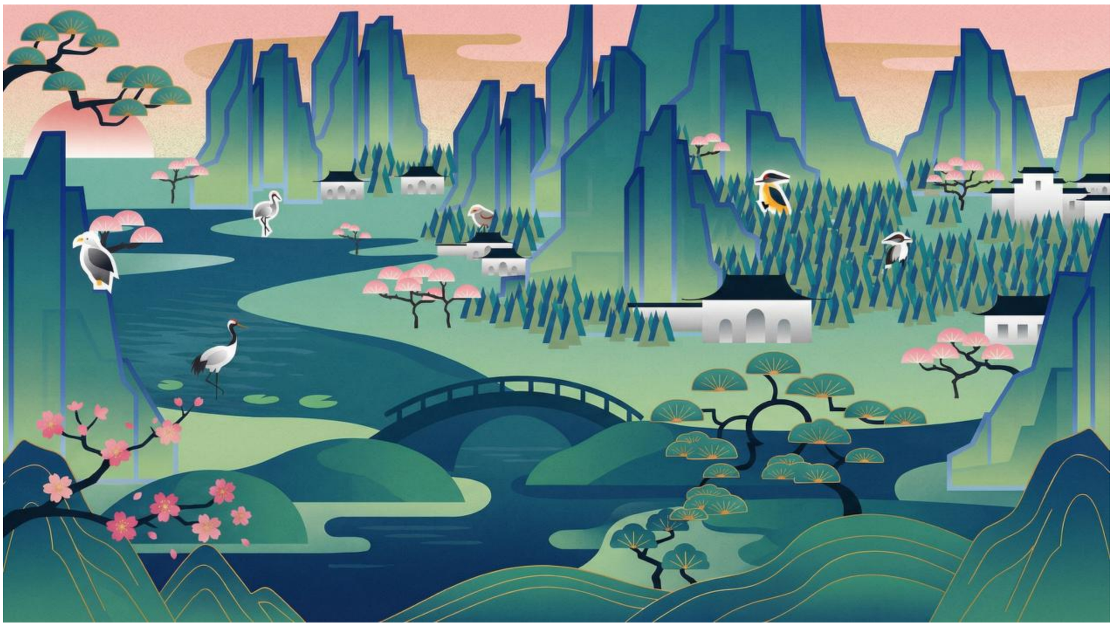

Fig. 2 Rendering of Sub-game 2 Interface

4.3 Sub-Game3 Narratives

Core Objective

Help children understand the connotation of "thematic relevance," practice effective selection and distractor elimination in distracting environments, and form a judgment logic adapted to themes.

Gameplay Design

The game presents a clear thematic scene (e.g., a garden), and children select highly theme-relevant objects from continuously updated candidate elements to supplement the scene; after each selection, the "thematic fit" of the scene changes intuitively, and children adjust their subsequent choices based on this change; when the thematic fit reaches a preset threshold and maintains for a certain period, the current level is completed and the next phase is entered. The core of the gameplay is "continuous selection" rather than single-point judgment, transforming "thematic relevance" from an abstract concept into an operable rule that children can independently induce.

Difficulty Progression Logic

Difficulty progression focuses on three dimensions: thematic clues gradually transition from explicit to implicit (e.g., from "garden" to "rainy day"), increasing the depth of reasoning; the proportion of distractors is gradually increased, enhancing children's ability to distinguish between "superficial relevance" and "substantive relevance"; the scene's target threshold and the update rate of candidates are gradually raised, increasing information processing pressure and training the stability of strategies. The core of the design is to replace "binary right/wrong judgment" with "phased achievement," allowing children to form theme-related rules through continuous selection and develop the ability to adhere to and revise rules in distracting environments.

Fig. 3 Rendering of Sub-game3 Interface

4.4 Sub-Game4 Narratives

Core Objective

Guide children to establish predictive abilities through the connection between "operation and result," revise hypotheses through attempts, and initially understand simple causal chains (e.g., A→B→C).

Gameplay Design

Children select or combine natural elements (e.g., sun, water droplets, stones) to trigger visual changes in target objects (e.g., plant growth, pattern unfolding), thereby perceiving the "correspondence between operation and result"; gradually eliminate ineffective or negatively acting elements through multiple attempts to form stable causal judgments; when the target state is achieved (e.g., mature plant) or the specified number of effective combinations is completed, the next level is entered. The core advantage of G4 is "highly visualized results," which naturally adapts to children's learning path of "proposing hypotheses - rapid verification - revising cognition."

Difficulty Progression Logic

Difficulty progression does not increase the cognitive load and is mainly achieved through three dimensions: adding ineffective or negatively acting elements to enhance comparison and elimination abilities; transitioning from single-element causal relationships to dual-element combined causality to increase the difficulty of combinatorial reasoning; expanding from one-step causal chains (e.g., watering → plant growth) to multi-step causal chains (e.g., sun exposure → soil warming → watering → rapid growth), testing predictive abilities and memory load. The core of the design is to transform inductive reasoning into an "experiential experimental process," enabling children to establish explainable and predictable causal rules through independent operations rather than mechanically memorizing answers.

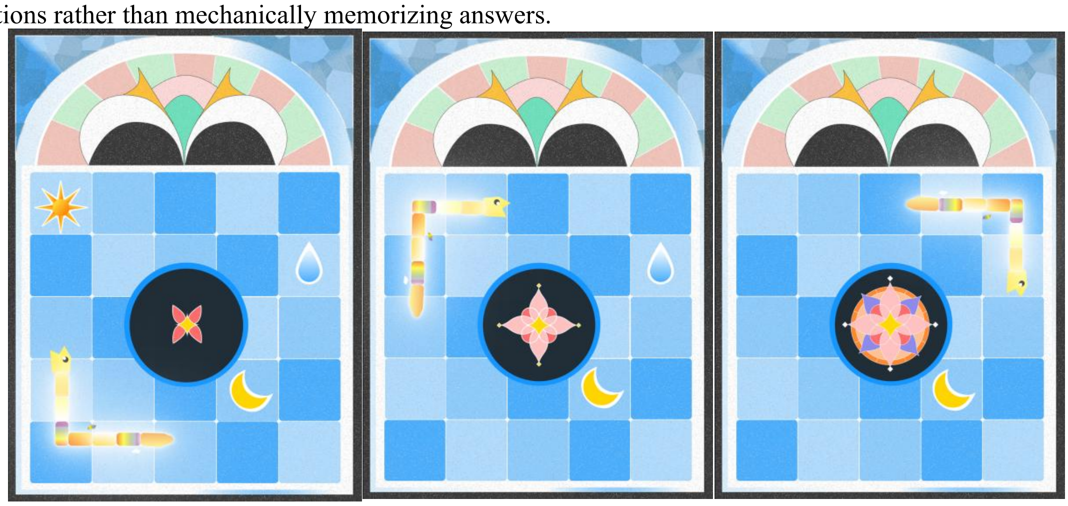

Fig. 4 Rendering of Sub-game4 Interface

4.5 Unified Design Principles

To reduce learning costs and support continuous participation, the four sub-games follow unified principles in process organization: each round contains only one key judgment to avoid attention loss caused by long processes; task objectives adopt a phased achievement logic to ensure children always have a clear direction; difficulty gradually transitions from "explicit clues and few distractions" to "implicit clues and more distractions," with consistent mechanisms and progressively increasing load. This not only covers the ability differences of children aged 3–6 but also provides a foundation for subsequent personalized support and dynamic adjustments.

In summary, this chapter transforms four types of inductive reasoning training into implementable game forms at the gameplay mechanism level, unifying them into a core structure of "short rounds, low reading load, and repeated trial and error." This clarifies the core objects for subsequent formative evaluation and iteration: the evaluation focus is not on the completeness of interface details, but on the clarity of each game cycle, children's continuous participation, and the ability of difficulty variables to support the progressive learning process from initiation to improvement.

V. RESULTS AND DISCUSSION

5.1 Expert Evaluation Results

Three cross-disciplinary experts evaluated the four inductive-reasoning sub-games using a 5-point Likert scale (32 items per expert). As shown in Table 3, 89 out of 96 expected item ratings were usable (92.7% completeness). Missing values mainly occurred when experts selected "NA/unable to judge" for items that required more concrete evidence of scaffolding or feedback implementation beyond the static materials.

Table 3. Expert dataset and completeness (Likert 1–5; NA excluded)

Expert source	Usable ratings / 32	Notes
Expert A (Preschool education researcher)	26	6 items marked NA (related to dynamic feedback details)
Expert B (Educational game designer, Xiaotiancai Education Team)	31	1 item marked NA; detailed open-ended feedback provided
Expert C (Senior interaction designer, children's digital products)	32	Complete; focused on interface intuitiveness comments
Overall	89 / 96	92.7% completeness

Across all usable items, the expert panel rated the suite as moderately strong (grand mean = 3.64, SD = 1.13). At the sub-game level, causal induction (G4) and perceptual similarity (G1) received relatively higher ratings, suggesting more legible rules and interpretable "try–observe–infer" feedback loops. Category induction (G2) and thematic induction (G3) scored lower and showed greater variability, indicating that experts diverged primarily when correct play depended on scene parsing (G2) or abstract constructs such as "stability" (G3).

Table 4 highlights representative items that help localize revision priorities. The lowest-performing anchors were scene readability and rule/criteria clarity in G2, implying that children may struggle before meaningful category induction can occur. Expert open-ended comments further noted that G2's unclear region partitioning and high scene density made category boundaries ambiguous, and some category criteria may be mismatched to the target age range (3–6 years old). For G3, weaker ratings clustered around the comprehensibility of the stability indicator and the legibility of action–effect mapping; 2 out of 3 experts pointed out that the shaking effect may be too subtle for young children without additional cues and suggested replacing it with a more direct visual indicator (e.g., a progress bar). Meanwhile, the highest means appeared in G4's exploration payoff, reinforcing the strength of its immediate and intuitive outcome feedback.

Table 4. Cross-item diagnostic panel (representative items explaining key weaknesses)

Construct	**Representative item**	**Mean**	**SD**	**n (non-NA)**
Entry clarity	Goal is immediately understandable	3.67	0.88	6
Scene readability	Region boundaries are clear (G2)	2.67	0.58	3
Rule/criteria clarity	Classification rule is understandable (G2)	2.33	1.53	3
Grounding abstract constructs	Stability meter is understandable (G3)	2.67	1.53	3
Feedback legibility	Action → change is clear (G3)	2.67	1.53	3
Motivation / replay	Fun & replayability (G1)	3.00	1.00	3
Exploration payoff	Outcome changes are intuitive (G4)	4.50	0.71	2

Open-ended comments further triangulated these patterns, with one expert explicitly prioritizing G2 for redesign due to unclear region partitioning, high scene density, and potentially age-mismatched category judgments. Additional suggestions emphasized lightweight scaffolding, clearer incentive structure, eye-friendly visual strategies, and more salient feedback cues.

5.2 User Questionnaire Results

A questionnaire-based acceptability assessment was conducted based on the static interface package (n = 26). Respondents were adults closely connected to the target children, including parents (57.69%), kindergarten teachers (23.08%), after-school education practitioners (7.69%), and others (11.54%). The children they most frequently interacted with were mainly aged 3–4 (46.15%) and 4–5 (38.46%). Participants rated each sub-game on a 5-point Likert scale covering task clarity, ease of getting started, feedback clarity, replay suitability, frustration concern, and autonomy without extensive adult instruction.

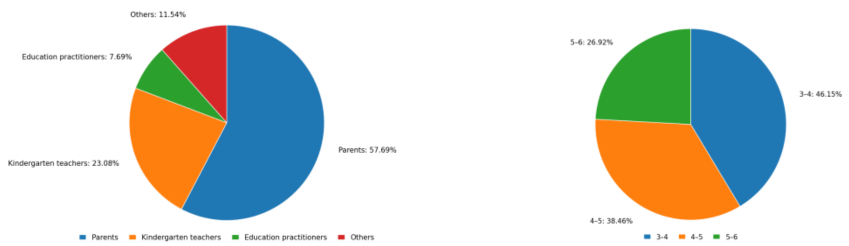

Fig. 5 Participant Identity Composition & Age Distribution of Contacted Children

Overall ratings showed higher adoption intention than perceived attractiveness: overall attractiveness averaged 3.73, willingness to try in home/classroom settings averaged 4.27, and perceived feasibility for deployment as a real app averaged 3.88. This pattern indicates broad acceptability of the concept, while suggesting that sustained use may depend on lowering first-time comprehension costs and strengthening progression and motivation mechanisms.

Fig. 6 Site Photographs of the Children's Experience Product Experiment

Table 5. Summary of user Likert ratings across four sub-games (means)

Dimension (Mean)	G1 Appearance similarity	G2 Category induction	G3 Thematic association	G4 Causal induction
Task is easy to understand	3.42	3.15	3.42	3.73
Easy to get started	3.96	3.50	3.85	3.96
Feedback is clear	3.54	3.73	3.73	3.88
Suitable for replay	3.92	3.77	3.69	4.04
Frustration concern (higher = more concern)	3.58	3.46	3.77	3.42
Can continue without long adult instruction	3.58	3.35	3.42	3.73

At the sub-game level, G4 performed best across the understanding–feedback–replay chain and showed the highest autonomy score, suggesting an easier-to-grasp action–outcome loop. G2 scored lowest on task clarity and autonomy, indicating that the primary bottleneck lies in boundary and criteria comprehension rather than feedback visibility—consistent with expert comments. G3 showed the highest frustration concern; user open-ended feedback (34.6% of respondents) noted that "the stability shaking is hard to interpret, and repeated incorrect choices lead to confusion", calling for lower error cost and lightweight scaffolding. For G4, although quantitative scores were the highest, 15.4% of users mentioned that "the outcome feedback is too fast for young children to observe and infer the causal relationship", indicating a need to optimize pacing and state cues. Open-ended responses also echoed the need to clarify G2 boundaries and improve feedback salience for all sub-games.

5.3 Design Implications and Discussion

Expert review and user evidence converge on five iteration priorities, with triangulated support from both quantitative data and qualitative comments:

(1) Reduce first-time comprehension cost across all sub-games using minimal, low-reading scaffolding (one-line prompts, pointing cues, brief exemplar demonstrations). This addresses the moderate entry clarity scores (mean=3.67) and user feedback about "needing adult explanation to understand the goal initially."

(2) Prioritize G2 redesign by strengthening region boundaries (e.g., using distinct colors or outlines), lowering scene density (reducing irrelevant elements), and clarifying age-appropriate category criteria (e.g., starting with concrete categories like "animals/plants" instead of abstract ones). This targets the lowest scores in G2's scene readability (mean=2.67) and rule clarity (mean=2.33).

(3) Ground the "stability" construct in G3 with self-evident indicators (e.g., a progress bar or glowing effect) and stronger action–effect mapping. This responds to the low comprehensibility scores of the stability meter (mean=2.67) and comments about "unrecognizable shaking effects."

(4) Polish G4 and G1 for retention by improving pacing/state cues (G4: slowing feedback speed to allow observation) and sharpening correctness cues with gentler error recovery (G1: enhancing error prompts to highlight key features).

(5) Unify progression and incentives across sub-games through consistent milestones and light rewards (e.g., completion feedback, points/badges) to support replay without diluting inductive reasoning training goals. This aligns with the gap between attractiveness (mean=3.73) and adoption intention (mean=4.27).

VI. CONCLUSION

This study presents the design, formative evaluation, and iterative implications of a four-mechanism inductive reasoning game concept tailored for children aged 3–6 years, with a focus on early-stage feasibility and acceptability evidence using static UI materials. The research contributes a developmentally aligned mapping between inductive mechanisms and game interactions, and demonstrates a pragmatic concept-validation pathway that can reduce late-stage redesign risk in educational game development.

The core findings of this study can be summarized as follows. First, the four sub-games demonstrate moderate-to-high feasibility and acceptability among cross-disciplinary experts and users closely connected to the target group, with a grand mean of 3.64 (expert evaluation) and 3.73–4.27 (user acceptability dimensions). Second, G4 and G1 were comparatively stronger in perceived rule legibility and action–outcome mapping, while G2 and G3 exposed higher first-time comprehension costs. Third, triangulated feedback converged on five actionable optimization strategies, emphasizing reduced entry cost, targeted G2 boundary/criteria redesign, explicit visualization of abstract constructs in G3, pacing refinement, and unified progression mechanisms—providing a clear roadmap for prototype development and subsequent effectiveness studies.

The theoretical contributions of this study lie in integrating developmental psychology theories of inductive reasoning with educational game design principles, systematically mapping four core inductive mechanisms to age-appropriate game interactions. This framework enriches the theoretical foundation for designing cognitive training tools for young children, highlighting the importance of aligning game mechanics with the developmental sequence of inductive capabilities (from perceptual dependence to abstract thinking). Practically, the study establishes a pragmatic pre-implementation evaluation pathway using static materials, which can be adopted by designers and researchers to validate educational game concepts under resource constraints, reducing the risk of late-stage redesign and improving development efficiency.

6.1 Limitations

Despite its contributions, this study has several limitations that should be acknowledged. First, the evaluation relies solely on static UI materials (interface images and description cards), which cannot capture dynamic interaction fluency, real-time feedback effectiveness, or children's actual operational behaviors. Second, the user sample reflects caregivers' and educators' perceptions rather than direct testing with the target children (aged 3–6), which limits ecological validity because adult judgments may not fully match children's actual comprehension and engagement. Third, the study adopts a cross-sectional formative design and does not evaluate learning gains or long-term retention. Fourth, the expert panel ($n = 3$), though cross-disciplinary, is small, which may restrict generalizability of expert feedback.

6.2 Future Work

Future research will build on the current findings to address these limitations and advance the game concept. First, develop a fully interactive prototype based on the five optimization strategies identified in this study, incorporating enhanced region boundaries and clearer category cues for G2, explicit progress visualization for G3's stability construct, adjusted feedback pacing for G4, and unified progression and reward mechanisms across all sub-games. Second, conduct usability testing with the target child population (aged 3–6) using mixed methods (behavioral observation, engagement duration, optional eye-tracking where feasible, and age-appropriate interview prompts) to capture children's direct experience and iteratively refine the prototype. Third, implement a short-term intervention study to explore the impact of the optimized game on children's inductive reasoning skills (e.g., pre/post tasks aligned with the four mechanisms) and compare outcomes with a control condition using conventional training materials. Finally, explore adaptive support strategies (e.g., difficulty adjustment and just-in-time prompts) that personalize scaffolding based on children's performance while keeping cognitive demands age-appropriate.

REFERENCES

[1] V.M. Sloutsky, A.V. Fisher, Induction and categorization in young children: A similarity-based model, Journal of Experimental Psychology: General, 133(2), 2004, 166-188.

[2] S.A. Gelman, E.M. Markman, Categories and induction in young children, Cognition, 23(3), 1986, 183-209.

[3] V.M. Sloutsky, H. Kloos, A.V. Fisher, When looks are everything: Appearance similarity versus kind information in early induction, Psychological Science, 18(2), 2007, 179-185.

[4] J.E. Opfer, M.J. Bulloch, Causal relations drive young children's induction, naming, and categorization, Cognition, 105(1), 2007, 206-217.

[5] S.F. Griffith, M.B. Hagan, P. Heymann, B.H. Heflin, D.M. Bagner, Apps as learning tools: A systematic review, Pediatrics, 145(1), 2020, e20191579.

[6] K. Kiili, Digital game-based learning: Towards an experiential gaming model, The Internet and Higher Education, 8(1), 2005, 13-24.

[7] A. Wirth, et al., Evaluating educational apps for preschoolers: Differences and agreements between the assessments of experts, parents, and their children, Computers in Human Behavior, 160, 2024, 108361.

[8] J. Piaget, The origins of intelligence in children, International Universities Press, 1952.

[9] V. Nagalingam, R. Ibrahim, R.C.M. Yusoff, EDUGXQ: User experience instrument for educational games' evaluation, International Journal of Advanced Computer Science and Applications, 11(1), 2020, 562-569.

[10] K.N. Truong, G.R. Hayes, G.D. Abowd, Storyboarding: An empirical determination of best practices and effective guidelines, in: Proceedings of the 6th ACM Conference on Designing Interactive Systems (DIS '06), ACM, 2006, 12-21.

[11] R.A. Virzi, Refining the test phase of usability evaluation: How many subjects is enough?, Human Factors, 34(4), 1992, 457-468.

[12] D.F. Polit, C.T. Beck, S.V. Owen, Is the CVI an acceptable indicator of content validity? Appraisal and recommendations, Research in Nursing & Health, 30(4), 2007, 459-467.

[13] M.S.B. Yusoff, ABC of content validation and content validity index calculation, Education in Medicine Journal, 11(2), 2019, 49-54.