E-ISSN: 3031-8629
P-ISSN: 3032-325
Halaman 202-220
Volume 2 Nomor 1 Tahun 2025

Nusantara Board Games for Higher-Order Thinking in English Language Learning

Lisma Lia Karsa[1], Heri Hidayatullah[2], I Made Permadi Utama[3]
[1,2,3]English Language Education Department, FBMB, Universitas Pendidikan Mandalika
Jl. Pemuda No. 59 A, Mataram, Indonesia 83125
E-Mail: lismaliakarsa30@gmail.com

Abstrak

Penelitian ini mengkaji permainan papan bertema Nusantara yang dikembangkan untuk mendukung keterampilan berpikir tingkat tinggi dalam pembelajaran bahasa Inggris. Penelitian sumber memadukan pengembangan produk dan evaluasi kelas dengan melibatkan 30 peserta didik yang ditempatkan pada kelompok eksperimen dan pembanding. Produk mengintegrasikan narasi budaya Indonesia dengan tugas analisis, evaluasi, dan kreasi. Hasil utama menunjukkan rata-rata kelompok eksperimen meningkat dari 65 menjadi 82, sedangkan kelompok pembanding meningkat dari 66 menjadi 70. Penilaian ahli terhadap kelayakan produk dilaporkan sebesar 89%. Artikel ini mempertahankan ketiga bentuk bukti tersebut secara terpisah dan tidak menggabungkan validasi produk dengan hasil belajar. Pembahasan menekankan keselarasan antara keputusan dalam permainan, penggunaan bukti dari narasi, tuntutan bahasa Inggris, serta indikator berpikir tingkat tinggi. Temuan menunjukkan potensi pedagogis permainan berbasis budaya, tetapi kekuatan klaim efektivitas tetap dibatasi oleh dokumentasi desain dan asesmen yang tersedia. Pengembangan berikutnya perlu mendokumentasikan revisi produk, memperjelas karakteristik kelompok, dan menggunakan tugas analitis pada narasi baru yang dinilai secara independen.

Kata kunci: Nusantara board game; higher-order thinking; cultural narratives; English learning; educational development

Abstract

This study examined a Nusantara-themed board game developed to support higher-order thinking in English language learning. The source research combined product development and classroom evaluation involving 30 learners in experimental and comparison groups. The product integrated Indonesian cultural narratives with tasks intended to elicit analysis, evaluation, and creation. The main reported results show that the experimental-group mean increased from 65

to 82, whereas the comparison-group mean increased from 66 to 70. Expert evaluation of product feasibility was reported at 89%. The present article keeps these forms of evidence analytically separate rather than merging product validation with learning outcomes. The discussion focuses on alignment among game decisions, textual evidence, English-language demands, and higher-order thinking criteria. The findings indicate pedagogical promise for culturally grounded game tasks, but the strength of the effectiveness claim remains bounded by the available design and assessment documentation. Further development should retain a transparent revision record, clarify cohort characteristics and scoring procedures, and use independently rated analytical tasks based on unfamiliar narratives to examine transfer beyond the practiced game content.
Keywords: Nusantara board game; higher-order thinking; cultural narratives; English learning; educational development

INTRODUCTION

Higher-order thinking in English learning involves moving beyond recall toward analyzing relationships, evaluating claims, and creating defensible responses. In the revised taxonomy, these operations are cognitively distinct from remembering and understanding, so an instructional activity should be judged by what learners are required to do rather than by the attractiveness of its format. A board game can therefore support higher-order learning only when its decisions, explanations, and follow-up questions actually demand higher-level processing. This alignment between intended cognition, task design, and assessment provides the conceptual basis for the present article. (Anderson & Krathwohl, 2001; Krathwohl, 2002).

Game-based learning can provide repeated opportunities to make decisions, receive feedback, interact with peers, and persist with a task. Yet game features such as points, turns, competition, or visual appeal do not automatically produce learning. Their educational value depends on whether successful play requires the knowledge and reasoning targeted by instruction. This distinction is especially important when a study claims higher-order thinking, because engagement can coexist with tasks that remain largely reproductive. (Plass et al., 2015; Wouters et al., 2013).

Cultural familiarity offers another resource. Folklore and local narratives can lower the burden of establishing a context because learners may already know the characters, events, or social values involved. At the same time, culturally relevant instruction is strongest when familiar material becomes a resource for inquiry rather than an object to be memorized. UNDIKMA-related studies on culture-based English materials and culture-based instructional models similarly emphasize connecting local content with critical engagement, interaction, and language learning rather than treating culture as decoration. (Ladson-Billings, 1995; Prayati et al., 2020; Haerazi et al., 2021).

This relationship between language learning and higher-order thinking has also been examined in English education research at Universitas Pendidikan Mandalika. Critical-reading syllabus development has foregrounded evaluative reading needs; mobile-assisted writing has been investigated in relation to critical-thinking ability; and Value Line Learning has been used to support critical thinking in speaking and writing. Together, these studies suggest that higher-order outcomes become more interpretable when the target thinking operation is made explicit in both instruction and assessment. (Imran & Hidayatullah, 2020; Haerazi et al., 2020a; Hidayatullah & Qomariyah, 2023).

Board games are particularly relevant because they organize interaction around a sequence of turns and choices. Hidayatullah and Haerazi's work with board games in speaking instruction showed how games can create structured opportunities for language production, although their outcome was speaking rather than higher-order thinking. The present Nusantara project extends the instructional logic in a different direction by linking game decisions to culturally grounded narratives and reasoning tasks. The key issue is therefore not whether games are generally effective, but whether the Nusantara game requires learners to analyze, justify, evaluate, and create while using English. (Hidayatullah & Haerazi, 2022).

The source study combines educational product development with classroom evaluation. These activities generate different kinds of evidence. Expert judgment concerns the product's suitability, clarity, or alignment, while learner outcomes concern what participants demonstrate after instruction. Positive expert ratings can support

continued refinement but cannot substitute for evidence of student learning. Likewise, a favorable class score does not show that every design feature was equally effective. Keeping these evidential functions separate is essential for interpreting the study proportionately. (Black & Wiliam, 1998).

Accordingly, this article examines how the Nusantara board game can connect cultural narratives with analysis, evaluation, and creation; how the main reported classroom comparison should be interpreted; and what assessment evidence would be needed to support stronger claims about higher-order thinking. It retains the aggregate results documented in the source while deepening the discussion through relevant game-based, culturally responsive, critical-thinking, and UNDIKMA English-education scholarship. The contribution is a bounded account of a locally grounded instructional product rather than a universal claim that board games automatically generate higher-order thinking. (Haerazi et al., 2020b; Muliani et al., 2024).

METHOD

The source research used a mixed developmental and classroom-evaluation approach. The main methods account identifies 30 learners participating in experimental and comparison groups. The Nusantara board game incorporated Indonesian cultural narratives and English-language tasks intended to engage higher-order thinking. Because grade-level descriptions are not fully consistent across sections of the source, the present article retains the broader learner description rather than selecting one unsupported label.

The reported evidence included higher-order-thinking tasks, classroom observation, learner questionnaires, interviews, and expert validation. These sources addressed different questions: task performance concerned learning outcomes; observation and learner responses described implementation and experience; and expert review concerned product feasibility. The article retained the principal pretest-posttest group means and the reported 89% expert evaluation. It did not reconstruct

individual scores or merge supplementary summary profiles whose relationship to the principal groups was insufficiently documented.

The quantitative results were interpreted descriptively from the means reported in the source. The analysis therefore calculated only transparent differences that follow directly from those means: a 17-point gain for the experimental group, a 4-point gain for the comparison group, and a 13-point difference between the two gains. These calculations were not treated as an adjusted causal effect, because the available manuscript did not provide the full information required to verify allocation procedures, variability estimates, or a new inferential model.

The qualitative discussion focused on alignment among the game's cultural narratives, the decisions required during play, and the intended cognitive processes of analysis, evaluation, and creation. Expert evaluation was interpreted as product evidence rather than student-achievement evidence. No new participants, interviews, validation values, ethical approval identifiers, or assessment events were introduced during article preparation. Recommendations concerning revised task cards, rubrics, transfer tasks, and revision logs are presented as future design implications rather than procedures already completed in the source study.

RESULTS AND DISCUSSION

The main reported comparison shows that the experimental-group mean increased from 65 before instruction to 82 after instruction. The comparison-group mean increased from 66 to 70. These values correspond to descriptive gains of 17 and 4 points, respectively, producing a 13-point difference between the gains. The experimental group therefore shows the more favorable trajectory in the reported classroom comparison. This calculation summarizes the documented aggregate means; it is not an adjusted causal estimate or a substitute for the original statistical analysis.

Expert evaluation of the product was reported at 89%. This percentage concerns the product's assessed feasibility within the source's validation process. It is not the proportion of learners who mastered higher-order thinking and should not be combined

numerically with the pretest-posttest means. The distinction is important because development studies often produce evidence about product quality and learner performance at the same time, even though the two forms of evidence answer different questions.

Table 1. Reported evidence and interpretive boundaries in the Nusantara board-game study

Evidence source	Reported result	What it supports	Interpretive boundary
Experimental-group means	65 → 82 (gain = 17)	Favorable development on the reported learning measure	Does not identify which game component produced the change
Comparison-group means	66 → 70 (gain = 4)	Reference trajectory during the same study	Group comparability and allocation require the source design record
Difference in gains	13 points	Descriptive contrast between reported changes	Not an adjusted causal effect or standardized effect size
Expert validation	89%	Product feasibility/suitability in the validation procedure	Not a student mastery percentage or learning effect

The larger experimental-group gain provides a reason to investigate the game further, but the educational explanation depends on what learners were required to do. A task that asks for an alternative ending with a reason can engage creation and evaluation, whereas a question asking for a character's name primarily assesses recall. The game's instructional value is therefore located in its task design. The source's favorable aggregate result is consistent with the possibility that the game created useful

opportunities for higher-order engagement. It does not identify which individual game component produced the difference. (Krathwohl, 2002; Plass et al., 2015).

Cultural familiarity can support access to the task while also limiting it if familiar stories are treated as fixed answers. The more productive approach asks learners to compare motives, examine consequences, or consider another perspective. This allows the narrative to function as a shared resource for inquiry. It also avoids reducing cultural learning to recognition of symbols or memorization of a moral. The present product's next development cycle can make these distinctions explicit in its task cards and scoring criteria. (Prayati et al., 2020; Haerazi et al., 2021; Muliani et al., 2024).

The separation between feasibility and effectiveness is important for the contribution of a development study. A favorable expert rating can justify refinement and classroom testing, but it cannot substitute for evidence that learners independently performed the intended cognitive operations. Conversely, a score increase does not show that every design feature was suitable or accessible. The two forms of evidence should inform each other without being collapsed. A revision log linking expert comments to product changes would make that relationship clearer. (Black & Wiliam, 1998).

The source also raises the practical issue of evaluating reasoning in a foreign language. A learner may have a defensible interpretation but lack the language to express it fully. A rubric should therefore distinguish the depth and coherence of reasoning from surface language accuracy, while still recognizing English use as part of the task. Limited language support can make the reasoning visible without supplying the argument itself. This would strengthen the validity of future higher-order assessment and clarify the instructional value of the Nusantara game. (Haerazi et al., 2020a).

The Nusantara design links game activity with narrative interpretation and cultural content. These are related resources, but their educational value depends on the task a learner must complete during play. A move that only requires recalling a character's name does not have the same cognitive demand as evaluating the

character's decision or constructing a defensible alternative ending. The distinction is central to a higher-order-thinking claim. The article therefore interprets the game through the reasoning required by its activities rather than assuming that the presence of a board, competition, or folklore automatically produces higher-order learning. (Plass et al., 2015; Hidayatullah & Haerazi, 2022).

The reported mean changes were 17 points for the experimental group and four points for the comparison group. Their descriptive difference in change is therefore 13 points, which is not the same as the 12-point difference between the reported posttest means. These quantities answer different descriptive questions and should not be interchanged. Neither quantity alone resolves baseline comparability or other design limitations. The reported pattern is favourable to the game condition, while a causal interpretation still depends on the underlying allocation, measurement, and implementation evidence. (Black & Wiliam, 1998).

Cultural familiarity can make a narrative more accessible, but familiarity should not be confused with the thinking skill being assessed. Learners may respond more successfully because they already know the story, because the language is easier, or because the task genuinely prompts better analysis. These possibilities are not mutually exclusive. A stronger evaluation would make the role of prior familiarity explicit and include a task that requires reasoning beyond recalling a known plot. The present discussion treats cultural content as a potentially meaningful resource without assigning it an independently measured effect. (Prayati et al., 2020; Muliani et al., 2024).

The game can support analysis when learners must explain relationships within a narrative. They might consider how an event changes a character's options, whether a decision follows from the available information, or what evidence supports an interpretation. These examples clarify an analytical demand rather than adding new observed tasks to the source study. The important design principle is that a successful move should require a reasoned response that can be examined. Otherwise, the competitive structure may reward speed or chance more directly than the higher-order

thinking named in the research objective. (Hidayatullah & Qomariyah, 2023; Imran & Hidayatullah, 2020).

Evaluation requires criteria against which a decision can be judged. Asking whether a character acted well can invite a personal preference, whereas asking whether the action was justified by the evidence encourages a more explicit judgement. Learners can disagree while still producing a defensible evaluation if they state their criteria and support their reasoning. A game-based activity should make that possibility acceptable. Treating only one unexamined opinion as the correct answer would weaken the evaluative dimension even if the class remained highly engaged. (Krathwohl, 2002; Hanan et al., 2020).

Creative work also requires a distinction between novelty and quality. An alternative ending can be unusual while remaining inconsistent with the story's characters or unsupported by its events. A more useful task asks learners to create a possibility and explain how it relates to the narrative. This links creativity with coherence rather than treating any different answer as equally successful. The source's emphasis on narrative tasks makes that distinction relevant, although the article does not invent a separate creativity effect beyond the reported evidence. (Haerazi et al., 2020b).

Language proficiency can influence performance on a higher-order task administered in English. A learner may have a defensible interpretation but lack the vocabulary to express it fluently. Conversely, a fluent response can sound convincing while offering little analysis. Assessment should therefore clarify how language use and reasoning are weighted, especially when the research addresses both English learning and higher-order thinking. The article keeps those dimensions analytically separate rather than treating a single total score as proof of equivalent development in each. (Haerazi et al., 2020a; Sweller, 1988).

The expert-validation result of 89% concerns the product appraisal reported in the source. It is not a percentage of learners who improved and cannot be added to the achievement results as another effect-size estimate. Expert review can support judgements about content, design, and appropriateness, while learner assessments

address a different question. This distinction is essential in research that combines development with classroom evaluation. The two forms of evidence complement each other when their purposes remain clear, rather than being merged into an undifferentiated claim of effectiveness. (Black & Wiliam, 1998).

Collaboration during gameplay can expose learners to alternative reasoning, but it can also obscure individual understanding. A group may arrive at an answer through one participant's contribution while other members follow the move without examining it. Requiring a brief individual explanation after discussion can make learning more visible. The task should be short enough to preserve the game's flow but substantive enough to show what the learner understands. This offers a practical response to the tension between shared success and individual evidence in a culturally contextualised game. (Gillies, 2016; Murphy et al., 2009).

UNDIKMA scholarship provides several particularly relevant comparison points. Hidayatullah and Haerazi used board games to structure speaking practice, showing that a game can organize repeated language participation without implying that all game formats operate identically. Prayati and colleagues developed culture-based English materials linked to students' thinking and classroom interaction, while Haerazi and colleagues' C-BIM work examined culture-based materials in relation to critical thinking. Hidayatullah and Qomariyah further demonstrated that critical-thinking outcomes depend on a strategy that makes evaluative participation explicit. These studies strengthen the rationale for the Nusantara game while also showing why the present higher-order-thinking outcome requires its own aligned evidence. (Hidayatullah & Haerazi, 2022; Prayati et al., 2020; Haerazi et al., 2021; Hidayatullah & Qomariyah, 2023).

Engagement is a necessary consideration but not a sufficient learning criterion. Learners may enjoy winning, moving pieces, or interacting with friends without attending closely to the narrative problem. A game designer needs to connect the rewarding action with the intended intellectual operation. For example, progress can depend on an explanation rather than only a selected answer, while feedback can

identify why the explanation is defensible. The article presents this as a design implication grounded in the study's objective, not as a measured motivational mechanism. (Fredricks et al., 2004; Plass et al., 2015; Wouters et al., 2013).

The study supports further investigation of culturally contextualised games through transparent and aligned evaluation. The reported achievement pattern and product appraisal are promising, but they do not remove the need to clarify participant context, task scoring, and the relationship between English proficiency and reasoning. Future work can retain the game's cultural relevance while using unfamiliar narrative tasks and independent responses to assess transfer. That would show whether learners apply the intended thinking operations beyond the immediate game. The present account contributes a basis for that next stage rather than a universal claim about all board-game learning. (Haerazi, 2023; Muliani et al., 2024).

The reported groups began one point apart in the pretest, with the experimental mean lower than the comparison mean. At posttest the experimental mean was 12 points higher, producing the descriptive 13-point difference in change. This reconstruction clarifies the relationship among the reported means without turning the contrast into a fully identified causal estimate. A difference-in-changes description does not by itself establish the assumptions of a causal difference-in-differences design. The source pattern remains favourable, but its interpretation must stay connected to the actual allocation and assessment conditions. (Black & Wiliam, 1998).

An illustrative game prompt can ask learners to defend a character's decision using two pieces of narrative evidence. Another prompt can ask them to propose a different decision and explain how the story would change. The first foregrounds evaluation, while the second connects creation with coherence. These examples are proposed task refinements, not additional items claimed to have been administered. They show how a game's success condition can be aligned with a thinking operation rather than simply with recall or speed. (Krathwohl, 2002).

Cultural content should invite interpretation rather than require learners to accept a single unexamined moral. A familiar story can support discussion of

perspective, consequence, and the reasons a character acts as described. Learners may draw on cultural knowledge while still being asked to justify their reading of the text. This preserves the relationship between cultural relevance and critical engagement. The article does not claim that every participant experienced the content in the same way, because familiarity and identification can vary within the same classroom. (Ladson-Billings, 1995; Muliani et al., 2024).

A product-validation framework and a learner-assessment framework need different criteria. Experts can appraise whether the game's content and design plausibly support the intended task, while learners' responses show whether they actually perform the reasoning. A well-designed product can still require adjustment after classroom use. The source's combination of development and outcome reporting is therefore most useful when the two forms of evidence remain distinguishable. Future work could document which expert recommendations were supported, challenged, or refined by the learners' actual responses. (Black & Wiliam, 1998).

The Nusantara game combines cultural content with a game structure, and these components should be distinguished analytically. Familiar folklore may reduce the background knowledge needed to understand a narrative, while the game rules determine what students must actually do with that understanding. A culturally familiar story can still support only recall if the questions ask for names and events. The higher-order-thinking rationale becomes stronger when decisions in the game require students to compare explanations, justify a judgment, or construct an alternative ending. (Prayati et al., 2020; Haerazi et al., 2021).

The reported group means describe a larger improvement in the experimental group, but they do not identify which aspect of the intervention produced that pattern. Cultural familiarity, repeated exposure to narrative content, peer discussion, and the challenge structure could each contribute. The study can therefore support discussion of the combined instructional design without assigning the observed difference exclusively to cultural content or to competition. A component comparison would be needed to examine those more specific explanations. (Plass et al., 2015).

Expert validation and classroom performance also provide different kinds of evidence. Expert judgment can identify whether materials appear coherent, culturally appropriate, and aligned with the intended cognitive tasks. Learner performance concerns what participants demonstrate under the conditions of the implemented assessment. Agreement between these sources can strengthen the case for further use, but a high validation percentage is not itself a measure of students' higher-order-thinking gains. (Black & Wiliam, 1998).

An inclusive game should make reasoning visible for more than the fastest or most outspoken participant. A group can rotate the responsibility to propose an answer, challenge its justification, and connect it to the narrative evidence, followed by a short individual response. Such structures retain the social appeal of a board game while creating opportunities to inspect each learner's understanding. This is a practical implication of the design rather than a claim that the original study separately measured the effectiveness of role rotation. (Gillies, 2016; Hattie & Timperley, 2007).

The reported outcome should remain bounded by the assessment and the available research record. A score increase on a task closely related to the game's stories may represent useful proximal learning while leaving transfer to unfamiliar narratives uncertain. The interpretation is stronger when the rubric distinguishes language proficiency from the quality of analysis and justification. Without that distinction, an apparent thinking difficulty may reflect limitations in expressing an idea in English. Future assessment can therefore combine concise language support with explicit criteria for reasoning. (Imran & Hidayatullah, 2020; Hanan et al., 2020).

The regional comparison is also useful at the level of instructional mechanism. Mobile-assisted writing research at UNDIKMA connected technology use with critical-thinking ability, and genre-based work linked creativity to learners' writing performance. Haerazi's task-based flipped-learning study demonstrates another route for connecting cultural topics, metacognitive regulation, and communicative learning. The Nusantara board game adds a physical, turn-based decision environment to this local body of work. Its distinct contribution should therefore be evaluated through the reasoning required

during play and the evidence that learners can transfer those operations beyond the familiar game narratives. (Haerazi et al., 2020a; Haerazi et al., 2020b; Haerazi, 2023).

A refinement cycle can focus on the quality of explanations rather than adding more competitive features. After each challenging decision, players can state a reason and identify a detail from the narrative that supports it. A peer can then offer an alternative, and the group can compare which argument better addresses the task. This procedure gives feedback an intellectual function instead of treating it solely as confirmation of a correct answer. It also provides a visible basis for the teacher to judge whether the game is engaging the intended thinking process. (Hattie & Timperley, 2007; Plass et al., 2015).

Future studies should retain a record of product changes between validation rounds. Changes to instructions, language level, visual organization, or task difficulty can affect the interpretation of learner outcomes. Reporting only the final percentage hides the reasoning that makes a development study useful to other designers. A revision log can show which problem was identified, what was changed, and what evidence informed the decision. Such documentation strengthens the contribution of a locally grounded instructional product even before large-scale effectiveness testing becomes available. (Black & Wiliam, 1998).

Assessment alignment is the central condition for interpreting the higher-order-thinking claim. A correct answer to a game question can result from recall, chance, peer influence, or genuine analysis. A stronger task therefore requires the learner to state a claim, identify relevant narrative evidence, and explain why that evidence supports the decision. Evaluation tasks can additionally require criteria and comparison, while creation tasks can require a coherent alternative that remains accountable to the narrative. These distinctions make the cognitive process observable rather than inferred from the label attached to the card. (Anderson & Krathwohl, 2001; Krathwohl, 2002).

Language support should make reasoning visible without supplying the reasoning itself. Learners may need key vocabulary, sentence starters, or time to formulate a response in English. Such support is pedagogically appropriate when the

target includes both language and higher-order thinking, but the scoring rubric should still distinguish language control from analytical quality. Otherwise, a learner with a defensible argument may appear weak because of limited English expression, while a fluent learner may receive an inflated higher-order score for a response that is linguistically smooth but conceptually shallow. (Sweller, 1988; Haerazi et al., 2020a).

Transfer should be examined with a narrative that is not reproduced from the practiced game. A learner who can justify a decision only when the story, cards, and peer prompts are familiar has demonstrated useful proximal learning, but not yet flexible higher-order performance. A delayed task using a comparable but unfamiliar narrative would help determine whether learners can independently select evidence, evaluate alternatives, and construct a reasoned response after the immediate game context has been removed. (Black & Wiliam, 1998; Hanan et al., 2020).

The cultural dimension can likewise be extended beyond recognition. Muliani, Hidayatullah, and Qomariyah's illuminated Sasak stories illustrate the pedagogical value of local narratives for English learning, while the present board-game design adds explicit decisions and higher-order prompts. This suggests a productive design principle: cultural familiarity can provide the shared content, but the intellectual demand should come from perspective taking, justification, comparison, and consequence. The resulting task treats local culture as a resource for thought and communication rather than as a fixed set of facts to reproduce. (Muliani et al., 2024; Ladson-Billings, 1995).

Future product development should also preserve a transparent revision trail. Expert comments, learner difficulties, changes to rules, language level, visual layout, and task demands can all alter the final intervention. Recording the problem identified, the revision made, and the evidence used for the decision would strengthen the developmental contribution and make replication more meaningful. Such documentation is especially important when an expert-validation percentage is reported, because readers need to understand how validation informed the version actually used in the classroom. (Black & Wiliam, 1998; Hattie & Timperley, 2007).

Taken together, the reported pattern supports cautious optimism. The experimental group improved more than the comparison group on the source measure, the product received a favorable expert appraisal, and the cultural-game format provides plausible opportunities for reasoning and interaction. At the same time, the strongest educational claim concerns the combined instructional design rather than any single feature. More transparent scoring, learner-level trajectories, independent transfer tasks, and a fully documented product-revision history would allow subsequent research to determine whether the observed advantage reflects durable higher-order thinking in English rather than successful performance within a familiar game environment. (Plass et al., 2015; Wouters et al., 2013; Prayati et al., 2020).

CONCLUSION

The experimental-group mean increased from 65 to 82, compared with 66 to 70 in the comparison group, while expert feasibility was reported at 89%. These results support the pedagogical promise of the Nusantara board game but represent different forms of evidence. The game's higher-order contribution depends on tasks that require analysis, evaluation, and creation rather than recall alone. Future development should separate reasoning from English-language accuracy, document product revisions, report learner-level evidence, and assess transfer through independently scored tasks based on unfamiliar cultural narratives.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74. https://doi.org/10.1080/0969595980050102

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109. https://doi.org/10.3102/00346543074001059

Gillies, R. M. (2016). Cooperative learning: Review of research and practice. Australian Journal of Teacher Education, 41(3), 39–54. https://doi.org/10.14221/ajte.2016v41n3.3

Haerazi. (2023). Task-based flipped learning integrated with metacognitive skills to improve prospective English teachers' intercultural communicative competence. IJEE (Indonesian Journal of English Education), 10(2), 323–339. https://doi.org/10.15408/ijee.v10i2.32696

Haerazi, Dehghani, S., Rachmawati, U., & Irwansyah, D. (2021). The C-BIM model in improving reading, writing, and critical thinking skills: Outcome and perception. Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika, 5(2), 152–167. https://doi.org/10.36312/esaintika.v5i2.503

Haerazi, Irawan, L. A., Suadiyatno, T., & Hidayatullah, H. (2020b). Triggering preservice teachers' writing skills through genre-based instructional model viewed from creativity. International Journal of Evaluation and Research in Education, 9(1), 234–244. https://doi.org/10.11591/ijere.v9i1.20395

Haerazi, Utama, I. M. P., & Hidayatullah, H. (2020a). Mobile applications to improve English writing skills viewed from critical thinking ability for pre-service teachers. International Journal of Interactive Mobile Technologies, 14(7), 58–72. https://doi.org/10.3991/ijim.v14i07.11900

Hanan, A., Muhlisin, M., & Suadiyatno, T. (2020). Actuating reading comprehension through genre-based directed reading thinking activities (G-BDRTA) viewed from critical thinking skills. Journal of Languages and Language Teaching, 8(4), 402–411. https://doi.org/10.33394/jollt.v8i4.2778

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Hidayatullah, H., & Haerazi, H. (2022). Exploring the use of various board games to enhance speaking skills viewed from students' phonology awareness. Journal of Language and Literature Studies, 1(2), 93–102. https://doi.org/10.36312/jolls.v1i2.614

Hidayatullah, H., & Qomariyah, S. S. (2023). Toward EFL students' critical thinking: How effective Value Line Learning strategy is? Jo-ELT (Journal of English Language Teaching), 10(1), 114–119. https://doi.org/10.33394/jo-elt.v10i1.7854

Imran, F., & Hidayatullah, H. (2020). EFL critical reading syllabus and materials for students of the English Department. Jo-ELT (Journal of English Language Teaching), 7(1), 51–56. https://doi.org/10.33394/jo-elt.v7i1.2747

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. Theory Into Practice, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. American Educational Research Journal, 32(3), 465–491. https://doi.org/10.3102/00028312032003465

Muliani, M., Hidayatullah, H., & Qomariyah, S. S. (2024). Illuminated Sasak stories: A bridge to English language learning. Jo-ELT (Journal of English Language Teaching), 11(2), 214–223. https://doi.org/10.33394/jo-elt.v11i2.13092

Murphy, P. K., Wilkinson, I. A. G., Soter, A. O., Hennessey, M. N., & Alexander, J. F. (2009). Examining the effects of classroom discussion on students' comprehension of text: A meta-analysis. Journal of Educational Psychology, 101(3), 740–764. https://doi.org/10.1037/a0015576

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. Educational Psychologist, 50(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Prayati, Z., Haerazi, H., Irawan, L. A., & Vikasari, R. M. (2020). Developing culture-based English instructional materials to improve students' thinking, ELT achievement, and classroom interaction. Lingua Cultura, 14(2), 211–218. https://doi.org/10.21512/lc.v14i2.6650

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology, 105(2), 249–265. https://doi.org/10.1037/a0031311