ETDC
Educational Technology in Developing Countries

Vol. 1, No.1 (June 2026), 1-22
P-ISSN: 3163-3927 | E-ISSN: 3163-401X
DOI: https://doi.org/10.64929/etdc.v1i1.60

Enhancing Conceptual Understanding in Islamic Education Through a Hypercontent Digital Flipbook in Blended Learning: Evidence from Indonesian Madrasas

OPEN ACCESS

Miftah Hur Rahman Zh[1]*, Cynthia Mayang Sari[2], Zahid Zufar At Thaariq[3], Ilham Sumalee[4], Athia Nur Kamilah[5]

[1] Darunnajah University Jakarta, Indonesia
[2] Universitas Negeri Jakarta, Indonesia
[3] Cukurova University, Adana, Turkey
[4] Universitas Muhammadiyah Ponorogo, Indonesia
[5] Universiti Utara Malaysia, Malaysia

*Correspondence: miftahurrahmanzh@darunnajah.ac.id

Received: 19-03-2026	Accepted: 10-05-2026	Published: 16-05-2026

Abstract: Conceptual understanding in Qur'anic studies requires learners to integrate Arabic text, recitation, exegesis, historical context, and contemporary application, an inherently multilayered and nonlinearly structured knowledge domain that linear print materials struggle to support. This study developed and evaluated a hypercontent digital flipbook within a blended learning structure to enhance conceptual understanding in Qur'anic studies at an Indonesian madrasa. Recognizing that linear print materials inadequately reflect the multilayered, nonlinear nature of religious knowledge, we designed a web-based flipbook that integrates Arabic text, translation, concise *tafsir*, audio recitation, animated explanations, and gamified assessment. Following the ADDIE model, we validated the artifact across four expert domains (instrument, learning strategy, media, content), conducted a small-group usability trial (n=20), and implemented a single-group pretest-posttest design with 36 Grade XI students from a state Islamic senior high school in East Java. Expert validation yielded ratings between 78% and 98.5%, while student acceptance ranged from 90.56% to 91% across trials. Pretest mean scores of 68.33 (61.11% mastery) increased to 80.28 (91.67% mastery) at posttest, with paired-samples $t(35) = 22.23$, $p < .001$, mean increase = 11.94 points, and Cohen's dz = 3.70. Hake's normalized gain of 0.38 indicated moderate educational improvement. However, the pre-experimental design, identical pretest-posttest items, and single-site sample preclude causal attribution. We articulate a candidate mechanism whereby nonlinear navigation supports conceptual integration across content layers. Future research requires quasi-experimental dismantling designs, multi-site replication, and delayed retention measures.

Keywords: blended learning; conceptual understanding; digital flipbook hypercontent; Islamic education; multimedia learning.

Copyright © 2026 by Author(s)
This work is licensed under a **Creative Commons Attribution-ShareAlike 4.0 International License**.

INTRODUCTION

Indonesian secondary schooling sits in a paradox that frames the present study. The country enrolls more than 9.7 million students in formal Islamic schools known as *madrasah*, including approximately 1.4 million students in *Madrasah Aliyah* at the senior secondary level (Indonesian Ministry of Religious Affairs, 2024), making the system one of the largest faith-based education networks in the world. However, in the most recent Program for International Student Assessment (PISA 2022), Indonesian 15-year-olds scored 359 in reading, 366 in mathematics, and 383 in science, all substantially below the OECD averages of 476, 472, and 485, respectively, and all marking declines from 2018 (OECD, 2023). More striking still, more than 60 percent of Indonesian students performed at PISA Level 1 or below across all three domains, indicating capacity for surface recall but limited ability to integrate concepts, interpret complex material, or apply knowledge in new situations. The pattern recurs in classroom-level indicators. In the present study setting, only 22 of 36 Grade XI students (61.11 percent) reached the Minimum Competency Achievement threshold of 70 on a routine pretest of Al-Qur'an and Hadith content. These figures identify a problem of practice that is empirical rather than rhetorical: students complete the curriculum but do not consistently integrate what they read, a pattern that has acquired new urgency as Indonesian education enters its current digital-transformation phase (Rahmatullah et al., 2022).

Two strands of intervention research bear directly on this problem. First, blended learning, the structured combination of face-to-face and technology-mediated instruction (Garrison & Vaughan, 2008; Graham, 2006), has accumulated robust meta-analytic support over the past decade. Anthony et al. (2022) reported consistent positive effects on student outcomes when blended designs are implemented with pedagogical coherence. In a larger review, Topping et al. (2022) found that blended learning outperformed regular instruction in 85 percent of comparisons. Müller and Mildenberger (2021) further confirmed that replacing classroom time with well-designed online environments produces neutral-to-positive learning effects in higher education, building on earlier meta-analytic syntheses (Means et al., 2013). However, these reviews share a striking blind spot for the present study, since almost none of the synthesized studies were conducted in faith-based educational settings, and the Indonesian *madrasah* context appears in fewer than 5 percent of the empirical work that grounds these claims (Sihombing & Fatra, 2021; Bahri et al., 2021; Dwianto et al., 2017; Suana et al., 2017; Wahyuni et al., 2019). The applicability of meta-analytic effect sizes to Indonesian Islamic education, therefore, rests on an assumption of transferability that has not been empirically tested. That gap matters when 1.4 million *Madrasah Aliyah* students are the intended beneficiaries of any policy that follows from this evidence base.

Second, the broader hypertext and hypermedia literature, originating with Nelson (1965) and Conklin (1987) and consolidated through Landow (1992), DeStefano and LeFevre (2007), and Salmerón et al. (2018), has accumulated four decades of evidence on how nonlinear digital architectures affect comprehension. The synthesis is more nuanced than enthusiasm for digital media might suggest. For instance, DeStefano and LeFevre (2007) showed that nonlinear navigation can either support or impair comprehension depending on cognitive load, and Scheiter and Gerjets (2007) documented that learner control in hypermedia produces benefits only when learners possess the metacognitive regulation to use it effectively. Indonesian flipbook studies, by contrast, have tended to treat hypercontent as a primarily local construct describing media that combine text, images, audio, and video, without engaging the international hypertext-comprehension literature in any depth (Hasrah et al, 2025; Widyaningrum & Ardiansyah, 2026; Simbolon et al., 2021; Nababan, 2023). The

result is a body of empirical work that documents flipbooks' acceptability and engagement effects but rarely engages the harder cognitive question that the international literature has been wrestling with for 40 years: under what conditions does nonlinear architecture support genuine conceptual integration rather than surface processing?

Three understanding gaps, therefore, motivate the present study, framed not as the output of a systematic review but as a qualitative reading of the available literature in this area. The first is a population-coverage gap. Indonesian flipbook research published in indexed journals in recent years has predominantly addressed STEM content in general secondary schools rather than *madrasah* education, and very few studies have specifically targeted Qur'anic-content learning at the *Madrasa Aliyah* level (Hasrah et al, 2025; Zh et al., 2025; Simbolon et al., 2021; Dermentzi et al., 2022; Abdullah et al., 2021). The cognitive demands and the cultural-religious constraints of the *madrasah* Qur'anic-studies context differ substantially from those of the general secondary STEM context, and findings from one are not straightforwardly transferable to the other. The second is an outcome-construct gap. The Indonesian flipbook literature has tended to prioritize engagement, satisfaction, and feasibility outcomes over conceptual or higher-order learning gains. Where pretest-posttest designs have been used, effect-size reporting has rarely extended beyond Cohen's *d* for independent groups, leaving the strength of any association difficult to interpret in international meta-analytic terms (Zh et al., 2025; Dermentzi et al., 2022). The third is a theoretical-grounding gap. Although hypercontent and digital flipbook studies in Indonesia draw on local conceptualisations and on Mayer's multimedia learning principles, they have rarely engaged with the foundational hypertext-comprehension literature (Conklin, 1987; DeStefano & LeFevre, 2007; Landow, 1992; Nelson, 1965; Salmerón et al., 2018), even though this body of work provides the most directly relevant theoretical resources for understanding why nonlinear multimedia architecture might or might not support conceptual integration. Each gap is conceptually specifiable, addressable in a single study, and consequential for the field's ability to make informed claims about what hypercontent flipbooks can do.

These three gaps matter substantively for both research and practice. For research, the limited engagement with international theoretical resources and the predominance of engagement-and-feasibility outcomes mean that the Indonesian flipbook literature has not yet generated the kind of evidence that international meta-analyses can readily absorb, leaving *madrasah* educational research relatively peripheral to global conversations about blended and digital learning. For practice, the absence of theoretically grounded design knowledge means that practitioners and policymakers advocating for flipbook adoption often rely on engagement and acceptability data to support claims about learning gains, an inferential leap that the available evidence does not strongly support. The cost of this gap is not abstract. As Indonesia's Ministry of Religious Affairs continues to invest in digital instructional materials for the *madrasah* system, decisions are being made without the kind of rigorous, construct-aligned evidence that the international literature would expect for comparable-scale investments in secular contexts. Closing the gap, therefore, serves both scholarly and practical interests, and the present study is one step in that direction rather than a definitive resolution.

The present study addresses these gaps through four research questions, each formulated to generate evidence that the field currently lacks. RQ1 (validity) asks how experts in instrument design, learning strategy, media technology, and Qur'anic content evaluate the validity of a hypercontent digital flipbook designed to support conceptual understanding in *Madrasah Aliyah*. RQ2 (acceptance) asks how students and teachers evaluate the usability and

pedagogical fit of the artifact during small- and large-group implementation trials. RQ3 (association) asks whether pretest-posttest changes in students' conceptual understanding of Al-Qur'an and Hadith content are of meaningful magnitude under the intervention, framed as association rather than causal effect, given the design constraints. RQ4 (mechanism, exploratory) asks what theoretically grounded mechanism, drawing on hypertext-comprehension research, plausibly accounts for any observed associations and what design conditions would be required to test that mechanism rigorously in subsequent research. The study's unique contribution is to operationalize hypercontent through international hypertext-comprehension theory rather than as a local construct, to target conceptual understanding (rather than engagement) as the primary outcome, and to report effect sizes (Cohen's *d*, Hedges' g_av, Hake's normalized gain) and a non-parametric sensitivity check that the existing Indonesian flipbook literature has not provided. Across the literature located for this study in Scopus and Google Scholar, we did not identify a prior empirical study at the intersection of *madrasah* Qur'anic studies and hypercontent design that brings these three elements together, though we make no claim to have searched exhaustively.

METHOD

Research Design and Procedural Overview

The study followed the ADDIE instructional development model (Branch, 2009), with the Evaluation phase implemented as a single-group pretest-posttest design. ADDIE was selected because it offers a tightly specified workflow appropriate for a master 's-thesis-derived study with a fixed timeline. We acknowledge that design-based research traditions offer stronger theoretical commitments to iterative refinement and are preferable for studies aiming to generate transferable design principles. The chosen evaluation design is explicitly pre-experimental, with no control group, no random assignment, and no blinding. Causal inferences are therefore not warranted, and the analysis treats the pretest-posttest comparison as an associational, feasibility-oriented signal. Table 1 summarises the ADDIE phases implemented in the present study, specifying deliverables, duration, and methods for each phase to support transparency and replicability.

Table 1. ADDIE Phases as Implemented in the Present Study

Phase	Activities	Deliverable	Duration	Method
Analysis	Curriculum review; teacher interview; learner needs analysis	Needs profile and content blueprint	2 weeks	Document analysis; semi-structured interview (*n* = 2 teachers)
Design	Storyboarding, learning objectives mapping, and selection of media affordances	Design specification document	3 weeks	Mayer multimedia principles applied; nonlinear architecture mapped to content layers
Development	Flipbook construction; Wordwall game design; audio recording; testing	Functional prototype	6 weeks	Heyzine for shell; Wordwall for assessment; iterative debug cycles
Implementation	Small-group trial (*n* = 20); large-group trial (*n* = 36); blended sequence	Use logs and acceptance data	4 weeks	Asynchronous-then-synchronous syntax; teacher-facilitated synchronous sessions
Evaluation	Expert validation (4 domains); pretest and posttest; acceptance survey	Validation results, paired-samples scores, and qualitative feedback	2 weeks	Likert validation; identical-item pretest and posttest; user-experience survey

Participants and Setting

The study was conducted at Madrasah Aliyah Negeri (MAN) Kota Batu, a state Islamic senior high school operating under the Indonesian Ministry of Religious Affairs in East Java. MAN Kota Batu is a state (not private) *madrasah*, urban in location, with mid-resource infrastructure, including stable classroom internet and a teacher cohort with prior digital-tool exposure through provincial professional development. Findings should be interpreted within this profile and cannot be assumed to transfer to private, rural, or lower-resource *madrasah* without further empirical work. Twenty students participated in the small-group usability trial, selected purposively to represent a range of prior achievement (high, mid, and low tertiles) and to balance gender within the available cohort. The full implementation involved 36 students from one intact class. Participation was voluntary; informed consent was obtained from students and their parents or guardians, and no students withdrew. The school's Minimum Competency Achievement (*Kriteria Ketuntasan Minimal*, KKM) threshold for Al-Qur'an and Hadith was 70 points. No a priori power analysis was conducted. The analysis is therefore framed as descriptive and feasibility-oriented, and a precise estimation of population effect sizes is not claimed.

Product Specification: Hypercontent Digital Flipbook

The artifact was a web-accessible digital flipbook that embedded multiple content layers within a nonlinearly navigable interface. The technical realization used Heyzine for the flipbook shell and Wordwall for embedded gamified assessment. Each design choice was tied to a specific theoretical commitment drawn from multimedia learning, cognitive load, and

hypertext comprehension theory. Table 2 maps each major design feature to the theoretical principle that motivated it and to the anticipated cognitive function, thereby supporting the reader's interpretation of the artifact as a theory-driven design rather than an arbitrary collection of media.

Table 2. Hypercontent Flipbook Design Features Mapped to Theoretical Principles

Design Feature	Theoretical Principle	Reference	Anticipated Cognitive Function
Arabic text with Indonesian translation displayed contiguously	Modality and contiguity principles	Mayer (2024)	Reduce split-attention; support verbal-visual integration
Audio recitation accompanying written text	Dual-channel processing	Mayer (2024)	Distribute load across auditory and visual channels
Concise *tafsir* summaries	Germane cognitive load	Sweller et al. (2019)	Support schema construction; reduce extraneous load
Hyperlinked nonlinear navigation	Hypertext architecture	DeStefano and LeFevre (2007); Landow (1992)	Enable learner-controlled integration across content layers
Game-based formative assessment via Wordwall	Retrieval practice and testing effect	Castro-Alonso et al. (2021)	Strengthen retrieval pathways through low-stakes recall
Cross-device responsive design	Resource-aware design for developing-country contexts	Rasheed et al. (2020)	Support equitable access across heterogeneous student devices

Blended Learning Syntax

The blended learning syntax operationalized two distinct phases with distinct pedagogical functions. In the asynchronous phase, students accessed the flipbook independently on any internet-connected device, working through self-paced reading, video viewing, audio listening, note-taking, and an initial diagnostic quiz. In the synchronous phase, the classroom teacher facilitated whole-class discussion, mind-mapping of conceptual relationships, group presentation, and reflection on application to daily life. The synchronous phase was deliberately designed not as content transmission but as the integration moment in which teacher and peers helped consolidate the relational knowledge students had begun to construct asynchronously. Figure 1 visualizes the syntax of the teacher-student roles in each phase. The full intervention spanned four weeks, comprising four synchronous 90-minute sessions, with asynchronous engagement scheduled between sessions.

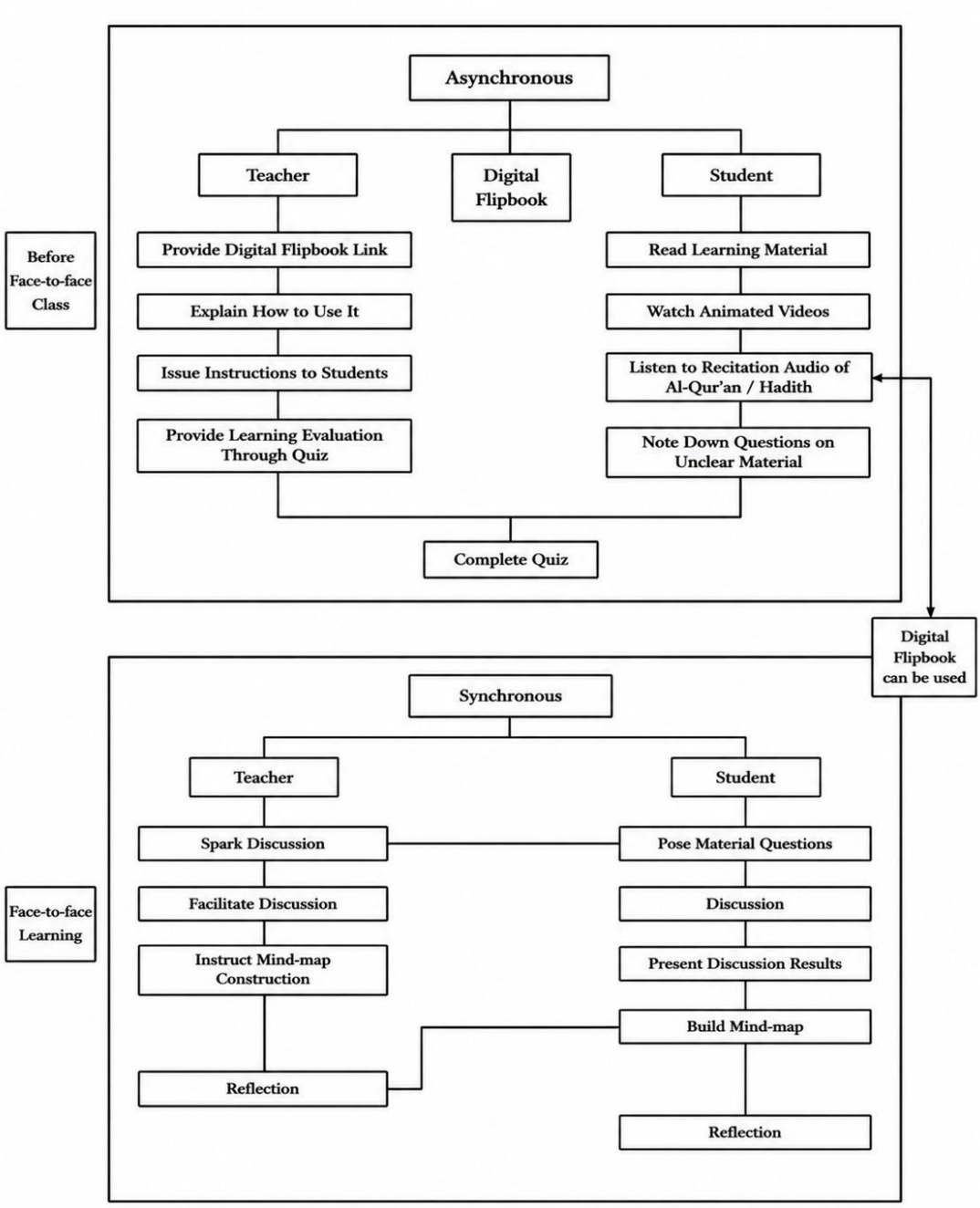

Figure 1. Blended learning syntax with teacher and student roles in asynchronous and synchronous phases.

Instruments and Validation Framework

Four expert validation instruments were administered to evaluate distinct quality dimensions of the developed artifact. Each instrument used a five-point Likert scale (1 = not at all valid; 5 = highly valid), along with open-ended fields for qualitative feedback. Single-rater validation is a substantial methodological limitation that we acknowledge openly. International validation standards require multi-expert panels (typically three to five raters per dimension) to report Aiken's V or the Content Validity Index, together with interrater reliability indices. The validity classifications reported in the Results section should therefore be read as expert ratings from individual reviewers rather than as psychometrically robust

validity indices. The four domains have differing epistemic status and are not aggregated, as detailed in Table 3.

Table 3. Validation Framework Across Four Quality Dimensions

Domain	Items	Expert Profile	Epistemic Status
Instrument validation	10 (clarity, content adequacy, relevance, language)	Methodological expert (*n* = 1)	Measurement-quality judgment
Learning strategy validation	10 (blended-syntax alignment with instructional objectives)	Learning-design expert (*n* = 1)	Instructional-design fit
Media validation	13 (interface, multimedia functionality, cross-device compatibility, interactivity)	Media-technology expert (*n* = 1)	Technical usability judgment
Content validation	18 (content adequacy, presentation, language, pedagogical alignment)	Al-Qur'an and Hadith subject-matter expert (*n* = 1)	Disciplinary-content judgment

Conceptual Understanding Assessment

Conceptual understanding was operationalized through a 20-item multiple-choice test targeting Bloom's understanding and application levels in the *halal-thayyib*, *syukur*, and ethical application content areas. Items were developed against a content blueprint covering five subtopics, with item distribution proportional to curriculum weighting. The pretest and posttest used identical items administered four weeks apart. Multiple-choice items are widely critiqued in the assessment literature as suboptimal measures of conceptual understanding because they reward recognition over construction (Arif & Abd Aziz, 2023). The use of identical items further introduces a substantial testing effect risk, in that students may improve in part because they remember items rather than because they have integrated the content more deeply. Both limitations are treated as substantial rather than incidental, and the test's reliability (KR-20 or Cronbach's alpha) was not formally estimated because raw item-level data were unavailable. This is a reporting limitation that should be addressed in any future replication.

Data Analysis

Validity percentages were computed as P = (sum of R divided by N) multiplied by 100 percent, with the conventional Indonesian instructional-development classification: A (76 to 100 percent, highly valid or effective), B (51 to 75 percent), C (26 to 50 percent), and D (0 to 25 percent). We retain this rubric for transparency with the source thesis, but note that international standards typically use Aiken's V (≥ 0.7) or the Content Validity Index (≥ 0.78). For the pretest-posttest comparison, normality was assessed using the Shapiro-Wilk test, which is more sensitive than the Kolmogorov-Smirnov test for samples with $n < 50$. The paired-samples t-test served as the primary analysis, complemented by the Wilcoxon signed-rank test as a sensitivity check on the same paired data.

We report Cohen's d_z as the primary standardized effect-size measure for paired data, accompanied by Hedges' g_{av} (small-sample-corrected) and Hake's normalized gain ⟨g⟩ for educational interpretation. A bootstrap 95 percent confidence interval (10,000 resamples) is reported for d_z. We also examined heterogeneity in within-person change (the range and

distribution of individual gains) following recent guidance on heterogeneous treatment effects in educational research (Wang & Hofkens, 2020).

Research Ethics

The research protocol received departmental approval from the Faculty of Education Sciences, Universitas Negeri Malang. We acknowledge that formal Research Ethics Committee approval was not obtained at the time of fieldwork, and we recommend explicit Research Ethics Committee approval for future studies aiming at international publication. Informed consent was obtained from all participating students and from their parents or guardians prior to data collection. Participation was voluntary throughout, and students were informed they could withdraw at any time without academic consequence. All student data are reported using pseudonyms (Student 1 through Student 36) to protect confidentiality. Data are stored on password-protected institutional servers and can be made available to the corresponding author upon reasonable request, subject to participant confidentiality.

RESULTS

This section reports findings from the four expert validations, the small-group and large-group implementation trials, and the pretest-posttest assessment of conceptual understanding. Following the decision documented in the Method section, the four validation domains are reported separately rather than aggregated into a single mean. The pretest-posttest analysis is reported with multiple effect-size estimates and a non-parametric sensitivity check, in line with contemporary reporting standards for educational intervention research.

Table 4. Expert Validation Results Across Four Domains

Validation Domain	Score	Max	Validity (%)	Category
Instrument (*n* = 1)	39	50	78%	A: Valid / Feasible
Learning Strategy (*n* = 1)	38	40	95%	A: Highly Effective
Media / Technology (*n* = 1)	64	65	98.5%	A: Highly Effective
Content / Subject Matter (*n* = 1)	73	90	81.11%	A: Highly Valid

All four domains received ratings in the upper band of the rubric, with media-technology highest at 98.5 percent and content lowest at 81.11 percent. The content expert recommended strengthened language accessibility for students with lower Indonesian-language reading proficiency and additional contextualization of historical references, and these recommendations were incorporated into the version used for the implementation trials. Because each rating reflects one rater, these results are best read as encouraging single-expert signals warranting further development rather than as established validity indices in the contemporary psychometric sense. Replication with multi-rater panels reporting Aiken's V or the Content Validity Index would be required to support stronger validity claims for the instruments and the artifact.

Table 5. Student and Teacher Acceptance During Implementation Trials

Trial Phase	*n*	Acceptance (%)	Category	Key Feedback
Small Group Trial	20	91%	Excellent	Engaging interface, intuitive navigation
Large Group Trial	36	90.56%	Excellent	Rich content, occasional loading delay
Teacher Response	2	Qualitative	—	Supports differentiated pacing

Both student trials produced ratings in the upper band of the rubric, with qualitative feedback highlighting the engaging interface and intuitive navigation paths between content layers. Qualitative feedback in the large-group trial flagged occasional loading delays on slower internet connections, an infrastructural condition that becomes relevant to the scalability discussion below. Teacher feedback emphasized that the asynchronous-then-synchronous structure supported differentiated pacing, allowing students who needed more time with foundational material to take it without holding back peers who could move more quickly. The acceptance instrument captured user satisfaction and perceived usability rather than theoretically grounded technology acceptance, and the results are reported descriptively rather than as tests of any acceptance theory.

Table 6. Pretest and Posttest Descriptive Results (*N* = 36)

Measure	Pretest	Posttest	Difference	Significance
Mean Score (*SD*)	68.33 (9.06)	80.28 (7.61)	+11.94 (3.22)	see Table 7
Median	70.00	80.00	—	
Q1, Q3 (IQR)	60, 75 (15)	75, 85 (10)	—	
Min, Max	50, 80	65, 90	—	
Mastery Rate (KKM ≥ 70)	61.11%	91.67%	+30.56 pp	

Table 7. Inferential Tests and Effect Sizes for Posttest Versus Pretest

Statistic	Value	95% CI / Interpretation
Mean Increase (raw points)	11.94	[10.85, 13.04]
t (*df* = 35)	22.23	$p < .001$ (two-tailed)
Wilcoxon W (sensitivity)	0	$p < .001$
Cohen's dz (paired)	3.70	[3.07, 4.85] (bootstrap, 10,000 resamples)
Hedges' g_av (corrected)	1.40	Very large by Cohen's benchmarks
Hake's normalized gain ⟨*g*⟩	0.38	Medium gain (Hake's benchmarks)
Pre-post correlation (*r*)	.94	Very high; relevant to testing-effect concern

Three patterns warrant explicit attention. First, parametric and non-parametric tests converge unambiguously. Both the paired-samples t-test and the Wilcoxon signed-rank test indicate that posttest scores were systematically higher than pretest scores, and the Wilcoxon W of 0 means that every single student improved between the two test administrations. Second, effect-size estimates differ substantially in magnitude across metrics. Cohen's dz of 3.70 is very large, Hedges' g_av of 1.40 is also large, but Hake's normalized gain of 0.38 lies in the medium range. The divergence reflects that dz is sensitive to the very small standard deviation of difference scores (3.22), itself driven by the high pre-post correlation, while normalized gain accounts for the ceiling room remaining at pretest. Third, the high pre-post correlation of .94, combined with an identical-item design, raises a substantive concern that

part of the observed change reflects testing effects rather than genuine conceptual restructuring. We therefore present these results as a within-subjects association of meaningful magnitude under a pre-experimental design rather than as causal evidence of intervention effectiveness.

DISCUSSION

Interpreting the Pretest-Posttest Association in Light of Design Constraints

The four research questions yield distinct evidentiary strands that, taken together, support a feasibility claim rather than an effectiveness claim. For RQ1, expert validation across all four domains fell within the upper band of the Indonesian instructional-development rubric, with the highest rating for media-technology quality and the lowest for content. These ratings should not be over-aggregated into a single mean validity score. However, they collectively support feasibility for further development with the language-accessibility refinements recommended by the content expert. For RQ2, both student trials produced acceptance ratings in the upper band, with qualitative feedback indicating engagement and a perceived match between the asynchronous-then-synchronous pacing model and the diverse pacing needs of an intact class. For RQ3, all 36 students improved between pretest and posttest, with a mean increase of 11.94 points and effect-size estimates of dz = 3.70, g_av = 1.40, and Hake's normalized gain of 0.38.

We deliberately avoid the language of triangulation when synthesizing these results. Expert validation, user acceptance, and pretest-posttest change measure substantively distinct constructs and do not converge on a single epistemic question that triangulation would help address. Instead, they constitute a multiple-outcome evaluation rather than methodological triangulation. The honest synthesis is that the artifact appears feasible to develop and acceptable to its intended users. That within-person change in test scores was systematically positive, conditional on the design's well-known threats to causal inference. Several alternative explanations for the observed change deserve explicit acknowledgment rather than dismissal in passing. Testing effects, supported by the very high pre-post correlation under identical items, are a particularly strong candidate for part of the observed gain because students may improve through familiarity with item content and structure rather than through deeper understanding.

Novelty effects from introducing any salient new educational medium are well documented in educational technology research and often produce short-term gains that fade as novelty wears off (Topping et al., 2022). Hawthorne effects, in which students perform better when they perceive themselves as participating in a research program, cannot be ruled out under the present design. Teacher-attention effects are also plausible because the same teachers facilitated the synchronous phase, and any increased attention or preparation effort during the research period would, in itself, contribute to measured gains independently of the artifact. The pre-experimental design cannot disentangle these candidate explanations from the contribution of the hypercontent flipbook itself, and any honest interpretation of the results must hold these alternatives open rather than dismiss them.

The divergence between effect-size estimates is itself informative and worth examining. The very large Cohen's dz of 3.70 is partly an artifact of an unusually small standard deviation of difference scores, which in turn reflects the very high pre-post correlation. Hake's normalized gain of 0.38, in the medium band, by Hake's own benchmarks for educational interventions, may better represent the educational interpretation of these results, in the sense that students made meaningful progress. However, ceiling room remained at posttest,

and the gain is not extraordinary by the standards of well-designed multimedia interventions reported in international meta-analytic work (Anthony et al., 2022; Topping et al., 2022). Considering the parametric, non-parametric, and educationally interpretable effect-size measures together yields a coherent overall picture. The intervention is associated with positive change, the change is statistically detectable under multiple analytical lenses, but the magnitude is moderate rather than transformative, and a substantial portion of it may reflect testing and novelty effects that the design cannot isolate. We invite readers to keep these caveats in mind throughout the remainder of this discussion.

Situating the Study Within International Hypertext-Comprehension Research

The novelty of the present study lies in three connected analytical moves rather than in any single first claim. First, the study integrates hypercontent design, understood within its proper international lineage from Nelson (1965) through Conklin (1987), Landow (1992), DeStefano and LeFevre (2007), and Salmerón et al. (2018), with a blended learning framework that has been theorized in contemporary educational technology research (Anthony et al., 2022; Halverson & Graham, 2019). This integration is rare in the available literature. Most Indonesian flipbook studies have approached hypercontent as a primarily local construct without engaging the international hypertext-comprehension literature in any depth. In contrast, most international hypertext research has been conducted in secular Western contexts and has not engaged with the specific cognitive demands of religious-text learning. Bringing these two literatures into conversation through a single empirical study contributes to both, importing international theoretical resources into the Indonesian flipbook literature and extending the conversation on hypertext comprehension into a content domain it has not previously addressed.

Second, the study foregrounds conceptual understanding as the target outcome and offers an explicit theoretical operationalization of relational knowledge that supports schema construction (Sweller et al., 2019). This is a substantively more demanding outcome than the engagement and feasibility outcomes that have dominated the Indonesian flipbook literature for the past several years (Zh et al., 2025; Dermentzi et al., 2022) and one that aligns with international scholarly conversations about what learning technologies should ultimately deliver, including game-based and retrieval-oriented approaches that target conceptual gain (Castro-Alonso et al., 2021; Widiyatmoko et al., 2022). Operationalizing conceptual understanding theoretically rather than rhetorically forces the Method and Results sections to address harder questions about whether instrumentation actually measures what is claimed, and the present study's frank acknowledgment of multiple-choice limitations and the risk of identical-item testing effects provides methodological transparency that the field would benefit from more broadly.

Third, the study introduces a mechanistic question (RQ4) into a literature that has often stopped at descriptive validation. We articulate, rather than empirically test, a candidate mechanism by which nonlinear navigation might support conceptual integration in multilayered religious knowledge domains, and we identify the controlled-design conditions under which the mechanism could be tested rigorously in subsequent research. The proposed mechanism is summarised graphically in Figure 2 and verbally below: nonlinear navigation between text, translation, exegesis, audio, and application examples enables learner-controlled integration across content layers in ways that linear print or single-layer digital materials do not. We were unable to locate a prior study combining hypercontent design grounded in international hypertext theory, blended learning grounded in contemporary

blended learning theory, conceptual understanding as a theoretically operationalized outcome, and articulated mechanistic question in a single empirical study at the intersection of Indonesian *madrasah* Qur'anic studies, hypercontent design, and blended learning theory. We present this as the outcome of a bounded rather than exhaustive search of the available literature.

Theoretical, Methodological, and Practical Considerations

We separate theoretical, methodological, and practical implications because each rests on different evidentiary support. Theoretically, the present study applies rather than extends the cognitive theory of multimedia learning (Mayer, 2024) to a faith-based content domain that has been largely beyond the empirical reach of the international multimedia-learning literature. We did not test specific multimedia design principles such as modality, redundancy, or segmenting under controlled conditions. Therefore, the appropriate theoretical claim is that multimedia design principles can be implemented coherently in religious-content learning, not that the theory has been extended into new territory. A genuine extension would require designs that systematically vary specific principles and measure their effects in this content domain. The study also contributes to ongoing conversations about hypertext-supported comprehension by articulating a candidate mechanism in a non-Western religious-content context. However, it does not establish new boundary conditions for hypertext-comprehension theory in any rigorous sense.

Methodologically, the study illustrates both the feasibility and limitations of single-group pretest-posttest designs in instructional development research intended for international publication. The reported associations are substantial in magnitude and warrant further investigation through more rigorous designs. However, the present design cannot disentangle the contribution of the hypercontent artifact from that of the blended structure itself, the asynchronous-then-synchronous sequencing that has well-documented effects on learning (Halverson & Graham, 2019). Future work should employ at a minimum a quasi-experimental design with a comparison group receiving the blended structure without hypercontent features, isolating the artifact's contribution from the structural contribution. Ideally, a dismantling design that systematically varies specific media affordances within the hypercontent design would allow the field to identify which features carry the most pedagogical weight. The methodological lesson here is an honest one. Feasibility-and-association studies remain valuable as developmental scaffolding, but they cannot answer the questions the field most needs answered.

In practice, the validated artifact and the documented blended-syntax model provide a starting point for *madrasah* teachers who wish to explore digital media in Qur'anic studies but lack the instructional design resources to develop their own materials. The flipbook's cross-device compatibility, integration of audio recitation alongside written text, inclusion of concise *tafsir* summaries and historical-context boxes, and embedded gamified retrieval practice address real classroom constraints that teachers in this context have voiced, drawing on broader conversations about smart and personalized learning environments (Cheung et al., 2021; Shemshack & Spector, 2020). We are deliberately cautious, however, about scaling claims beyond what the evidence supports. A single demonstrably acceptable pilot at one state *madrasah* does not warrant system-level recommendations across Indonesia's much larger *madrasah* network. The educational technology literature is replete with cautionary tales of pilots that did not scale beyond their initial sites for reasons that became apparent only at scale, including persistent challenges in the online component of blended designs

(Rasheed et al., 2020; Topping et al., 2022). Any scaling would depend on conditions the present study did not establish, including stable internet infrastructure across receiving sites, sufficient teacher capacity for blended facilitation rather than content transmission, and content review by religious authorities competent to assess fidelity to the discipline.

The intervention also carries trade-offs that warrant explicit acknowledgment rather than celebration. Increased screen time has been shown to affect student well-being and attention, effects that may not be fully captured by outcome measures focused on test performance. The digital divide between students with reliable home internet access and those without could exacerbate existing achievement gaps if the asynchronous component becomes an inequality multiplier rather than a leveler. Questions about the role of religious authorities in content validation deserve serious attention rather than dismissal. For example, who has the authority to determine that a *tafsir* summary is faithful to the source tradition, and how should that authority be incorporated into design and review processes for digital materials in Islamic education? These practical implications point toward genuinely difficult questions that the present study can identify but cannot resolve, and they should be central to any expansion of the work into other sites or audiences.

A Multi-Strand Framework for Faith-Based Educational Technology

Figure 2 synthesizes the contribution of the present study to educational technology research in resource-constrained and faith-based educational settings. The figure presents three connected strands, labeled cognitive, pedagogical, and contextual, each carrying part of the contribution and each constraining what the study can legitimately claim from a single-site, pre-experimental investigation. The framework is offered as a synthesis tool and roadmap for future investigation rather than as a tested model, and we explicitly state that empirical support for the proposed mechanism at the center of the figure remains a matter for further research. The cognitive strand contributes to the application of multimedia learning principles in nonlinear, multilayered domains of religious content where Arabic script, recitation, exegesis, and contextual application coexist. This is a content domain that has been largely outside the empirical reach of the international multimedia-learning literature, which has historically focused on STEM and language content where the relevant cognitive structures are better characterized, including ubiquitous and guided digital learning environments (Hwang et al., 2021).

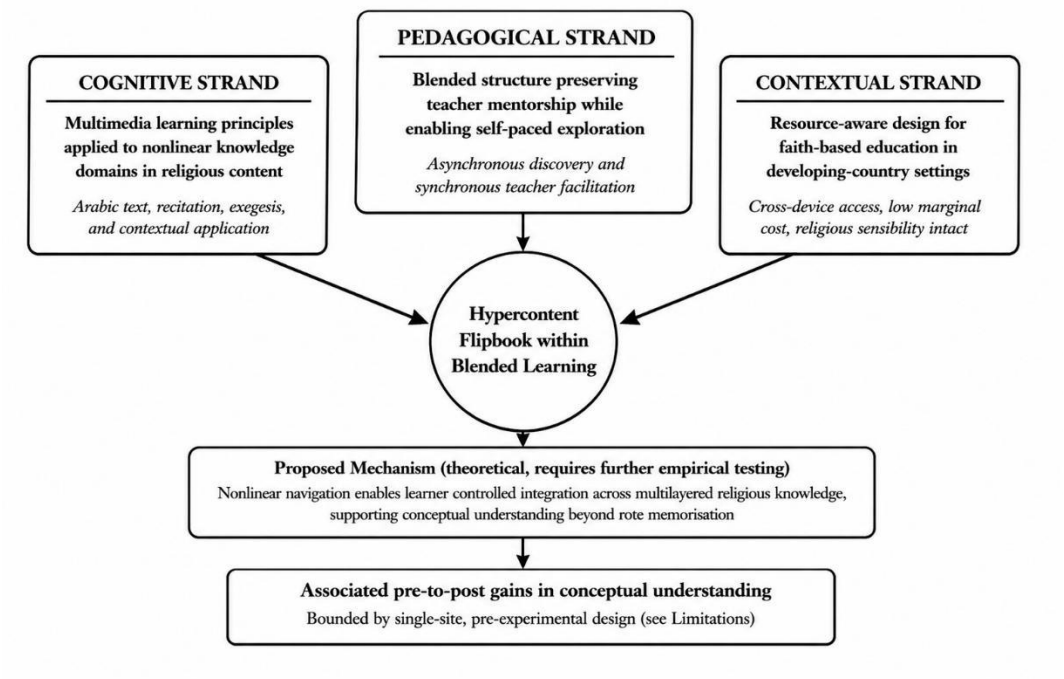

Figure 2. Multi-strand framework synthesizing the contributions of the present study to educational technology in resource-constrained, faith-based settings.

The pedagogical strand contributes to blended learning theory by demonstrating an asynchronous-then-synchronous sequence in which the synchronous phase is reserved for integration and consolidation rather than for content transmission. This structural choice aligns with contemporary continuum-based conceptualizations of blended learning that view synchronous time as a scarce resource best deployed for the harder integrative work that asynchronous time cannot easily support (Anthony et al., 2022; Schmidt & Glaser, 2021; Müller & Mildenberger, 2021; Nasution et al., 2021; Singh et al., 2021). The choice also explicitly preserves the teacher-student mentoring relationship that has been central to *madrasah* educational tradition for generations, rather than treating the digital component as a substitute for the teacher. In a context where some discussions of educational technology have implicitly framed the teacher as a bottleneck to be routed around, the present design treats the teacher as the integrative anchor that the asynchronous content makes all the more essential.

The contextual strand contributes resource-aware design considerations relevant to faith-based education in developing-country settings, considerations that much international educational technology research has not had to address. Cross-device compatibility for heterogeneous student devices, low marginal cost of digital reproduction, integration of religiously significant elements such as Arabic text and recitation, and teacher-led integration that does not require teachers to produce digital content from scratch are all design considerations that matter substantively in the Indonesian *madrasah* context and in comparable settings elsewhere. Design knowledge for faith-based, resource-constrained educational technology remains underdeveloped in the international literature, and the present study contributes a small but specific addition to that knowledge base by documenting design choices and their rationales (Rasheed et al., 2020). Across the three strands, the proposed mechanism, that nonlinear navigation supports conceptual integration across multilayered religious knowledge, is a hypothesis that the present study can articulate

clearly but cannot test definitively. The contribution of the work, properly bounded, is to make the hypothesis empirically tractable for subsequent controlled studies.

To make the contextual strand concrete, Figure 3 presents two representative screens from the developed flipbook: the chapter-selection main page and a content layer combining a concept map, Arabic text, translation, and concise exegesis. These screens illustrate the nonlinear architecture in operation rather than describe it abstractly, and they show how the multilayered structure of the content (text, translation, exegesis, application) is reflected in the artifact's visual organization. The interface design draws on established principles of multimedia learning while accommodating the specific affordances of Qur'anic-studies content, including the need to display Arabic script with appropriate typographic respect and to integrate audio recitation as a first-class content element rather than an optional add-on. Readers interested in the full set of screens, including the entry animation and the gamified evaluation interface, can consult the master's thesis from which this study is derived.



Figure 3. Representative screens from the hypercontent digital flipbook developed in the study

Several limitations of this study should be explicitly acknowledged, as they constrain what the work can claim and shape the design of the controlled investigations that should follow. The single-group pretest-posttest design without a control group means that the observed pretest-posttest association cannot be causally attributed to the intervention. Maturation, history, testing effects, novelty effects, Hawthorne effects, and teacher-attention effects all remain plausible alternative explanations. The 20-item multiple-choice test is a suboptimal measure of conceptual understanding because multiple-choice formats reward recognition over construction and can be answered correctly with partial knowledge. The test's reliability (KR-20 or Cronbach's alpha) was not formally estimated because raw item-level data were unavailable. The use of identical pretest and posttest items creates a substantial testing-effect risk consistent with the very high pre-post correlation observed. A single expert rated each of the four expert validation domains. To support psychometrically robust validity claims, Aiken's V or the Content Validity Index, accompanied by interrater reliability indices, would be required.

The sample of 36 students from one intact class at one state *madrasah* constrains generalisability to other educational settings, and no a priori power analysis was conducted to justify the sample size. Setting characteristics, including state status, urban location, East Javanese cultural context, and mid-resource infrastructure, all matter substantively when interpreting the findings or considering whether they might transfer elsewhere. The intervention combined hypercontent media features with a blended asynchronous-then-synchronous sequencing, and the design cannot isolate the contribution of hypercontent features from that of the blended structure. A dismantling study would be required to do so rigorously. Several reporting limitations should also be noted. For example, no formal Research Ethics Committee approval was obtained at the time of fieldwork (only departmental approval was secured), the test instrument blueprint was not exhaustively reported in the source thesis, and raw item-level response data were not available for psychometric analysis. These reporting gaps should be addressed proactively in any future replication.

CONCLUSION

This instructional development study has shown that a hypercontent digital flipbook within a blended learning structure is feasible to implement in a state *madrasah* context, is acceptable to students and teachers in practice, and is associated with substantial pretest-to-posttest score gains. However, the study cannot claim that the intervention causes those gains, that the multimedia learning theory it uses has been extended into new theoretical territory, or that the model is ready for system-wide scaling across the Indonesian *madrasah* network or elsewhere. Stronger claims would require controlled comparative designs, dismantling studies to identify which design features carry the pedagogical weight, and multi-site replication to establish boundary conditions for transferability.

Four future directions emerge in order of methodological priority. First, quasi-experimental or randomized designs comparing a hypercontent flipbook within a blended structure against the same blended structure without hypercontent features. Second, dismantling studies that systematically vary specific multimedia design principles within the hypercontent design. Third, multi-site replication across diverse *madrasah* contexts, including private and state *madrasahs*, rural and urban settings, and low- and mid-resource settings. Fourth, longer-term retention and transfer studies extending well beyond the immediate

posttest to assess whether observed gains reflect durable conceptual change rather than short-term performance inflated by testing or novelty effects.

The broader takeaway is methodological. Instructional development research in faith-based and resource-constrained settings can and should engage international theoretical and methodological resources without abandoning context-specific concerns. Explicit acknowledgment of design limitations strengthens rather than weakens the contribution of feasibility-oriented studies. The Indonesian *madrasah* system serves a substantial student population with content that demands genuine conceptual integration rather than rote memorization, and the field will be best served by research programs that combine theoretical rigor with practical attention to the conditions under which Indonesian students and teachers actually work.

ACKNOWLEDGMENT

The authors thank the principal, teachers, and Grade XI students of MAN Kota Batu for their participation in this study and for their patience with the demands of an instructional development research program. We thank the four expert reviewers from the Department of Educational Technology at Universitas Negeri Malang for their careful evaluation of the artifact and the validation instruments. We thank the two anonymous reviewers whose constructive critique on an earlier version of this manuscript substantially improved its theoretical grounding, methodological transparency, and presentation, and whose feedback shaped the framing of the limitations and the conclusion in particular. The first author thanks Universitas Negeri Malang for supervisory support during the master's thesis from which this study is derived, and Darunnajah University Jakarta for institutional support that enabled manuscript preparation. Any remaining errors of fact or interpretation are the responsibility of the authors alone.

REFERENCES

Abdullah, W., Degeng, I. S., & Praherdhiono, H. (2021). The Influence of Blended Learning to Understanding Concept. *Turkish Journal of Computer and Mathematics Education*, *12*(3), 3756-3761. https://turcomat.org/index.php/turkbilmat/article/view/1658

Anthony Jr, B., Kamaludin, A., Romli, A., Raffei, A. F. M., Phon, D. N. A. E., Abdullah, A., & Ming, G. L. (2022). Blended learning adoption and implementation in higher education: A theoretical and systematic review. *Technology, Knowledge and Learning*, *27*(2), 531-578. https://doi.org/10.1007/s10758-020-09477-z

Arif, M., & Abd Aziz, M. K. N. (2023). Islamic religious education learning model in the 21st century: Systematic literature review. *Indonesian Journal of Islamic Education Studies (IJIES)*, *6*(2), 237-262.https://doi.org/10.33367/ijies.v6i2.4417

Bahri, A., Idris, I. S., Muis, H., Arifuddin, M., & Fikri, M. J. N. (2021). Blended learning integrated with innovative learning strategy to improve self-regulated learning. *International Journal of Instruction*, *14*(1), 779-794. https://doi.org/10.29333/iji.2021.14147a

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. https://doi.org/10.1007/978-0-387-09506-6

Castro-Alonso, J. C., De Koning, B. B., Fiorella, L., & Paas, F. (2021). Five strategies for optimizing instructional materials: Instructor-and learner-managed cognitive load. *Educational Psychology Review*, *33*(4), 1379-1407. https://doi.org/10.1007/s10648-021-09606-9

Cheung, S. K., Kwok, L. F., Phusavat, K., & Yang, H. H. (2021). Shaping the future learning environments with smart elements: challenges and opportunities. *International Journal of Educational Technology in Higher Education*, *18*(1), 16. https://doi.org/10.1186/s41239-021-00254-1

Conklin, J. (1987). Hypertext: An introduction and survey. *Computer (United States)*, *20*(9). https://doi.org/10.1109/MC.1987.1663693

Dermentzi, E., Zotou, M., Tambouris, E., & Tarabanis, K. (2022). Using the problem based learning method and educational technologies to teach open data: A design-based research approach. *Education and information technologies*, *27*(6), 8859-8882. https://doi.org/10.1007/s10639-022-10995-9

DeStefano, D., & LeFevre, J. A. (2007). Cognitive load in hypertext reading: A review. *Computers in human behavior*, *23*(3), 1616-1641. https://doi.org/10.1016/j.chb.2005.08.012

Dwianto, A., Wilujeng, I., Prasetyo, Z., & Suryadarma, I. (2017). The Development of Science Domain Based Learning Tool Which is Integrated with Local Wisdom to Improve Science Process Skill and Scientific Attitude. *Jurnal Pendidikan IPA Indonesia, 6*(1), 23- 31. https://doi.org/10.15294/jpii.v6i1.7205

Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. John Wiley & Sons. https://doi.org/10.1002/9781118269558

Graham, C. R. (2006). Blended learning systems: Definition, current trends, and future directions. In C. J. Bonk & C. R. Graham (Eds.), *The handbook of blended learning* (pp. 3–21). Pfeiffer.

Halverson, L. R., & Graham, C. R. (2019). Learner engagement in blended learning environments: A conceptual framework. *Online learning*, *23*(2), 145-178. https://doi.org/10.24059/olj.v23i2.1481

Hasrah, H., Wiratman, A., & Kaharuddin, M. (2025). Development of FlippingBook-Based E-Modules Integrated with Islamic Values to Enhance Interest and Learning Outcomes. *Educational Journal of Learning Technology*, *2*(3), 249-260. https://doi.org/10.58230/edutech.v2i3.54

Hwang, G. J., Wang, S. Y., & Lai, C. L. (2021). Effects of a social regulation-based online learning framework on students' learning achievements and behaviors in mathematics. *Computers & Education*, *160*, 104031. https://doi.org/10.1016/j.compedu.2020.104031

Indonesian Ministry of Religious Affairs. (2024). *Education statistics of madrasah in Indonesia 2023/2024*. EMIS Pendis Kemenag. https://emispendis.kemenag.go.id/dashboard/

Landow, G. P. (1992). *Hypertext: The convergence of contemporary critical theory and technology*. Johns Hopkins University Press.

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review*, *36*(1), 8. https://doi.org/10.1007/s10648-023-09842-1

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers college record*, *115*(3), 1-47. https://doi.org/10.1177/016146811311500307

Müller, C., & Mildenberger, T. (2021). Facilitating flexible learning by replacing classroom time with an online learning environment: A systematic review of blended learning in higher education. *Educational research review*, *34*, 100394. https://doi.org/10.1016/j.edurev.2021.100394

Nababan, D. (2023). Building hypercontent-based teaching modules: Improving Christian education learning outcomes for adolescents. *ISAR Journal of Arts, Humanities and Social Sciences*, *1*(3), 60-64.

Nasution, A. K. P., Surbakti, A. H., Zakaria, R., Wahyuningsih, S. K., & Daulay, L. A. (2021, February). Face to face learning vs blended learning vs online learning (student perception of learning). In *Journal of Physics: Conference Series* (Vol. 1783, No. 1, p. 012112). IOP Publishing. https://doi.org/10.1088/1742-6596/1783/1/012112

Nelson, T. H. (1965, August). Complex information processing: a file structure for the complex, the changing and the indeterminate. In *Proceedings of the 1965 20th national conference* (pp. 84-100). https://doi.org/10.1145/800197.806036

OECD (2023), *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*, PISA, OECD Publishing, Paris, https://doi.org/10.1787/53f23881-en

Rahmatullah, A. S., Mulyasa, E., Syahrani, S., Pongpalilu, F., & Putri, R. E. (2022). Digital era 4.0: The contribution to education and student psychology. *Linguistics and Culture Review, 6*(S3), 89–107. https://doi.org/10.21744/lingcure.v6nS3.2064

Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2020). Challenges in the online component of blended learning: A systematic review. *Computers & education*, *144*, 103701. https://doi.org/10.1016/j.compedu.2019.103701

Salmerón, L., Strømsø, H. I., Kammerer, Y., Stadtler, M., & Van den Broek, P. (2018). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & P. van den Broek (Eds.), *Learning to read in a digital world* (pp. 91-120). John Benjamins. https://doi.org/10.1075/swll.17.04sal

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review*, *19*(3), 285-307. https://doi.org/10.1007/s10648-007-9046-3

Schmidt, M., & Glaser, N. (2021). Investigating the usability and learner experience of a virtual reality adaptive skills intervention for adults with autism spectrum disorder. *Educational Technology Research and Development*, *69*(3), 1665-1699. https://doi.org/10.1007/s11423-021-10005-8

Shemshack, A., & Spector, J. M. (2020). A systematic literature review of personalized learning terms. *Smart Learning Environments, 7*(1), Article 33. https://doi.org/10.1186/s40561-020-00140-9

Sihombing, A. A., & Fatra, M. (2021). Distance Learning During the Pandemic Era: Online Learning Experiences of State Madrasah Tsanawiyah Students During Covid-19 in Indonesia. *Analisa: Journal of Social Science and Religion*, *6*(1), 95-112. https://doi.org/10.18784/analisa.v6i01.1235

Simbolon, R., Saragih, A. H., & Situmorang, J. (2021, March). Development of Hypercontent Based Learning Model for High School Students. In *Journal of Physics: Conference Series* (Vol. 1811, No. 1, p. 012122). IOP Publishing.

Singh, J., Steele, K., & Singh, L. (2021). Combining the Best of Online and Face-to-Face Learning: Hybrid and Blended Learning Approach for COVID-19, Post Vaccine, & Post-Pandemic World. *Journal of Educational Technology Systems*, *50*(2), 140-171. https://doi.org/10.1177/00472395211047865

Suana, W., Maharta, N., Nyeneng, I. D. P., & Wahyuni, S. (2017). Design and implementation of Schoology-based blended learning media for basic physics I course. *Jurnal Pendidikan IPA Indonesia, 6*(1), 170–178. https://doi.org/10.15294/jpii.v6i1.8648

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

Topping, K. J., Douglas, W., Robertson, D., & Ferguson, N. (2022). Effectiveness of online and blended learning from schools: A systematic review. *Review of Education, 10*(2), Article e3353. https://doi.org/10.1002/rev3.3353

Wahyuni, S., Sanjaya, I. G. M., Erman, E., & Jatmiko, B. (2019). Edmodo-based blended learning model as an alternative of science learning to motivate and improve junior high school students' scientific critical thinking skills. *International Journal of Emerging Technologies in Learning, 14*(7), 98–110. https://doi.org/10.3991/ijet.v14i07.9980

Wang, M.-T., & Hofkens, T. K. (2020). Beyond classroom academics: A school-wide and multi-contextual perspective on student engagement. *Adolescent Research Review, 5*(4), 419–433. https://doi.org/10.1007/s40894-019-00115-z

Widiyatmoko, A., Taufiq, M., Purwinarko, A., Wusqo, I. U., & Darmawan, M. S. (2022). The effect of environmental pollution game-based learning on improving students' conceptual understanding and environmental awareness. *Journal of Innovation in Educational and Cultural Research, 3*(4), 691–700. https://doi.org/10.46843/jiecr.v3i4.344

Widyaningrum, R., & Ardiansyah, A. A. (2026). Development of a Hypercontent Book for the Program Evaluation Course in Educational Technology. *Tadris: Jurnal Keguruan dan Ilmu Tarbiyah*, *11*(1), 285-297. https://doi.org/10.24042/tadris.v11i1.28948

Zh, M. H. R., Ananta, M. P., Nasution, N. R., & Haq, M. A. (2025). The Effect of a Digital Flipbook Integrated with Gamified Wordwall Assessments on Learning Motivation and Outcomes in Islamic Education. *International Journal of Learning and Education*, *1*(2), 95-105. https://najahajournal.najahaofficial.id/index.php/IJLE/article/view/300