Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

When Engagement Is Mistaken for Critical Reading: A Construct Audit of Technology-Supported EFL Research

Dedi Rahman Nur
Universitas Widya Gama Mahakam Samarinda, Indonesia
d.blues84@gmail.com
Correspondence author Email: d.blues84@gmail.com

Paper received: February-2024; Accepted: July-2024; Publish: August-2024

Abstract

Technology-supported EFL studies often report engagement, motivation, usefulness, and perceived learning gains. These outcomes matter, but they do not necessarily show that students can evaluate evidence, question sources, detect bias, or defend a judgment. This study audits the constructs measured in five project manuscripts involving Blooket, Quizizz, Canva, Kahoot!, and Wayground. The corpus represents 133 participant-records across manuscripts, which were not pooled because the samples, instruments, and contexts differ. A comparative secondary construct audit classified each manuscript against seven dimensions: technology acceptance, engagement, language processing, feedback and self-monitoring, social-dialogic interaction, multimodal meaning-making, and epistemic critical reading. Engagement was directly represented in all five manuscripts. Acceptance, language processing, and feedback were also common. Social-dialogic interaction appeared directly in only one manuscript, multimodal meaning-making in one, and none directly measured source credibility, evidence quality, bias, counterargument, or justified critical judgment. The pattern suggests a measurement gap rather than a failure of the technologies themselves. Positive learner experience is best treated as an enabling condition, not a proxy for critical reading. The study proposes a Critical Reading Technology Measurement Framework that separates acceptance and engagement from comprehension, monitoring, dialogue, multimodal interpretation, and epistemic evaluation. Future technology-supported EFL research should pair perception measures with performance tasks that make critical-reading operations observable and assessable.

Keywords: critical reading; educational technology; EFL; construct audit; measurement; digital literacy

1. Introduction

Digital platforms now sit inside routine EFL teaching. Students answer game-based questions, arrange images and text, monitor scores, and receive feedback within seconds. These activities can make language learning more visible and more participatory. Recent reviews show that gamification frequently improves motivation and engagement, although effects vary with task design, game elements, and context (Chan & Lo, 2024; Li et al., 2023). Quizizz research, for example, has linked gamified formative assessment with stronger motivation and engagement (Zhang & Crawford, 2024). Leaderboards can also raise participation and performance under some conditions (Cigdem et al., 2024). These findings explain why

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

technology studies often begin with learner perception. They do not, however, settle what students actually learn to do with texts.

That distinction becomes consequential when a study invokes critical reading. Critical reading is not simply sustained attention to a text, nor is it equivalent to understanding vocabulary or reporting that a platform is useful. A critical reader has to make claims about meaning and then test those claims against evidence. Online reading adds further demands because source expertise, credibility, multimodal framing, and platform cues may all shape judgment. Research on credibility evaluation shows that readers must examine author expertise, benevolence, and evidence quality rather than rely on a single surface cue (Kiili et al., 2023). New media literacy likewise includes critical forms of consuming and producing, not merely functional digital access (Vergili & Kara, 2024). Fake-news detection provides a practical example: critical-thinking dispositions and new media literacy predict whether students can discriminate unreliable information (Orhan, 2023).

Measurement is therefore not a technical afterthought. It defines the claim a study is allowed to make. Higher-education research increasingly warns against treating self-report as equivalent to demonstrated competence. Rubric-based work on digital competence argues that self-assessment should be paired with external performance evidence when researchers want to infer actual capability (González-Mujico, 2024). Scale development studies make the same point from another direction: a construct needs explicit dimensions, item coverage, and validity evidence before its score can carry a substantive interpretation (Mejías-Acosta et al., 2024). Critical-thinking instruments also need to distinguish analysis, reasoning, questioning, evaluation, decision making, and action rather than collapse them into a general impression (Galindo-Domínguez et al., 2023).

A similar problem appears in engagement research. Engagement is valuable, but it is heterogeneous. It can refer to behavior, emotion, cognition, social participation, persistence, or traces recorded by a platform. A recent systematic review of learning analytics found considerable variation in how higher-education studies conceptualize and operationalize student engagement (Bergdahl et al., 2024). Learning analytics reviews also show a gradual shift from reporting activity data toward pedagogically informed measures that support learning processes (Paulsen & Lindsay, 2024). The implication is straightforward: clicking, completing, scoring, and enjoying are evidence of participation or experience. They become evidence of critical reading only when a task requires critical reading and the measure captures that requirement.

The EFL context makes this boundary easy to blur. Vocabulary access, comprehension, motivation, and metacognitive monitoring genuinely support more demanding reading. Online-reading research with university students shows that metacognitive awareness is associated with English reading comprehension (Anggia & Habók, 2024). Technology-mediated self-study can also improve critical-thinking performance when the intervention explicitly teaches and tests critical-thinking skills (Algouzi et al., 2023). The important phrase is explicitly teaches and tests. A platform may reduce friction, provide practice, or increase willingness to

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

participate. Those benefits are enabling conditions. They do not by themselves demonstrate evaluation of claims or evidence.

Multimodality creates another measurement challenge. Digital composing asks learners to coordinate language, image, layout, and other semiotic resources. Research agendas for L2 digital multimodal composing call for closer attention to task design, feedback, collaboration, assessment, and critical digital literacies (Jiang & Hafner, 2025). Multimodal assessment can reveal conceptual understanding that monolingual written tasks may overlook, but validity and manageability remain central concerns (Jiang et al., 2024). Critical literacy work with digital rewriting goes further by asking learners to interrogate representation and reframe familiar narratives (Lam & Putri, 2024). Again, the technology does not supply the critical operation. The task and assessment must make that operation visible.

This issue is particularly relevant to the five manuscripts in the present project archive. Each examines a familiar educational technology: Blooket, Quizizz, Canva, Kahoot!, or Wayground. Together they offer a useful snapshot of how technology-supported EFL studies are commonly framed. Their instruments cover perceptions, motivation, engagement, vocabulary learning, comprehension, feedback, confidence, and interaction. The manuscripts were not originally designed as critical-reading studies. That fact should not be hidden or corrected retrospectively. It creates a legitimate secondary research question: if these instruments were used to support a claim about critical reading, what parts of the construct would they actually measure, and what parts would remain invisible?

The study therefore treats the manuscript archive as a measurement corpus rather than as a pooled intervention dataset. Its novelty lies in auditing the distance between technology-related measures and critical-reading claims. The purpose is threefold. First, it identifies which learning constructs are directly represented across the five manuscripts. Second, it examines where measurement thins out as the construct moves from engagement toward epistemic judgment. Third, it proposes a Critical Reading Technology Measurement Framework for future EFL research. The framework is intended to prevent a common inference error: mistaking a positive technology experience for evidence that critical reading has occurred.

2. Method

2.1. Research Design and Source Corpus

The study used a comparative secondary construct-audit design. The source material consisted of five unique research manuscripts in one project archive. The archive also contained a near-duplicate revision of the Canva manuscript and a Turnitin similarity report associated with the Kahoot! manuscript. Those two files were reviewed for provenance but were not counted as independent evidence. The unit of analysis was the measurement design reported in each unique manuscript: questionnaire indicators, item content described in the text or tables, qualitative coding categories, interview prompts when reported, and the authors' stated interpretation of those measures.

The five manuscripts reported 133 participant-records in total: Blooket (n = 20), Quizizz (n = 18), Canva (n = 34), Kahoot! (n = 20), and Wayground (n = 41). This number is

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

descriptive only. It is not treated as a pooled sample, and no claim is made that the records represent 133 unique individuals across all projects. The studies used different instruments, scales, participant contexts, and research purposes. Combining their participant-level outcomes would therefore create a false appearance of comparability. The corpus was instead compared at the construct level.

This choice follows a broader measurement principle: the strength of an inference depends on alignment among the construct, task, observed evidence, and interpretation. Work on digital competence has shown why performance rubrics may be needed when self-assessment cannot establish actual capability (González-Mujico, 2024). Digital-literacy research in higher education also illustrates that competence profiles can be uneven across dimensions, making a single favorable global impression insufficient (Pegalajar Palomino & Rodríguez Torres, 2023). The present audit applies that logic to technology-supported EFL reading research.

Table 1. Source Corpus and Original Measurement Focus

Source	Reported participants	Original data and focus	Key reported evidence
Blooket	20 EFL students	10-item, 4-point questionnaire; perception and vocabulary-related skill	Grand M = 3.15/4; 94.5% positive; perception M = 3.20; skill M = 3.11
Quizizz	18 undergraduates; Indonesia and Kazakhstan	Open-ended questionnaire; motivation, engagement, competition, self-evaluation	Codes included motivation, active participation, competition, progress monitoring, achievement motivation
Canva	34 undergraduates; Indonesia and Kazakhstan	10-item, 5-point questionnaire; ease/usefulness, learning support, creativity/confidence	Overall M = 4.03/5; Q7 text-image-design understanding M = 4.41; Q9 confidence M = 3.76
Kahoot!	20 Indonesian university EFL students	10-item, 5-point questionnaire; expository learning, participation, feedback, interaction	Item means 3.40-4.15; learning effectiveness/participation highest; peer interaction lowest
Wayground	41 EFL learners	10-item, 5-point questionnaire plus interview support; perceptions, benefits, problems	70.0% positive; engagement, motivation, memory, feedback; time pressure and connectivity challenges

2.2. Audit Dimensions and Coding Rules

Seven dimensions were defined before cross-case coding. Technology acceptance covered usability, usefulness, and willingness to use a platform. Engagement covered motivation, enjoyment, participation, persistence, and competitive involvement. Language processing covered vocabulary learning, comprehension, and text understanding. Feedback and self-monitoring covered immediate feedback, error recognition, progress checking, and self-

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

evaluation. Social-dialogic interaction covered peer explanation, discussion, collaborative reasoning, or structured exchange. Multimodal meaning-making covered explicit coordination of linguistic and visual or spatial resources. Epistemic critical reading covered source credibility, evidence quality, bias or positioning, comparison of perspectives, counterargument, and justified judgment.

The last dimension was deliberately demanding. It was informed by recent work showing that credibility evaluation involves explicit judgments about source and evidence (Kiili et al., 2023), while critical new media literacy extends beyond functional use toward critical consuming and producing (Vergili & Kara, 2024). It also reflects the view that thinking skills should be intentionally taught and assessed rather than inferred from technology use (Dori & Lavi, 2023). The audit did not classify vocabulary knowledge, motivation, or comprehension as critical reading merely because those constructs can contribute to it.

Each manuscript received one of three codes for every dimension. Direct meant that the construct was explicitly represented in an indicator, item, qualitative code, or reported interpretation. Adjacent meant that the manuscript contained related evidence, but the construct was not directly operationalized. Absent meant that the reported instrument provided no defensible evidence for the dimension. These codes describe coverage, not quality. A manuscript could be methodologically appropriate for its original research question while still receiving an Absent code for critical reading.

Table 2. Construct-Audit Framework and Coding Rules

Dimension	Direct evidence required	Examples not sufficient by themselves
Technology acceptance	Explicit usability, usefulness, preference, or continued-use item	Platform exposure
Engagement/motivation	Explicit participation, enjoyment, effort, persistence, or motivation measure	Completion count alone
Language processing	Vocabulary, comprehension, inference, or text-understanding measure	Liking the platform
Feedback/self-monitoring	Feedback use, error checking, progress monitoring, or self-evaluation	Receiving a score without reflective use
Social-dialogic interaction	Peer explanation, discussion, collaborative reasoning, or revision	Competition or co-presence alone
Multimodal meaning-making	Explicit analysis/integration of text, image, layout, or modes	Use of visual templates alone
Epistemic critical reading	Source credibility, evidence quality, bias, perspective comparison, counterargument, justified judgment	Motivation, comprehension, confidence, or high quiz scores alone

2.3. Coding and Cross-Case Analysis

Coding proceeded in three passes. The first pass extracted the stated constructs and the numerical or qualitative evidence attached to them. The second pass mapped those constructs

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

to the seven audit dimensions without renaming an original measure. The third pass compared patterns across manuscripts and recorded which dimensions became sparse or disappeared. No new psychometric scale was created, and no reliability coefficient was calculated for the audit because the source manuscripts did not share participant-level items. The resulting frequencies are counts of manuscript-level construct coverage, not participant scores.

To limit interpretive drift, coding decisions were anchored to explicit wording in the source files. For example, Blooket's feedback/error-identification item was coded Direct for feedback and self-monitoring. Quizizz competition was coded Adjacent rather than Direct for social-dialogic interaction because competing against peers does not require explaining a position to them. Canva's item about combining images, text, and design was coded Direct for multimodal meaning-making. Kahoot!'s peer-interaction item was Direct for social-dialogic interaction even though it received the lowest mean in that manuscript. Critical reading was coded Absent unless an instrument explicitly required credibility, evidence, bias, perspective, counterargument, or justified evaluation.

The approach is consistent with learning-analytics research that calls for stronger alignment between collected traces and learning design (Drugova et al., 2024). It also reflects warnings that dashboards and digital indicators become more educationally meaningful when they are tied to learning processes rather than reported as analytics for their own sake (Paulsen & Lindsay, 2024). The present audit asks an analogous question of questionnaires: what learning process does each item actually make observable?

2.4. Integrity and Boundary Conditions

The analysis used only aggregate information available in the manuscript files. No personal identifiers or raw respondent records were accessed. Because the source studies were designed for other questions, their measures were not retroactively relabeled as critical-reading measures. This boundary is central to the study's integrity. Secondary analysis can generate new interpretations, but it cannot generate observations that were never collected. Before journal submission, the author should also confirm that reuse of the source manuscript data is consistent with the original consent, ethics, authorship, and publication arrangements.

3. Findings and Discussion

3.1. A Corpus Built Around Positive Learning Experience

The first pattern is not subtle. The corpus is strongest where technology research is usually strongest: acceptance, engagement, and perceived learning support. Four manuscripts directly measured technology acceptance or usefulness, while Quizizz contained adjacent evidence through positive descriptions of its scoring experience. All five directly represented engagement or motivation. Four directly represented language processing or comprehension, with Quizizz offering adjacent vocabulary-learning evidence rather than a separate performance measure. Four directly represented feedback or self-monitoring. The profile is coherent with the original aims of the studies. These manuscripts were designed to understand how students experience educational technology, not to test an epistemic reading construct.

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

This emphasis mirrors the wider literature. Reviews of EFL gamification repeatedly find motivation and engagement among the most commonly studied outcomes (Chan & Lo, 2024). Meta-analytic evidence also supports overall learning benefits from gamification while stressing variation across designs (Li et al., 2023). Quizizz studies provide a clear example of the attraction: points, feedback, and game structures can make formative assessment feel motivating rather than punitive (Zhang & Crawford, 2024). These are educationally useful outcomes. The measurement problem begins only when they are used as substitutes for a more demanding construct.

Engagement itself is not a single thing. Bergdahl et al. (2024) show that learning-analytics studies operationalize engagement in varied ways, including behavioral, emotional, cognitive, and social dimensions. The present corpus leans most heavily toward behavioral and emotional manifestations: participation, enjoyment, motivation, recommendation, competition, and willingness to continue. Cognitive engagement is present indirectly where students report comprehension or self-evaluation. Social engagement is comparatively thin. This pattern matters because critical reading is unlikely to be captured by engagement items unless those items refer to actual interpretive or evaluative behavior.

3.2. Comprehension and Feedback Sit in the Middle of the Measurement Chain

Language processing is better represented than critical evaluation. The Blooket study includes vocabulary-related skill items and a perception/skill distinction. The Canva study directly asks whether combining text, images, and design supports narrative understanding. Kahoot! links platform use with expository-text learning. Wayground reports vocabulary memory and learning benefits. These measures occupy an important middle position. Readers cannot evaluate an argument they have not understood. At the same time, comprehension does not guarantee evaluation.

Online-reading research reinforces that distinction. Metacognitive awareness of strategies is associated with English reading comprehension, suggesting that monitoring and strategy use are meaningful components of digital reading (Anggia & Habók, 2024). The source manuscripts also contain several feedback-related mechanisms. Blooket includes error identification. Quizizz participants describe self-evaluation and progress monitoring. Kahoot! offers immediate assessment information. Wayground reports rapid feedback as a benefit. These mechanisms could support rereading and revision if the next task requires them.

Formative assessment research helps clarify the missing step. Feedback becomes formative when it changes subsequent learning activity, not simply when a score appears on screen. Studies of formative assessment in higher education emphasize reflection, feedback use, and adjustment as parts of the process (Parmigiani et al., 2024). In the five manuscripts, feedback is usually measured as useful or motivating. The instruments rarely ask what students did after discovering an error. A future critical-reading study would need to observe whether feedback prompted students to revisit evidence, reconsider a source, or revise an interpretation.

This point also explains why a high platform score cannot stand in for critical reading. A learner may answer a vocabulary question quickly and accurately while accepting a later

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

claim uncritically. Conversely, a learner may struggle with lexical retrieval yet display sophisticated source skepticism once language support is provided. Critical thinking and technology can be positively related in EFL contexts (Schenck, 2024), and explicit technology-mediated critical-thinking instruction can produce gains (Algouzi et al., 2023). But those studies measure the target intellectual behavior more directly than the perception instruments in this corpus.

3.3. Interaction and Multimodality Are Unevenly Represented

The audit becomes thinner at the social-dialogic and multimodal layers. Only Kahoot! directly includes peer interaction as a measured construct, and it is the lowest reported item in that manuscript (M = 3.40). Quizizz is Adjacent because students mention competition and friends, but competition does not necessarily involve explanation or collaborative reasoning. The remaining manuscripts do not directly operationalize dialogic interpretation. This gap is important. Critical reading often becomes visible when a learner must defend one interpretation against another and return to the text for support.

Research on student-centered digital learning consistently places interaction among learners, teachers, content, and technologies at the center of meaningful participation (Otto et al., 2024). Learning experience design research also shows that motivational conditions can predict elaboration and critical-thinking behaviors, not merely engagement (Wong & Hughes, 2023). The practical implication is that a technology study should distinguish social presence from epistemic dialogue. Two students may be active in the same game without ever comparing reasons. A direct measure would ask whether they cite evidence, challenge an interpretation, or revise a claim after peer critique.

Multimodal meaning-making appears directly only in the Canva manuscript. Its strongest item, Q7 (M = 4.41), concerns understanding narrative stories through the integration of images, text, and design. This is the clearest bridge toward a richer literacy construct in the corpus. L2 multimodal research treats modes as resources for meaning, not as decorative additions (Jiang & Hafner, 2025). Multimodal assessment can expose conceptual understandings that conventional writing may miss (Jiang et al., 2024). Critical multimodal work can go further by asking learners to expose assumptions and reconstruct representation (Lam & Putri, 2024). The Canva item, however, stops at perceived understanding. It does not ask whether students can identify framing, omission, or ideological positioning.

This is not a criticism of Canva or of the original questionnaire. It is a reminder that representational richness does not automatically produce criticality. A platform may make it easier to juxtapose images and words, while a task may still reward visual polish rather than interpretive reasoning. The measurement instrument must reveal which of those outcomes occurred.

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

Table 3. Cross-Manuscript Construct Coverage

Construct	Blooket	Quizizz	Canva	Kahoot!	Wayground	Direct coverage
Technology acceptance/usability	D	A	D	D	D	4/5
Engagement/motivation	D	D	D	D	D	5/5
Language processing/comprehension	D	A	D	D	D	4/5
Feedback/self-monitoring	D	D	—	D	D	4/5
Social-dialogic interaction	—	A	—	D	—	1/5
Multimodal meaning-making	—	—	D	—	—	1/5
Epistemic critical reading	—	—	—	—	—	0/5

Note. D = direct; A = adjacent; — = absent. Coverage is coded at manuscript level and is not a participant score.

3.4. Critical Reading Is the Missing Direct Construct

The most consequential result is the final row of Table 3. None of the five manuscripts directly measures epistemic critical reading. No reported instrument requires students to judge source credibility, evaluate evidence quality, identify bias, compare competing perspectives, construct a counterargument, or defend a judgment with evidence. The studies sometimes use language such as deeper understanding, effectiveness, or progress. Those phrases should not be stretched beyond the evidence that produced them.

Credibility research shows why this matters. Kiili et al. (2023) found that evaluating online texts involves coordinated judgments about source expertise, benevolence, evidence, genre, and credibility ranking. That kind of task produces observable evidence of evaluation. A statement such as 'this platform helped me understand the text' does not. Orhan (2023) similarly demonstrates that critical-thinking dispositions and new media literacy contribute to fake-news detection. Detecting unreliable information is a performance problem, not simply an attitude toward the technology through which information appears.

The same boundary appears in digital-competence measurement. González-Mujico (2024) argues that self-assessment alone can be speculative when used to infer actual digital competence, and proposes rubric-based performance evidence. Mejías-Acosta et al. (2024) show the work required to develop and validate a measurement scale around explicit competence dimensions. Galindo-Domínguez et al. (2023) likewise validate a multidimensional model for critical-thinking perceptions. Even perception-focused work, then, benefits from construct clarity. A technology questionnaire that never includes evaluation or justification cannot support a conclusion that evaluation or justification improved.

This does not mean perception data are weak data. They answer different questions. Students are uniquely positioned to report whether a platform feels usable, motivating, stressful, or supportive. Wayground's time-pressure and connectivity concerns, Canva's confidence score, Kahoot!'s low peer-interaction item, and Quizizz comments about competition all provide actionable design evidence. The problem is inferential substitution: reporting one construct and concluding another.

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

3.5. From Acceptance to Epistemic Judgment: A Measurement Framework

The audit suggests a layered model rather than a binary distinction between 'technology use' and 'critical reading.' The proposed Critical Reading Technology Measurement Framework contains six sequential but non-deterministic layers. Layer 1 is technology access and acceptance. Layer 2 is engagement and persistence. Layer 3 is language processing and comprehension. Layer 4 is feedback use and metacognitive monitoring. Layer 5 is dialogic and multimodal interpretation. Layer 6 is epistemic evaluation and justified judgment. A learner can show strength at an earlier layer without automatically reaching a later one.

The framework is intentionally compatible with recent digital learning research. Learning analytics reviews recommend moving from raw activity indicators toward theory-informed learning processes (Drugova et al., 2024; Paulsen & Lindsay, 2024). Digital-literacy research similarly distinguishes functional competence from more critical forms of participation (Vergili & Kara, 2024). Open educational resources can support complex thinking when learners must adapt, connect, and critically work with knowledge rather than merely retrieve it (Sanabria-Z et al., 2024). These studies point toward the same design principle: the intellectual operation has to be specified before the platform trace or questionnaire item can be interpreted.

For EFL research, the framework also prevents a false choice between language learning and criticality. Vocabulary and comprehension remain legitimate outcomes. They may be prerequisites for later evaluation, especially for learners working in another language. The framework simply requires researchers to name the level they measured. If a study measures vocabulary retrieval and enjoyment, it should report vocabulary retrieval and enjoyment. If it claims critical reading, it should add a performance task that requires source, evidence, perspective, or argument evaluation.

Digital multimodality can be placed within the same logic. Canva or another composing platform can support critical reading if students must explain how image, wording, layout, and omission frame a perspective. Jiang and Hafner (2025) explicitly identify critical digital literacies as a research direction for L2 multimodal composing. Lam and Putri (2024) show how rewriting can be designed around critical literacy rather than presentation alone. A direct assessment could compare the original and redesigned artifact, score the learner's rationale, and examine whether the redesign addresses representation rather than aesthetics.

Gamified systems require a different extension. Their strongest affordances are rapid response, feedback, repetition, and motivation. These strengths are well documented (Chan & Lo, 2024; Cigdem et al., 2024; Zhang & Crawford, 2024). To reach critical reading, the game cycle should be followed by a slower evidence cycle. Students might answer an inference question, inspect response distributions, locate supporting evidence, compare a second source, and defend or revise the answer. The critical measure would score the defense or revision, not the leaderboard position.

The framework also clarifies how motivation should be interpreted. Engagement can create the time and willingness needed for demanding reading, yet it remains an enabling condition. Wong and Hughes (2023) found that motivational traits were associated with

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

elaboration and critical-thinking behaviors in online learning, which suggests a pathway rather than equivalence. Bergdahl et al. (2024) likewise show why engagement should be unpacked rather than treated as a universal proxy for learning. In practice, an EFL instrument can retain motivation items while adding direct evidence of what motivated students actually do with texts.

Table 4. Critical Reading Technology Measurement Framework

Layer	Target construct	Suggested observable evidence	Example technology-supported task
1	Access and acceptance	Usability, perceived usefulness, willingness to use	Complete platform task and report barriers
2	Engagement and persistence	Participation, effort, time-on-task, persistence	Repeat a quiz or revision after feedback
3	Language processing	Vocabulary, literal/inferential comprehension	Explain an inference and identify textual basis
4	Monitoring and feedback use	Error detection, strategy report, revision after feedback	Record why an answer changed after feedback
5	Dialogic/multimodal interpretation	Peer defense, comparison, visual-verbal analysis	Compare two readings or redesign a representation with rationale
6	Epistemic critical judgment	Credibility, evidence quality, bias, counterargument, justified decision	Rank sources, evaluate evidence, rebut a claim, and justify judgment

3.6. Implications for Technology-Supported EFL Research

A first implication concerns instrument architecture. Perception questionnaires should be modular. One module can measure usability and acceptance. Another can measure engagement. A third can measure learning support. If critical reading is a research outcome, a separate performance module should assess the relevant critical operations. This reduces construct contamination and makes null findings more informative. A platform might improve engagement while leaving source evaluation unchanged. That result would be theoretically meaningful rather than disappointing.

A second implication concerns evidence triangulation. A credible critical-reading study could combine a short perception scale with an authentic reading task, a rubric, and process data. Performance assessment is especially important when self-report may overestimate competence (González-Mujico, 2024). Critical-thinking assessment research shows that explicit dimensions can be operationalized rather than left as broad labels (Galindo-Domínguez et al., 2023). For online reading, source evaluation tasks already demonstrate how credibility judgments can be observed directly (Kiili et al., 2023).

A third implication is to preserve the educational strengths already visible in the five manuscripts. The audit is not an argument against Blooket, Quizizz, Canva, Kahoot!, or

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

Wayground. Engagement, rapid feedback, multimodal composition, and perceived usefulness can support learning design. Student-centered digital practices are most productive when they connect learner activity with content, peers, and pedagogical purpose (Otto et al., 2024). Formative feedback can also create valuable opportunities for reflection and adjustment (Parmigiani et al., 2024). The recommendation is to add a critical operation after these affordances, not replace them.

A fourth implication concerns digital literacy. Students may be proficient users of platforms while remaining vulnerable to poor evidence or persuasive presentation. Digital literacy among university students can vary across skill domains (Pegalajar Palomino & Rodríguez Torres, 2023), and critical new media literacy requires more than functional consumption (Vergili & Kara, 2024). For reading research, technology competence and epistemic vigilance should therefore be assessed separately. A student who navigates Canva fluently may still need explicit support to question the authority implied by an image. A student who excels on Kahoot! may still need practice comparing contradictory sources.

A fifth implication concerns claims in discussion sections. Technology studies often contain a rhetorical escalation: students liked the platform, therefore the platform improved learning, therefore it developed higher-order thinking. The first inference may be supported by perception data. The second requires stronger learning evidence. The third requires direct higher-order performance evidence. Dori and Lavi (2023) emphasize that thinking skills need appropriate models, methods, and assessment tools. Schenck (2024) shows that technology and critical-thinking orientations can be related, but correlational relationships do not eliminate the need to measure critical performance itself.

3.7. Methodological Boundaries

The audit has clear limits. It works from manuscript-level reporting rather than raw instruments for every study. Some questionnaire wording is described through results tables or narrative summaries, so the audit cannot estimate item-level coverage with psychometric precision. The five manuscripts also differ in research design. One uses open-ended qualitative responses, while the others rely mainly on Likert questionnaires. The coverage matrix should therefore be read as a map of what the manuscripts make visible, not as a ranking of methodological quality.

The corpus is also local to one project archive. It does not represent the full technology-supported EFL literature. Wider reviews indicate that gamified EFL research includes varied tools, populations, and outcomes (Chan & Lo, 2024), while digital multimodal composing research spans writing development, collaboration, assessment, and critical literacies (Jiang & Hafner, 2025). A larger construct audit across published studies would be needed before claiming that the same measurement pattern characterizes the field as a whole.

Finally, the framework proposed here is conceptual. It has not yet been validated as a scale or rubric. Future research should develop task specifications for each layer, conduct expert review, test inter-rater reliability for performance scoring, and examine relationships among perception, comprehension, and epistemic judgment. The scale-development work of

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

Mejías-Acosta et al. (2024) offers one model for validation, while recent research on reusable resources and complex thinking suggests that platform affordances can be aligned with explicit higher-order competencies (Sanabria-Z et al., 2024). The next step is empirical testing, not stronger rhetoric.

4. Conclusion

The five technology-supported EFL manuscripts provide substantial evidence about how learners experience digital tools. They show strong coverage of engagement, motivation, acceptance, vocabulary or comprehension support, and feedback. Those constructs are educationally meaningful. They also reveal a measurement boundary. Social-dialogic reasoning and multimodal meaning are only unevenly represented, while epistemic critical reading is not directly measured in any manuscript.

The main contribution of this study is therefore methodological rather than causal. It identifies the point at which a defensible technology claim becomes an unsupported critical-reading claim. Positive perception should be interpreted as an enabling condition. Critical reading requires separate evidence that students can evaluate sources, judge evidence, recognize positioning, compare perspectives, construct counterarguments, and justify conclusions. The proposed Critical Reading Technology Measurement Framework offers one way to keep those layers distinct while still recognizing the value of engagement and comprehension.

For future EFL research, the practical recommendation is modest but important: retain the technology measures that answer questions about usability, motivation, and learning support, then add a performance task for the critical-reading construct being claimed. A short source-ranking task, an evidence-defense prompt, a multimodal critique, or a counterargument task may provide more valid evidence than adding more general perception items. Technology can help create the conditions for critical reading. Measurement must show whether students actually crossed that threshold.

DECLARATION OF GENERATIVE AI

During the preparation of this manuscript, the author used ChatGPT (OpenAI) to support language editing, structural organization, and readability. The source data, analytical decisions, numerical results, interpretations, and reference verification were reviewed by the author. The author takes full responsibility for the accuracy, originality, integrity, and content of the manuscript.

5. References

Algouzi, S., Alzubi, A. A. F., & Nazim, M. (2023). Enhancing EFL students' critical thinking skills using a technology-mediated self-study approach: EFL graduates and labor market in perspective. PLOS ONE, 18(10), e0293273. https://doi.org/10.1371/journal.pone.0293273

Anggia, H., & Habók, A. (2024). University students' metacognitive awareness of reading strategies (MARS) in online reading and MARS' role in their English reading comprehension. PLOS ONE, 19(11), e0313254. https://doi.org/10.1371/journal.pone.0313254

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

==========

Bergdahl, N., Bond, M., Sjöberg, J., Dougherty, M., & Oxley, E. (2024). Unpacking student engagement in higher education learning analytics: A systematic review. International Journal of Educational Technology in Higher Education, 21, 63. https://doi.org/10.1186/s41239-024-00493-y

Chan, S., & Lo, N. (2024). Enhancing EFL/ESL instruction through gamification: A comprehensive review of empirical evidence. Frontiers in Education, 9, 1395155. https://doi.org/10.3389/feduc.2024.1395155

Cigdem, H., Ozturk, M., Karabacak, Y., Atik, N., Gürkan, S., & Aldemir, M. H. (2024). Unlocking student engagement and achievement: The impact of leaderboard gamification in online formative assessment for engineering education. Education and Information Technologies, 29, 24835–24860. https://doi.org/10.1007/s10639-024-12845-2

Dori, Y. J., & Lavi, R. (2023). Teaching and assessing thinking skills and applying educational technologies in higher education. Journal of Science Education and Technology, 32, 773–777. https://doi.org/10.1007/s10956-023-10072-x

Drugova, E., Zhuravleva, I., Zakharova, U., & Latipov, A. (2024). Learning analytics driven improvements in learning design in higher education: A systematic literature review. Journal of Computer Assisted Learning, 40(2), 510–524. https://doi.org/10.1111/jcal.12894

Galindo-Domínguez, H., Bezanilla, M.-J., Campo, L., Fernández-Nogueira, D., & Poblete, M. (2023). A teachers' based approach to assessing the perception of critical thinking in Education university students based on their age and gender. Frontiers in Education, 8, 1127705. https://doi.org/10.3389/feduc.2023.1127705

González-Mujico, F. de L. (2024). Measuring student and educator digital competence beyond self-assessment: Developing and validating two rubric-based frameworks. Education and Information Technologies, 29, 13299–13324. https://doi.org/10.1007/s10639-023-12363-7

Jiang, L. (George), & Hafner, C. (2025). Digital multimodal composing in L2 classrooms: A research agenda. Language Teaching, 58(4), 528–546. https://doi.org/10.1017/S0261444824000107

Jiang, L., Li, Z., & Leung, J. S. C. (2024). Digital multimodal composing as translanguaging assessment in CLIL classrooms. Learning and Instruction, 92, 101900. https://doi.org/10.1016/j.learninstruc.2024.101900

Kiili, C., Räikkönen, E., Bråten, I., Strømsø, H. I., & Hagerman, M. S. (2023). Examining the structure of credibility evaluation when sixth graders read online texts. Journal of Computer Assisted Learning, 39(3), 954–969. https://doi.org/10.1111/jcal.12779

Lam, K. Y., & Putri, E. W. (2024). A study on the affordances of digital fairy-tale rewriting: A critical literacy approach to digital multimodal composing. System, 127, 103536. https://doi.org/10.1016/j.system.2024.103536

Li, M., Ma, S., & Shi, Y. (2023). Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: A meta-analysis. Frontiers in Psychology, 14, 1253549. https://doi.org/10.3389/fpsyg.2023.1253549

Mejías-Acosta, A., D'Armas Regnault, M., Vargas-Cano, E., Cárdenas-Cobo, J., & Vidal-Silva, C. (2024). Assessment of digital competencies in higher education students: Development and validation of a measurement scale. Frontiers in Education, 9, 1497376. https://doi.org/10.3389/feduc.2024.1497376

Orhan, A. (2023). Fake news detection on social media: The predictive role of university students' critical thinking dispositions and new media literacy. Smart Learning Environments, 10, 29. https://doi.org/10.1186/s40561-023-00248-8

Otto, S., Bertel, L. B., Lyngdorf, N. E. R., Markman, A. O., Andersen, T., & Ryberg, T. (2024). Emerging digital practices supporting student-centered learning environments in higher education: A review of literature and lessons learned from the COVID-19 pandemic. Education and Information Technologies, 29, 1673–1696. https://doi.org/10.1007/s10639-023-11789-3

Borneo Educational Journal (Borju)
https://jurnal.fkip-uwgm.ac.id/index.php/Borju
Volume 6, Issues 2, August, 2024
EISSN : 2655-9323
Section : Research Article
Page : 340-354
DOI : 10.24903/bej.v6i2.2500

==========

Parmigiani, D., Nicchia, E., Murgia, E., & Ingersoll, M. (2024). Formative assessment in higher education: An exploratory study within programs for professionals in education. Frontiers in Education, 9, 1366215. https://doi.org/10.3389/feduc.2024.1366215

Paulsen, L., & Lindsay, E. (2024). Learning analytics dashboards are increasingly becoming about learning and not just analytics - A systematic review. Education and Information Technologies, 29, 14279–14308. https://doi.org/10.1007/s10639-023-12401-4

Pegalajar Palomino, M. del C., & Rodríguez Torres, Á. F. (2023). Digital literacy in university students of education degrees in Ecuador. Frontiers in Education, 8, 1299059. https://doi.org/10.3389/feduc.2023.1299059

Sanabria-Z, J., Alfaro-Ponce, B., González-Pérez, L. I., & Ramírez-Montoya, M. S. (2024). Reusable educational resources for developing complex thinking on open platforms. Education and Information Technologies, 29, 1173–1199. https://doi.org/10.1007/s10639-023-12316-0

Schenck, A. (2024). Examining relationships between technology and critical thinking: A study of South Korean EFL learners. Education Sciences, 14(6), 652. https://doi.org/10.3390/educsci14060652

Vergili, M., & Kara, M. (2024). An investigation of students and teachers' new media literacy: The contributing characteristics with the moderator role of gender. Research and Practice in Technology Enhanced Learning, 19, 029. https://doi.org/10.58459/rptel.2024.19029

Wong, J. T., & Hughes, B. S. (2023). Leveraging learning experience design: Digital media approaches to influence motivational traits that support student learning behaviors in undergraduate online courses. Journal of Computing in Higher Education, 35, 595–632. https://doi.org/10.1007/s12528-022-09342-1

Zhang, Z., & Crawford, J. (2024). EFL learners' motivation in a gamified formative assessment: The case of Quizizz. Education and Information Technologies, 29, 6217–6239. https://doi.org/10.1007/s10639-023-12034-7