Journal of General Education and Humanities
Vol. 5, No. 4, August 2026, pp. 5437 – 5448, https://doi.org/10.58421/gehu.v5i4.1877
ISSN 2963-7147

Reading Comprehension Across Prior-Exposure Groups in Interactive Flat Panel-Based Gamified Vocabulary Learning: A Non-Experimental Comparative Study

Sri Agustina[1], Ahmad Munir[2], Muhaimin Abdullah[3]
[1,2,3]Master Program in English Education, Faculty of Languages and Arts, Universitas Negeri Surabaya, Indonesia

Article Info

Article history:
Received 2026-07-14
Revised 2026-08-10
Accepted 2026-08-26

Keywords:
Gamified vocabulary learning
Interactive flat panel
Non-experimental comparative research
Prior exposure
Reading comprehension

ABSTRACT

This quantitative non-experimental comparative study examined the reading comprehension performance of 112 eleventh-grade students at an Indonesian public senior high school based on their prior exposure to Interactive Flat Panel-based gamified vocabulary learning. Before data collection, students in XI P8 and XI P9 had participated in six instructional sessions conducted over three weeks, in which gamified vocabulary activities were integrated with analytical exposition, hortatory exposition, and discussion texts (n = 56). In contrast, students in XI P6 and XI P7 had not received the same instructional exposure (n = 56). During data collection, all participants completed the same researcher-developed 30-item reading comprehension test under comparable assessment conditions. Descriptive statistics and a two-tailed Mann–Whitney U test were employed. Students with prior exposure obtained higher scores (M = 87.13, SD = 9.92; Mdn = 88.00) than those without prior exposure (M = 61.25, SD = 17.10; Mdn = 55.50). The score distributions differed significantly, U = 361.50, z = −7.03, p < .001, with a large rank-biserial correlation, $|r_{rb}|$ = .769. Because exposure was not randomized and was associated with intact class membership, the findings indicate a substantial group difference related to prior exposure rather than a causal effect of Interactive Flat Panel use.

Corresponding Author:
Sri Agustina
Master Program in English Education, Faculty of Languages and Arts, Universitas Negeri Surabaya, Indonesia
Email: shachie25@gmail.com

1. INTRODUCTION

Reading comprehension is central to academic literacy because it enables learners to obtain information, interpret concepts, evaluate arguments, and construct coherent meaning from written texts. In an English as a Foreign Language (EFL) context, comprehension extends beyond decoding or sentence-by-sentence translation. It requires the coordinated use of word recognition, vocabulary knowledge, grammatical processing, background knowledge, inferencing, text integration, strategic monitoring, and self-regulation. The Active View of Reading situates these processes within interacting word-recognition,

language-comprehension, bridging, and active self-regulation systems [1]. Similarly, a systematic review of adolescent EFL reading identifies foundational language abilities, higher-order language processes, cognitive resources, inference, and self-regulation as interdependent contributors to comprehension [2]. Learners may therefore demonstrate different combinations of word-level and higher-order reading skills rather than a single uniform reading profile [3].

The Indonesian Academic Competency Test, or *Tes Kemampuan Akademik* (TKA), reflects this multidimensional view of reading. The published English framework covers daily, vocational, and academic communication contexts and requires students to identify explicit information, interpret vocabulary in context, infer implicit meaning, recognize relationships among ideas, and evaluate information or a writer's position [4]. Publicly reported TKA information provides a relevant national context for examining students' English reading performance [5]. Although national-level information does not represent every school, it highlights the importance of investigating classroom experiences that may be associated with differences in students' reading comprehension performance.

Vocabulary knowledge is one of the strongest linguistic foundations of second-language reading. A meta-analysis covering more than 100 studies reported a substantial relationship between L2 vocabulary knowledge and reading comprehension [6]. Word knowledge supports efficient lexical access, accurate sentence interpretation, proposition integration, and inferential processing [7]. Vocabulary breadth increases the proportion of a text that learners can recognize, whereas vocabulary depth supports the interpretation of polysemy, collocation, word families, and contextual meaning [8]. When lexical access is slow or inaccurate, learners must devote more cognitive resources to individual words, leaving fewer resources to integrate ideas across a text.

Technology-assisted vocabulary learning may support vocabulary development by providing repeated encounters, multimodal representation, retrieval practice, contextual cues, and immediate feedback. Meta-analytic evidence indicates an overall advantage for technology-assisted L2 vocabulary learning, although outcomes vary by technology type, learner age, instructional mode, and assessment target [9]. Meaning-focused exposure may also support incidental vocabulary learning, but the extent of learning depends on repetition, input mode, and assessment conditions [10]. Technology should therefore be viewed as an instructional resource whose value depends on how effectively it supports form–meaning connections, contextual use, retrieval, and transfer.

Gamified vocabulary learning organizes instructional practice through game-design elements such as points, challenges, competition, time limits, rewards, progress indicators, and immediate feedback. Reviews of gamification in English instruction show that it has been applied to vocabulary, reading, grammar, listening, speaking, writing, motivation, and classroom participation [11], [13]. Meta-analytic evidence indicates positive average effects on learning outcomes, although the magnitude and consistency of these effects depend on instructional design and implementation [14]. Gamification may also support intrinsic motivation, autonomy, and relatedness, while producing more limited changes in perceived competence [15]. Reviews of K–12 gamification and the wider educational literature therefore caution that enjoyment and visible participation should not automatically be interpreted as evidence of sustained academic improvement [16], [17].

Vocabulary-focused studies provide stronger support for using gamification. Gamified instruction has been associated with vocabulary development and learner autonomy in higher education [18], as well as vocabulary gains among Indonesian senior high school students [19]. Gamified quizzes have also been linked with improved EFL reading performance [20], [21]. However, much of the existing evidence comes from individually accessed mobile or web-based platforms. Such platforms may be difficult to implement in schools where smartphone use is restricted, internet connectivity is unstable, or teachers must monitor many separate student devices.

An Interactive Flat Panel (IFP) provides a different instructional structure. It is a large touch-sensitive display that combines presentation, annotation, multimedia, and direct manipulation of digital materials. As a shared classroom medium, an IFP allows teachers and students to view the same task, discuss responses, and take turns touching the screen to drag, match, classify, select, or annotate vocabulary items. These visible actions may coordinate attention, discussion, and immediate feedback. Emerging research suggests that integrating an IFP with cloud-based media can support interactive and collaborative classroom experiences [22]. Related research on interactive whiteboards has also reported improvements in collaborative language performance, metacognitive activity, and co-regulation [23]. These studies do not show that the device itself improves learning; rather, they indicate that a shared interactive display can support structured language processing when combined with appropriate pedagogy. Before data collection, students in XI P8 and XI P9 participated in six IFP-based gamified vocabulary sessions over three weeks, integrated with analytical exposition, hortatory exposition, and discussion texts.

In contrast, students in XI P6 and XI P7 had not received the same instructional exposure because access to Interactive Flat Panels differed across classes. In the prior-exposure classes, students regularly took turns interacting directly with the display during word–meaning matching, vocabulary classification, drag-and-drop tasks, contextual questions, timed group challenges, competition, and immediate feedback. These sessions formed part of regular classroom instruction and were not introduced by the researcher as an experimental intervention. Such repeated participation may have supported lexical familiarity and more efficient access to word meanings during reading. Nevertheless, reading comprehension performance is also influenced by grammatical knowledge, background knowledge, inference, strategy use, motivation, reading habits, and prior proficiency [1], [2], [24]–[26]. Therefore, differences between the classes cannot be attributed solely to prior IFP experience.

Despite the growing literature on gamified language learning, limited research has examined whether prior instructional exposure to shared-display IFP-based gamified vocabulary learning is associated with differences in reading comprehension performance across intact secondary-school classes. Previous studies have mainly investigated newly introduced interventions, individually accessed mobile or web-based applications, vocabulary gains, motivation, participation, or reading outcomes under more controlled conditions. Comparisons involving pre-existing intact classes remain educationally relevant, but they require causal caution because exposure status may overlap with class composition, prior proficiency, learning habits, and other unmeasured characteristics [27].

Accordingly, this study employed a quantitative non-experimental comparative, or ex post facto, design. The prior-exposure groups existed before data collection and were not created through random assignment or a researcher-administered intervention. The main objective was to compare the reading comprehension performance of eleventh-grade students with and without prior exposure to IFP-based gamified vocabulary learning. Specifically, the study described each group's reading comprehension performance and examined whether their reading-score distributions differed significantly. The authors interpreted any observed difference as associated with prior instructional exposure rather than as evidence that IFP use caused the outcome [27]. The study addressed the following research questions: What is the reading comprehension performance of eleventh-grade students with prior exposure to Interactive Flat Panel-based gamified vocabulary learning. What is the reading comprehension performance of eleventh-grade students without prior exposure to Interactive Flat Panel-based gamified vocabulary learning. Is there a significant difference in reading comprehension performance between eleventh-grade students with prior exposure to Interactive Flat Panel-based gamified vocabulary learning and those without prior exposure.

2. METHOD

2.1 Research Design and Setting

The study used a quantitative, non-experimental comparative (ex post facto) design. This design is appropriate when naturally existing groups are compared, and the researcher neither manipulates the grouping condition nor randomly assigns participants [27]. The grouping variable was prior exposure to IFP-based gamified vocabulary learning, categorized as with prior exposure and without prior exposure. Prior exposure was operationally defined as repeated participation in the specified IFP-based activities before the common data-collection session. The outcome variable was reading comprehension.

The study was conducted at SMAN 1 Pare, East Java, Indonesia, during the 2025/2026 academic year. The school had integrated IFP technology into classroom instruction, but the limited number of available panels meant that earlier instructional exposure differed across classes. Before data collection, XI P8 and XI P9 had received continuous IFP-based gamified vocabulary instruction integrated with reading texts. In contrast, XI P6 and XI P7 had not received the same instructional exposure. During data collection, students in all four classes completed the same reading-comprehension test under comparable assessment conditions. No additional instructional treatment was administered during the testing session. The researcher did not create the prior-exposure groups, introduce random assignment, or administer an experimental pretest-posttest treatment.

2.2 Participants and Sampling

The final analytical dataset comprised 112 eleventh-grade students from four intact classes, with 28 students from each class. XI P8 and XI P9 formed the group with prior exposure, whereas XI P6 and XI P7 formed the group without prior exposure. Each group therefore contained 56 students. The dataset contained no missing reading-comprehension scores.

The four classes were selected purposively because their prior instructional histories were clearly documented. Records were included when students were officially enrolled, had verifiable exposure status, completed the same reading-comprehension assessment under standardized conditions, and had a complete de-identified record. Because the classes were not randomly sampled or assigned, the findings are interpreted primarily in relation to the participating classes.

Table 1. Distribution of Participants

Group	Classes	Students per class	Total
Without prior exposure	XI P6 and XI P7	28 + 28	56
With prior exposure	XI P8 and XI P9	28 + 28	56
Total	Four intact classes	-	112

2.3 Prior-Exposure Context and Verification

Prior exposure referred to continuous participation before data collection in vocabulary-learning activities that combined game-design elements with an IFP and were integrated with reading texts. XI P8 and XI P9 had used the panel regularly in previous lessons and had become familiar with it as an interactive learning medium. Students took turns directly touching the screen to match word meanings, classify vocabulary, drag and drop responses, answer contextual questions, and participate in timed group challenges involving points, competition, peer discussion, and immediate feedback. The target vocabulary was connected to argumentative texts, including analytical exposition, hortatory exposition, and discussion texts. XI P6 and XI P7 had not received this IFP-based instructional exposure before data collection. Prior exposure consisted of 6 instructional sessions conducted over 3 weeks. Each session lasted approximately 90 minutes and involved gamified vocabulary activities connected with analytical exposition, hortatory exposition, and discussion texts.

Exposure status was verified through class membership, teacher confirmation, instructional records, and available learning documentation. XI P8 and XI P9 consistently constituted the group with prior exposure, whereas XI P6 and XI P7 consistently constituted the group without prior exposure. The label without prior exposure did not mean that XI P6 and XI P7 had received no vocabulary instruction; it indicated that they had not participated in the specified IFP-based gamified vocabulary instruction before the common reading-comprehension assessment. Because exposure status was attached to intact classes, prior exposure could not be separated completely from possible class-level differences.

2.4 Reading-Comprehension Instrument

Reading comprehension was measured using the same researcher-developed 30-item test for both groups. The test was aligned with selected competencies and cognitive demands in the published TKA English framework [4]. It assessed textual comprehension, inferential understanding (including vocabulary in context and relationships among ideas), and evaluation and appreciation. The term TKA-aligned denotes alignment with the public framework; the instrument did not reproduce confidential official questions.

The test consisted of 19 conventional multiple-choice items, 9 multiple-response items, and 2 classification items. A weighted scoring system was used. Each conventional multiple-choice item contributed 1 point for a correct answer. For the multiple-response items, 1 point was awarded for each correctly selected required alternative: eight items had three correct alternatives (24 points), and one item had four correct alternatives (4 points), producing a maximum of 28 points. For the two classification items, 1 point was awarded for each correctly classified Yes/No label, producing a maximum of 6 points. Partial credit was therefore awarded for correctly identified alternatives or labels. The maximum weighted raw score was 53, calculated as 19 + 28 + 6. Scores were converted to a 0-100 scale using the formula: final score = (obtained weighted raw score / 53) × 100. The same texts, items, instructions, time allocation, administration conditions, answer key, and conversion formula were applied to both groups.

Table 2. Weighted Scoring Procedures for the Reading-Comprehension Test

Item format	Number of items	Scoring procedure	Maximum score
Conventional multiple choice	19	1 point per correct item	19
Multiple response	9	1 point per correctly selected required alternative	28
Classification	2	1 point per correctly classified category/label (Y/N)	6
Total	30	Weighted scoring based on required responses	53

Final score = (obtained weighted raw score / 53) x 100

An experienced senior high school English teacher reviewed the instrument. The review covered alignment with the research objectives, representation of the reading-comprehension indicators, suitability of the texts, item clarity, answer-key accuracy, distractor quality, and item difficulty. All 30 items were judged appropriate, clear, and feasible, and the instrument was considered suitable for use without revision. The internal consistency of the instrument was examined using responses from the 112 participants in the main study, and item analysis was conducted before the group comparison. The analysis yielded a raw Cronbach's alpha coefficient of .602 and an alpha based on standardized items of .700. Because the test combined different item formats and heterogeneous scoring weights, these coefficients were interpreted cautiously alongside expert review, item analysis, and standardized administration, consistent with recommendations for the context-sensitive evaluation of L2 reading assessments [28].

Table 3. Reading-Test Blueprint

Reading component	Principal indicators	Items
Textual comprehension	Explicit information, factual details, key ideas, and supporting information	10
Inferential understanding	Implicit meanings, relationships between ideas, and vocabulary in context	10
Evaluation and appreciation	Ideas, the author's position, arguments, purpose, and message	10
Total		30

2.5 Data Collection and Ethical Procedures

After obtaining institutional and school permission, the researchers checked exposure documentation and identified the final eligible participants. Before data collection, XI P8 and XI P9 received continuous IFP-based gamified vocabulary instruction integrated with reading texts, whereas XI P6 and XI P7 did not receive the same instructional exposure. During data collection, students in all four classes completed the same reading-comprehension test under comparable assessment conditions. The researchers administered no additional instructional treatment during the testing session. Researchers scored reading responses using the predetermined key and converted scores to the 0-100 scale.

The researchers obtained official permission from the school before data collection. Students were informed about the study's purpose and procedures, the voluntary nature of participation, and data confidentiality. The researchers replaced student identities with numerical codes and reported findings in aggregate form.

2.6 Data Analysis

Data screening examined duplicate cases, missing values, invalid class or exposure codes, and reading scores outside the possible 0-100 range. Descriptive statistics were calculated for reading comprehension in each prior-exposure group. Class-level statistics were also examined because exposure status coincided with intact class membership. The reported indices included sample size, mean, standard deviation, median, interquartile range, mean rank, and score range.

Distributional assumptions were examined using descriptive indices, Shapiro-Wilk tests, histograms, Q-Q plots, and boxplots. Because reading scores departed significantly from normality in both groups, a two-tailed Mann-Whitney U test was used at alpha = .05. Rank-biserial correlation (r_rb), which ranges from -1 to +1, was reported as the effect-size measure; its absolute value indicates magnitude. Absolute values below .30 were interpreted as small, values from .30 to below .50 as moderate, and values of .50 or above as large.

3. RESULTS AND DISCUSSION

3.1 Data Screening and Class-Level Reading Performance

The final dataset contained 112 valid cases and no missing reading-comprehension scores. Each class contributed 28 students, and each prior-exposure category contained 56 students. Across all participants, the mean reading score was 74.19, the median was 80.00, and scores ranged from 30 to 100.

Table 4. Reading-Comprehension Scores by Class

Class	Exposure status	N	Mean	SD	Median	IQR	Min-Max
XI P6	Without prior exposure	28	60.43	16.70	56.00	31	35-96
XI P7	Without prior exposure	28	62.07	17.77	55.00	26	30-94
XI P8	With prior exposure	28	87.46	9.41	91.50	7	53-100
XI P9	With prior exposure	28	86.79	10.56	87.00	15	66-100

Class means were similar within each exposure category: XI P6 = 60.43 and XI P7 = 62.07; XI P8 = 87.46 and XI P9 = 86.79. This consistency suggests that the overall reading pattern was not driven by only one class, although class membership remains a potential confounding factor.

3.2 Reading-Comprehension Performance Across Prior-Exposure Groups

The 56 students with prior exposure obtained a mean reading-comprehension score of 87.13 (SD = 9.92). Their median was 88.00, with an IQR of 12.00 and scores ranging from 53 to 100, indicating higher and comparatively consistent performance. The 56 students without prior exposure obtained a mean score of 61.25 (SD = 17.10). Their median was 55.50, with an IQR of 29.00 and scores ranging from 30 to 96, indicating lower and more varied performance. The mean difference was 25.88 points, and the median difference was 32.50 points.

Shapiro-Wilk tests indicated non-normal distributions in the group without prior exposure, $W(56) = .929$, $p = .003$, and in the group with prior exposure, $W(56) = .894$, $p < .001$. Scores in the prior-exposure group were concentrated toward the upper end of the scale, whereas scores in the group without prior exposure were more widely dispersed. These results supported using the Mann-Whitney U test.

Table 5. Reading-Comprehension Scores by Prior-Exposure Status

Group	N	Mean	SD	Median	IQR	Mean rank	Min-Max
Without prior exposure	56	61.25	17.10	55.50	29.00	34.96	30-96
With prior exposure	56	87.13	9.92	88.00	12.00	78.04	53-100

3.3 Mann-Whitney U Test for Reading Comprehension

The Mann-Whitney U test showed a statistically significant difference in reading-score distributions between students with and without prior exposure, $U = 361.50$, $z = -7.03$, $p < .001$. Students with prior exposure had a substantially higher mean rank (78.04) than students without prior exposure (34.96). The calculator returned $r_{rb} = -.769$ because of the order in which the groups were entered; the absolute value, $|r_{rb}| = .769$, indicated a large difference favoring the prior-exposure group.

The test therefore indicated a statistically significant difference between the two groups. Higher reading scores occurred much more frequently in the prior-exposure group. This result represents a substantial difference in score distributions and rank tendencies; it does not, by itself, establish that prior IFP exposure caused the difference.

Table 6. Mann-Whitney U Test for Reading Comprehension

Comparison	Mean rank: without prior exposure	Mean rank: with prior exposure	U	z	p	Rank-biserial
Reading comprehension	34.96	78.04	361.50	-7.03	< .001	.769 (large)

3.4 Discussion

The findings show a clear difference in reading-comprehension performance between the two naturally occurring groups. Students with prior exposure achieved

substantially higher mean and median scores, and their scores clustered toward the upper end of the scale. In contrast, students without prior exposure showed a lower central tendency and considerably wider variation. The large rank-biserial correlation confirms that the difference was not only statistically significant but also substantial in magnitude.

A plausible theoretical pathway lies in the close relationship between vocabulary knowledge and reading comprehension. Vocabulary supports efficient lexical access, accurate sentence interpretation, proposition integration, and inferential processing [6], [7]. When readers can retrieve familiar word meanings with less effort, they have more attentional capacity to connect ideas, monitor comprehension, and evaluate claims. Technology-assisted vocabulary learning may support this process through repeated encounters, multimodal cues, contextualized practice, retrieval demands, and immediate feedback [9], [10]. Gamification may further increase the frequency and salience of practice through goals, challenges, competition, and feedback cycles [11]-[14].

The IFP's direct-touch features may have enriched this broader instructional experience by making vocabulary work visible, shared, and physically interactive. Students did not merely view words; they dragged items, matched meanings, classified terms, selected contextual answers, and explained choices to peers. Such activities may combine retrieval, discussion, teacher mediation, peer feedback, and immediate correction. This interpretation is compatible with evidence that shared interactive displays can support collaborative activity, metacognitive processing, and co-regulation [22], [23]. Nevertheless, the present study did not isolate the separate influence of the screen, game elements, teacher guidance, peer interaction, or repeated vocabulary practice.

Causal caution remains essential. XI P8 and XI P9 consistently constituted the prior-exposure group, whereas XI P6 and XI P7 consistently constituted the group without prior exposure. The students were members of four intact classes, the groups were not formed through random assignment, and no formal baseline reading pretest was available. Although similar class means within each exposure category suggest the result was not driven by one unusually strong or weak class, the groups may have differed beforehand in English proficiency, vocabulary knowledge, reading habits, attendance, tutoring, motivation, peer composition, scheduling, or other classroom characteristics. The finding should therefore be described as a strong group difference associated with prior exposure rather than proof that the IFP-based experience alone produced the difference.

The score distributions also require careful interpretation. The prior-exposure group was concentrated near the upper end of the reading scale, raising the possibility of a ceiling effect that may have reduced the instrument's ability to distinguish among stronger readers. The raw reliability estimate of .602 was modest, although the alpha based on standardized items reached .700. Reliability in L2 reading assessment can be influenced by item number, format, difficulty, scoring procedures, testing mode, and the population being assessed [28]. Accordingly, the results are meaningful, but their precision should be considered alongside the instrument's measurement limitations.

For classroom practice, the findings support equitable and pedagogically deliberate access to shared interactive media. Schools may consider a transparent rotation schedule so that instructional access is not concentrated in the same classes. Teachers should connect game elements to substantive language work by requiring students to explain contextual

meanings, justify choices, compare interpretations, and apply vocabulary while reading complete texts. Future research should include baseline reading and vocabulary measures, a larger number of classes, direct vocabulary outcomes, and stronger measurement evidence. Quasi-experimental, longitudinal, matched, or multilevel designs would help distinguish the contribution of prior exposure from pre-existing learner and class differences.

4. CONCLUSION

This non-experimental comparative study examined reading comprehension among 112 eleventh-grade students with and without prior exposure to IFP-based gamified vocabulary learning. Before data collection, XI P8 and XI P9 received continuous IFP-based gamified vocabulary instruction integrated with reading texts, whereas XI P6 and XI P7 did not receive the same instructional exposure. During data collection, students in all four classes completed the same reading-comprehension assessment under comparable conditions. No additional instructional treatment was administered during the testing session.

Students with prior exposure demonstrated higher reading-comprehension performance ($M = 87.13$, $SD = 9.92$; $Mdn = 88.00$) than students without prior exposure ($M = 61.25$, $SD = 17.10$; $Mdn = 55.50$). The difference was statistically significant, $U = 361.50$, $z = -7.03$, $p < .001$, with a large rank-biserial effect, $|r_{rb}| = .769$.

The findings indicate that prior exposure to IFP-based gamified vocabulary learning was associated with substantially higher reading-comprehension performance. However, the result cannot be interpreted as proof that IFP use caused the difference because exposure was not randomized, no formal baseline reading pretest was available, and exposure status overlapped with intact class membership. The study therefore supports careful and equitable use of IFP-based gamified vocabulary learning and provides a basis for more controlled research rather than a definitive causal claim.

ACKNOWLEDGMENT

The authors acknowledge the school leadership, English teachers, and participating students of SMAN 1 Pare for their support. The authors also acknowledge the academic guidance provided by the Master Program in English Education, Faculty of Languages and Arts, Universitas Negeri Surabaya.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

De-identified data and supporting materials may be made available by the corresponding author for legitimate academic purposes, subject to institutional approval, participant-confidentiality requirements, and applicable data-protection procedures.

REFERENCES

[1] N. K. Duke and K. B. Cartwright, “The science of reading progresses: Communicating advances beyond the Simple View of Reading,” Reading Research Quarterly, vol. 56, suppl. 1, pp. S25-S44, 2021, doi: 10.1002/rrq.411.

[2] G. Vettori, L. Casado Ledesma, S. Tesone, and C. Tarchi, “Key language, cognitive and higher-order skills for L2 reading comprehension of expository texts in English as foreign language students: A

systematic review," Reading and Writing, vol. 37, pp. 2481-2519, 2024, doi: 10.1007/s11145-023-10479-3.

[3] Y. Shi and B. Lee, "Chinese EFL middle school learners' reading skills: A latent profile analysis," The Asia-Pacific Education Researcher, vol. 34, no. 5, pp. 1677-1688, 2025, doi: 10.1007/s40299-025-00980-4.

[4] Pusat Asesmen Pendidikan, Kerangka Asesmen Tes Kemampuan Akademik Jenjang SMA/MA/SMK/MAK/Sederajat: Bahasa Inggris. Jakarta, Indonesia: Kementerian Pendidikan Dasar dan Menengah, 2025.

[5] Pusat Asesmen Pendidikan, "Hasil Tes Kemampuan Akademik 2025/2026," Kementerian Pendidikan Dasar dan Menengah, 2025. [Online]. Available: https://tka.kemendikdasmen.go.id/hasiltka/

[6] S. Zhang and X. Zhang, "The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis," Language Teaching Research, vol. 26, no. 4, pp. 696-725, 2022, doi: 10.1177/1362168820913998.

[7] C. Perfetti and J. Stafura, "Word knowledge in a theory of reading comprehension," Scientific Studies of Reading, vol. 18, no. 1, pp. 22-37, 2014, doi: 10.1080/10888438.2013.827687.

[8] W. Grabe and J. Yamashita, Reading in a Second Language: Moving from Theory to Practice, 2nd ed. Cambridge, U.K.: Cambridge University Press, 2022.

[9] A. Yu and G. Trainin, "A meta-analysis examining technology-assisted L2 vocabulary learning," ReCALL, vol. 34, no. 2, pp. 235-252, 2022, doi: 10.1017/S0958344021000239.

[10] S. Webb, T. Uchihara, and A. Yanagisawa, "How effective is second language incidental vocabulary learning? A meta-analysis," Language Teaching, vol. 56, no. 2, pp. 161-180, 2023, doi: 10.1017/S0261444822000507.

[11] H. Dehghanzadeh, H. Fardanesh, J. Hatami, E. Talaee, and O. Noroozi, "Using gamification to support learning English as a second language: A systematic review," Computer Assisted Language Learning, vol. 34, no. 7, pp. 934-957, 2021, doi: 10.1080/09588221.2019.1648298.

[12] S. Zhang and Z. Hasim, "Gamification in EFL/ESL instruction: A systematic review of empirical research," Frontiers in Psychology, vol. 13, Art. no. 1030790, 2023, doi: 10.3389/fpsyg.2022.1030790.

[13] S. Chan and N. Lo, "Enhancing EFL/ESL instruction through gamification: A comprehensive review of empirical evidence," Frontiers in Education, vol. 9, Art. no. 1395155, 2024, doi: 10.3389/feduc.2024.1395155.

[14] M. Li, S. Ma, and Y. Shi, "Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: A meta-analysis," Frontiers in Psychology, vol. 14, Art. no. 1253549, 2023, doi: 10.3389/fpsyg.2023.1253549.

[15] L. Li, K. F. Hew, and J. Du, "Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: A meta-analysis and systematic review," Educational Technology Research and Development, vol. 72, no. 2, pp. 765-796, 2024, doi: 10.1007/s11423-023-10337-7.

[16] H. Dehghanzadeh, M. Farrokhnia, H. Dehghanzadeh, K. Taghipour, and O. Noroozi, "Using gamification to support learning in K-12 education: A systematic literature review," British Journal of Educational Technology, vol. 55, no. 1, pp. 34-70, 2024, doi: 10.1111/bjet.13335.

[17] F. Gini, S. Bassanelli, F. Bonetti, R. H. Mogavi, A. Bucchiarone, and A. Marconi, "The role and scope of gamification in education: A scientometric literature review," Acta Psychologica, vol. 259, Art. no. 105418, 2025, doi: 10.1016/j.actpsy.2025.105418.

[18] B. Panmei and B. Waluyo, "The pedagogical use of gamification in English vocabulary training and learning in higher education," Education Sciences, vol. 13, no. 1, Art. no. 24, 2023, doi: 10.3390/educsci13010024.

[19] D. I. R. Mubarok, M. Bilqis, and D. Imamyartha, "Gamified English instruction mediated by Kahoot! at senior high school: Evidence for vocabulary gains," IJIET, vol. 9, no. 1, pp. 113-126, 2025, doi: 10.24071/ijiet.v9i1.10435.

[20] P. L. Anggraini, S. A. Karim, and R. Y. Radjaban, "EFL students' achievement in reading comprehension through gamification Kahoot," Surakarta English and Literature Journal, vol. 6, no. 2, pp. 288-300, 2023, doi: 10.52429/selju.v6i2.153.

[21] M. Salim et al., "Using gamified quizzes to enhance students' English reading comprehension skills," Diksi, vol. 33, no. 1, pp. 125-142, 2025, doi: 10.21831/diksi.v33i1.74744.

[22] A. S. Bahar and Iriandy, "Enhancing the digital classroom experience through the integration of Interactive Flat Panel and Tug of Knowledge as cloud-based learning media," Information Technology Education Journal, vol. 4, no. 4, pp. 784-794, 2025, doi: 10.59562/intec.v4i4.11189.

[23] M. F. Teng, "Interactive-whiteboard-technology-supported collaborative writing: Writing achievement, metacognitive activities, and co-regulation patterns," System, vol. 97, Art. no. 102426, 2021, doi: 10.1016/j.system.2020.102426.

[24] M. Shortt, S. Tilak, I. Kuznetcova, B. Martens, and B. Akinkuolie, "Gamification in mobile-assisted language learning: A systematic review of Duolingo literature from public release of 2012 to early 2020," Computer Assisted Language Learning, vol. 36, no. 3, pp. 517-554, 2023, doi: 10.1080/09588221.2021.1933540.

[25] A. Hamada, H. Shimizu, Y. Hoshino, S. Takaki, and Y. Ushiro, "Robust evidence for the Simple View of second language reading: Secondary meta-analysis of Jeon and Yamashita (2022)," Studies in Second Language Acquisition, vol. 46, no. 5, pp. 1355-1372, 2024, doi: 10.1017/S0272263124000226.

[26] L. Ma and Z. Zhao, "Reading motivation and reading comprehension achievement among English majors in China: A descriptive correlational study," Heliyon, vol. 11, no. 3, Art. no. e42427, 2025, doi: 10.1016/j.heliyon.2025.e42427.

[27] J. W. Creswell and T. C. Guetterman, Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research, 7th ed. Boston, MA, USA: Pearson, 2024.

[28] H. Zhao and V. Aryadoust, "A meta-analysis of the reliability of second language reading comprehension assessment tools," Studies in Second Language Acquisition, vol. 47, no. 1, pp. 388-416, 2025, doi: 10.1017/S0272263124000627.