

Kahoot!-Supported Descriptive-Text Instruction: Effects on Higher-Order Thinking and Learning Interest among Vocational EFL Students

Ardhaniah Dwi Erlinah¹, Hariyanto¹, Kustiyowati¹

¹Universitas PGRI Argopuro Jember, Indonesia

Received: April 14, 2026 | Revised: June 24, 2026 | Accepted: August 5, 2026

Correspondence: Ellanrachmawatie99@gmail.com

Abstract

The pedagogical value of gamified student-response systems for developing higher-order thinking skills (HOTS) remains underexplored, particularly in English as a foreign language (EFL) vocational education. This study examined whether Kahoot!-based instruction improved students' HOTS-oriented learning outcomes and interest in learning English during descriptive-text instruction. A quasi-experimental pretest-posttest control-group design involved 72 tenth-grade students from an Indonesian vocational high school. Two intact classes formed an experimental group ($n = 36$), which received Kahoot!-based instruction, and a control group ($n = 36$), which received conventional teaching. Data were collected using a 15-item essay test targeting analysis, evaluation, and creation (C4–C6) and a 20-item questionnaire measuring attention, enjoyment, and engagement. Independent-samples t -tests assessed group differences, and Cohen's d estimated effect magnitude. The groups did not differ significantly at pretest in HOTS-oriented outcomes ($p = .179$) or learning interest ($p = .097$). At posttest, the experimental group outperformed the control group in cognitive outcomes ($M = 84.08$ vs. 78.03), $t(70) = 5.276$, $p < .001$, $d = 1.244$, and reported greater learning interest ($M = 85.44$ vs. 79.72), $t(70) = 5.217$, $p < .001$, $d = 1.230$. These findings indicate that Kahoot!-based instruction was associated with statistically significant and educationally meaningful improvements in both outcomes. The study extends research on gamified EFL learning by demonstrating Kahoot!'s potential to support learning interest and performance on cognitively demanding descriptive-text tasks in vocational classrooms. Nevertheless, the single-school, non-randomized design warrants cautious generalization.

Keywords: descriptive text; EFL vocational education; gamified learning; higher-order thinking skills; Kahoot!; learning interest

1. Introduction

English proficiency is important in vocational education because it supports access to field-specific knowledge, communication in multilingual workplaces, and graduates' recruitment and career prospects (Avci & Engin-Demir, 2021; Çal et al., 2022; Chan, 2019; Kamil & Muhammad, 2021). Descriptive-text instruction can contribute to this preparation when students describe people, tools, products, facilities, services, and work environments for purposes connected to their occupational specializations; vocational English is generally perceived as more relevant when its texts and tasks reflect learners' study programs and anticipated workplaces (Hajar & Triastuti, 2021; Natsir et al., 2022; Skarpaas, 2024; Skarpaas & Hellekjær, 2021). However, genre pedagogy in Indonesian and other secondary EFL contexts is often implemented only partially, with instruction emphasizing model texts, generic structures, and lexicogrammatical features while giving less systematic attention to the complete teaching–learning cycle and the genre's social purpose (Kartika-Ningsih & Gunawan, 2019; Liu & Chen, 2022; Nadjib & Triastuti, 2023; Suharyadi & Basthomi, 2020). Although such instruction can establish foundational genre knowledge, purposeful writing also requires learners to generate and select relevant content, organize ideas coherently, evaluate linguistic and rhetorical alternatives, and adapt a text to its audience and purpose (Driscoll et al., 2020; Lee & Mak, 2018; Negretti & McGrath, 2018; Tardy et al., 2020). Consequently, recognizing the stages and grammatical features of a model description does not necessarily enable students to elaborate content, maintain coherence, or make effective lexical and grammatical choices in an original text (Nuralisaputri & Megawati, 2023; Purnamasari et al., 2021; Ratnaningsih & Azizah, 2019). This limitation



© 2026 The Author(s).

Open access under the Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

is particularly consequential in vocational education, where communication is occupation- and genre-specific and requires context-sensitive decisions about what information to include, how to express it, and how to address particular audiences (Al Hilali & McKinley, 2021; Chan, 2021; Hoang et al., 2024; Sureeyatanapas et al., 2024). Descriptive-text instruction should therefore move beyond the reproduction of decontextualized patterns and prepare learners to construct clear, accurate, and professionally meaningful descriptions.

Developing this capacity requires instruction that engages higher-order thinking rather than restricting learning to the recognition and reproduction of textual features. Explicit reasoning instruction, inquiry-oriented activities, and opportunities to examine alternatives can strengthen critical and higher-order thinking (Abrami et al., 2015; Loyens et al., 2023; Simonovic et al., 2023). In this study, higher-order thinking refers to analysis, evaluation, and creation at the C4–C6 levels of the revised Bloom’s taxonomy (Adams, 2015; Larsen et al., 2022; Stanny, 2016). In descriptive-text learning, these abilities can be elicited when students analyze how vocabulary, grammar, and supporting details construct an intended representation; compare alternative descriptions; evaluate the relevance, accuracy, and coherence of information; and create an original text for a defined audience and communicative purpose (Ebadi & Rahimi, 2018; Wu et al., 2024; Yin et al., 2023). The genre itself is therefore not inherently lower-order: its cognitive demand depends on task design, including the reasoning required, the support provided, and the number of information elements learners must integrate (Frear & Bitchener, 2015; Révész et al., 2017; Xu et al., 2022). These abilities are relevant to vocational students because workplace English and professional writing require learners to interpret technical content and adjust communicative choices to occupational situations, genres, and audiences (Kim, 2021; Moore & Morton, 2017; Ramamurthy et al., 2021). Nevertheless, teacher-led explanation, recall-oriented questioning, textbook-answer routines, and model reproduction can limit opportunities to justify decisions, consider competing interpretations, and construct texts independently; studies of EFL classrooms and materials continue to document the underrepresentation of analysis, evaluation, and creation and the practical difficulty of implementing higher-order tasks (Assaly & Jabarin, 2024; Erdiana & Panjaitan, 2023; Tyas et al., 2020).

Yet opportunities for higher-order thinking constitute only one part of the instructional challenge; students must also be willing to sustain the attention, effort, and participation that cognitively demanding tasks require. Learning interest is therefore a complementary—but analytically distinct—outcome. As an affective-motivational orientation toward particular content or activities, interest is associated with focused attention, positive affect, persistence, active participation, and learning performance (Harackiewicz et al., 2016; Nuutila et al., 2021; Philp & Duchesne, 2016; Rieger et al., 2022). By contrast, repetitive and predictable instruction with limited opportunities for contribution can generate boredom or disengagement and reduce meaningful involvement in EFL tasks (Derakhshan et al., 2021; Henry & Thorsen, 2020; Kruk et al., 2021; Pawlak et al., 2020). This distinction matters because analyzing details, evaluating alternatives, and constructing original descriptions require sustained cognitive investment; interest does not constitute these abilities, but it can encourage the information seeking, deeper processing, and persistence needed to exercise them (Bernacki & Walkington, 2018; McIntyre et al., 2021; Murayama et al., 2019; Oga-Baldwin et al., 2017). Higher-order thinking and learning interest should therefore not be conflated: the former concerns demonstrated cognitive performance, whereas the latter concerns learners’ focused attention and positive orientation toward the learning activity (Braun et al., 2020; Huber & Kuncel, 2016; Knogler et al., 2015; Tiruneh et al., 2018). Consistent with research on situational interest and engagement, the present study operationalizes learning interest through attention, enjoyment, and active engagement and examines it alongside, rather than as a proxy for, higher-order-thinking performance (Oga-Baldwin, 2019; Oga-Baldwin & Nakata, 2017; Renninger et al., 2019; Renninger & Hidi, 2020).

Against this background, Kahoot! offers a means of addressing these cognitive and affective instructional needs together, although its contribution depends on how teachers design and integrate it. As a gamified student-response system, Kahoot! enables teachers to present questions to the whole class, collect simultaneous responses, display answer distributions, and provide immediate feedback. Its scoring, time limits, audiovisual displays, leaderboards, competition, and rapid participation cycle have been associated with greater attention, enjoyment, motivation, and classroom participation (Cameron & Bizo, 2019; Ismail et al., 2019; Licorish et al., 2018; Turan & Meral, 2018). These features can support higher-order descriptive-text learning when the questions require students to analyze competing descriptions, discriminate between relevant and irrelevant

details, evaluate lexical and grammatical choices, and solve context-specific communicative problems. Research on online and multiple-choice assessment indicates that scenario-based, application-oriented, and deliberately constructed items can elicit higher-order processing; the response format alone does not determine cognitive demand (Haataja et al., 2023; Lenchuk & Ahmed, 2021; Maryani et al., 2021; Scully, 2017). Discussion and explanatory feedback following each response can extend the activity beyond answer selection by requiring students to justify choices, compare interpretations, and revise their reasoning (Knight & Brame, 2018; Molin et al., 2021; Tullis & Goldstone, 2020; Vickrey et al., 2015). Kahoot!, however, does not inherently foster higher-order thinking. It is often used for review or retrieval, while insufficient response time, time-based scoring, and an emphasis on rank can encourage anxiety, shallow coverage, or rapid guessing instead of considered reasoning (Liang, 2024; Owen & Licorish, 2020; Poblaciones et al., 2021; Wang & Tahir, 2020). More broadly, gamification produces variable effects because the value of its mechanics depends on their alignment with learning objectives, meaningful content, learner needs, interaction, and the wider instructional design (Bai et al., 2020; Dichev & Dicheva, 2017; Sailer et al., 2017; Sailer & Homner, 2020). Accordingly, Kahoot! is better understood as a design-dependent pedagogical scaffold than as a tool that inherently produces either engagement or higher-order learning.

Although this pedagogical rationale is plausible, evidence concerning Kahoot!'s contribution to cognitively demanding EFL learning remains less developed than evidence of its motivational benefits. Previous research has predominantly examined enjoyment, motivation, participation, classroom interaction, and general achievement, with Kahoot! commonly used for quizzes, review, or formative assessment. Consequently, it remains insufficiently understood whether deliberately designed Kahoot!-supported instruction can improve learners' demonstrated abilities to analyze, evaluate, and create within a specific EFL genre rather than merely enhance engagement or factual recall. Limited attention has also been given to higher-order thinking and learning interest as simultaneous but conceptually distinct outcomes, particularly in vocational EFL classrooms, where students must use English for context-specific communicative purposes. This study addresses these gaps by integrating C4–C6 questions, sufficient reasoning time, explanatory feedback, whole-class discussion, and follow-up writing tasks into Kahoot!-supported descriptive-text instruction. Its novelty lies not simply in introducing Kahoot! as a digital tool but in employing it as a structured pedagogical scaffold for higher-order thinking while separately assessing cognitive performance and learning interest in an Indonesian vocational-school context. Using a quasi-experimental pretest–posttest control-group design, the study aimed to determine whether tenth-grade vocational EFL students receiving Kahoot!-supported instruction demonstrated greater improvement in HOTS-oriented descriptive-text performance and learning interest, measured through attention, enjoyment, and engagement, than those receiving conventional instruction.

2. Method

2.1 Research Design and Participants

This study used a quasi-experimental pretest–posttest nonequivalent control-group design to examine differences in higher-order-thinking-skills (HOTS)-oriented descriptive-text performance and learning interest between students who received Kahoot!-supported instruction and those who received conventional instruction. Quasi-experimental designs are appropriate for evaluating educational interventions when existing groups must be retained and random assignment of individual participants is impractical or infeasible, although the absence of random assignment increases the importance of addressing selection differences and alternative explanations (Denny et al., 2023; Gopalan et al., 2020).

The study was conducted at an Indonesian vocational high school during the 2025 academic year. The target population comprised tenth-grade students, from whom two intact classes containing 72 students were selected purposively while retaining their existing class membership. One class was designated as the experimental group ($n = 36$), and the other was designated as the control group ($n = 36$). Because students were not randomly assigned individually, the groups were treated as nonequivalent at the design stage. Both groups completed the outcome measures before and after the intervention. Pretest scores were analyzed to describe initial between-group differences, but nonsignificant pretest results were not interpreted as evidence that the groups were equivalent because failure to reject a difference does not establish the absence of a meaningful difference; such a conclusion

would require an appropriately specified equivalence test (Lakens, 2017). Because each condition was represented by only one intact class, class membership and instructional condition were perfectly aligned. Under educational evidence standards, this constitutes a single-unit confound because differences attributable to the intervention cannot be separated fully from differences associated with the particular class or teacher (Page et al., 2020; What Works Clearinghouse, 2017). Accordingly, the findings should be interpreted as comparisons between the two participating classes rather than as class-independent causal estimates of the effect of Kahoot!.

Table 1.

Pretest–Posttest Nonequivalent Control-Group

Group	Pretest	Instruction	Grades taught
Experimental	O ₁	X	O ₂
Control	O ₁	—	O ₂

Note. O₁ = pretest; X = Kahoot!-supported descriptive-text instruction; — = conventional descriptive-text instruction; O₂ = posttest.

2.2 Instruments

Two instruments were administered at pretest and posttest: a 15-item constructed-response test assessing HOTS-oriented descriptive-text performance and a 20-item questionnaire assessing interest in learning English.

Table 2.

Design Research Variables and Instruments

Outcome variable	Indicators	Instrument	Number of items	Scoring
HOTS-oriented descriptive-text performance	Analyzing (C4), evaluating (C5), and creating (C6)	Constructed-response test	15	0–4 points per item; raw total of 0–60, converted to a 0–100 scale
Interest in learning English	Attention, enjoyment, and engagement	Questionnaire	20	Five-point responses from 1 (strongly disagree) to 5 (strongly agree); total score of 20–100

The cognitive instrument was a 15-item constructed-response test designed to measure students' ability to analyze, evaluate, and create in the context of descriptive texts. These cognitive categories were based on the revised Bloom's taxonomy (Anderson & Krathwohl, 2001). Because cognitive complexity cannot be determined solely from the action verbs used in an item, the tasks were designed to require students to demonstrate the intended reasoning processes through textual analysis, evidence-based judgment, and original production (Stanny, 2016). Students read descriptive texts or prompts and produced written responses rather than selecting answers from predetermined alternatives. The C4 items required students to identify relationships among textual details and analyze how language features contributed to a description. The C5 items required students to judge the relevance, accuracy, or effectiveness of descriptive content and justify their judgments with textual evidence. The C6 items required students to produce original descriptive responses or paragraphs for a specified context. The instrument therefore assessed receptive analysis and productive language use rather than descriptive-text comprehension alone.

Table 3.

Distribution of HOTS Test Items

HOTS level	Cognitive indicator	Number of items
C4 (Analyzing)	Identify, classify, and analyze information and language features in descriptive texts	5
C5 (Evaluating)	Evaluate content and language use and justify interpretations with textual evidence	5
C6 (Creating)	Produce original descriptive responses and construct descriptive paragraphs	5
Total		15

Each response was scored from 0 to 4 using item-specific analytic criteria. The criteria evaluated the cognitive operation required by each item—including depth of analysis, accuracy and justification of evaluation, or relevance and originality of creation—together with the clarity of the written response. Analytic rubrics can

support more consistent assessment of complex performance when the criteria and performance levels are defined clearly, although their effectiveness also depends on appropriate rater preparation and application (Jonsson & Svingby, 2007). Raw scores could range from 0 to 60 and were converted to a 0–100 scale for analysis and reporting. The same scoring criteria and item-specific rubrics were applied consistently to the pretest and posttest responses.

The 15 scored items yielded Cronbach's $\alpha = .938$. This coefficient was interpreted as a sample-specific estimate of consistency among item scores rather than as evidence of unidimensionality, overall instrument validity, or agreement between scorers (Taber, 2018). Inter-rater reliability is a separate form of reliability that requires ratings from two or more scorers to be compared using an appropriate agreement or consistency coefficient (Hallgren, 2012). Because a separate inter-rater reliability coefficient was not available, the consistency of human scoring could not be evaluated empirically.

Table 4.
Representative HOTS Test Items

HOTS level	Cognitive indicator
C4 (Analyzing)	Identify and explain three descriptive features used by the author to describe the object.
C5 (Evaluating)	Evaluate whether the author's description effectively helps readers visualize the object, and justify your judgment with evidence from the text.
C6 (Creating)	Write a short descriptive paragraph about a tourist attraction in your local area using appropriate descriptive language features.

The learning-interest questionnaire comprised 20 items addressing attention, enjoyment, and engagement as indicators of students' affective orientation toward learning English. Students responded on a five-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Item coding was oriented consistently so that higher values indicated greater learning interest. The item scores were summed to produce a total score ranging from 20 to 100. Because the study treated learning interest as one overall outcome, separate dimension scores were not used in the inferential analyses. The total scale yielded Cronbach's $\alpha = .909$, which was interpreted as evidence of internal consistency among the questionnaire items in the present sample, subject to the recognized limitations of coefficient alpha (Taber, 2018). Before data collection, the instruments were reviewed by two doctoral-level specialists, one in English language education and one in educational technology. Expert evaluation of item relevance, clarity, representativeness, and alignment with the intended construct constitutes validity evidence based on test content, but it represents only one component of a broader validity argument for the proposed interpretations and uses of test scores (Kane, 2013; Sireci & Faulkner-Bond, 2014).

A feasibility percentage was calculated for each instrument using the formula (*total score awarded / maximum possible score*) $\times 100$. The study applied the following predetermined interpretation criteria: 81%–100% = *very appropriate*, 61%–80% = *appropriate*, 41%–60% = *fairly appropriate*, 21%–40% = *less appropriate*, and 20% or below = *not appropriate*. The HOTS test obtained a feasibility rating of 82%, and the learning-interest questionnaire obtained a rating of 97%; both therefore fell within the study's *very appropriate* category. These percentages are reported as evidence of expert-perceived content suitability rather than as conclusive proof of instrument validity. Because the retained validation summary did not contain item-level expert ratings, a content-validity index or Aiken's *V* could not be reconstructed. Aiken's *V* requires the ratings assigned by individual experts to individual items rather than only an overall percentage (Aiken, 1980). Internal-consistency analysis yielded Cronbach's $\alpha = .938$ for the HOTS test and $\alpha = .909$ for the learning-interest questionnaire.

2.3 Research Procedure

The study consisted of three stages: pretesting, intervention, and posttesting. At pretest, both groups completed the HOTS-oriented descriptive-text test and the learning-interest questionnaire under the same administration conditions. The intervention was subsequently delivered during the school's scheduled English lessons within the descriptive-text unit. Both groups addressed the same descriptive-text learning objectives, but their modes of instructional delivery differed. In the experimental group, Kahoot! was integrated as a scaffold for cognitively demanding learning rather than being used solely for factual recall. The activities included C4–C6 questions, sufficient reasoning time, displays of class-response patterns, immediate explanatory feedback, and teacher-led

discussions in which students justified their answers and considered alternative interpretations. Kahoot! activities were followed by written tasks requiring students to apply their reasoning when constructing original descriptions. These procedures were intended to combine active participation and timely feedback with higher-order cognitive engagement. The control group studied the same descriptive-text content through conventional teacher-led instruction without Kahoot! or another gamified digital response system. Instruction consisted of teacher explanation, textbook-based analysis of descriptive texts, oral questioning, individual written exercises, and descriptive-writing practice. After the instructional phase, both groups completed the same two outcome instruments under administration procedures consistent with those used at pretest. The defining treatment contrast was the integration of Kahoot!-supported higher-order questions, feedback, and discussion in the experimental condition versus conventional instruction in the control condition. No numerical treatment-fidelity index was retained. Because implementation-fidelity evidence helps establish whether an intervention was delivered as intended and whether the planned difference between conditions was maintained, the absence of such an index limits independent verification of the instructional contrast (O'Donnell, 2008).

2.4 Data Analysis

The two outcome variables were analyzed separately. All 72 participants contributed complete pretest and posttest scores; consequently, no missing-value imputation was required, and no cases were excluded from the reported analyses. Means and standard deviations were calculated for each group at pretest and posttest. Shapiro–Wilk tests were used to assess whether the score distributions showed statistically detectable departures from normality within each group (Shapiro & Wilk, 1965). Levene's test was used to assess the equality of variances between groups (Gastwirth et al., 2009). For both procedures, $p > .05$ was interpreted only as indicating that the corresponding null hypothesis had not been rejected; it was not treated as proof that the assumption was satisfied exactly because nonsignificant results do not establish the truth of a null hypothesis (Wasserstein et al., 2019).

Two-tailed independent-samples t tests compared the experimental and control groups at pretest and posttest, with the significance level set at $\alpha = .05$. The pretest tests described initial group differences, whereas the posttest tests examined between-group differences following instruction. These separate posttest comparisons did not statistically adjust posttest scores for pretest performance. Analytic approaches such as ANCOVA and change-score analysis explicitly incorporate baseline information but rely on different assumptions when applied to nonrandomized two-wave data (Lüdtke & Robitzsch, 2025). Cohen's d , calculated using the pooled within-group standard deviation, quantified the standardized magnitude of each between-group difference (Lakens, 2013). Ninety-five percent confidence intervals were calculated for the posttest mean differences and effect sizes to communicate the precision and range of effect estimates rather than relying only on null-hypothesis significance tests (Cumming, 2014). The analyses treated students as the observational units and did not model classroom-level variation. Because the instructional condition coincided with class membership and only one class represented each condition, a separate class effect could not be estimated. The resulting p values, confidence intervals, and effect sizes should therefore be interpreted cautiously as student-level comparisons conditional on these two participating classes rather than as estimates that isolate a generalizable treatment effect (Page et al., 2020; What Works Clearinghouse, 2017). Analyses were conducted in IBM SPSS Statistics; the software version was not retained in the available analytical record.

3. Results

3.1 Assumption Checks and Baseline Comparisons

Shapiro–Wilk tests were conducted to examine the distributions of the posttest scores within each group. As shown in Table 5, none of the tests was statistically significant. The results therefore provided no evidence of substantial departures from normality for either outcome in either group.

Table 5.
Shapiro–Wilk Tests for Posttest Scores

Outcome	Group	W	df	p
HOTS-oriented performance	Control	.969	36	.391
HOTS-oriented performance	Experimental	.961	36	.227
Learning interest	Control	.962	36	.249

Outcome	Group	<i>W</i>	<i>df</i>	<i>p</i>
Learning interest	Experimental	.974	36	.538

Levene's tests were used to examine equality of variance between the two groups. The tests were nonsignificant for HOTS-oriented performance, $F(1, 70) = 0.453$, $p = .503$, and learning interest, $F(1, 70) = 0.549$, $p = .461$. Consequently, the equal-variance assumption was not statistically contradicted for either posttest comparison.

Table 6.

Levene's Tests for Equality of Posttest Variances

Outcome	Levene statistic	<i>df1</i>	<i>df2</i>	<i>p</i>	Decision
HOTS-oriented performance	0.453	1	70	.503	Assumption not violated
Learning interest	0.549	1	70	.461	Assumption not violated

Independent-samples *t* tests were subsequently conducted to examine initial between-group differences. As presented in Table 7, the experimental group obtained a slightly higher HOTS-oriented pretest score ($M = 60.06$, $SD = 3.13$) than the control group ($M = 59.06$, $SD = 3.13$). The one-point difference was not statistically significant, $t(70) = 1.357$, $p = .179$, 95% CI $[-0.47, 2.47]$. Similarly, the experimental group's learning-interest pretest score ($M = 70.56$, $SD = 2.52$) was one point higher than that of the control group ($M = 69.56$, $SD = 2.52$), but the difference did not reach statistical significance, $t(70) = 1.681$, $p = .097$, 95% CI $[-0.19, 2.19]$. Thus, no statistically significant baseline differences were detected on either measured outcome. Nevertheless, these nonsignificant results should not be interpreted as formal proof that the nonequivalent groups were identical at baseline.

Table 7.

Baseline Between-Group Comparisons

Outcome	Experimental <i>M</i> (<i>SD</i>)	Control <i>M</i> (<i>SD</i>)	Mean difference	95% CI	<i>t</i> (70)	<i>p</i>
HOTS-oriented performance	60.06 (3.13)	59.06 (3.13)	1.00	$[-0.47, 2.47]$	1.357	.179
Learning interest	70.56 (2.52)	69.56 (2.52)	1.00	$[-0.19, 2.19]$	1.681	.097

Note. Mean differences were calculated as experimental minus control.

3.2 Posttest Outcomes

Table 8 presents the posttest comparisons and standardized between-group differences. For HOTS-oriented descriptive-text performance, the experimental group obtained a higher posttest mean ($M = 84.08$) than the control group ($M = 78.03$). The estimated mean difference was 6.06 points, 95% CI $[3.77, 8.35]$, and was statistically significant, $t(70) = 5.276$, $p < .001$. The corresponding standardized mean difference was large, Cohen's $d = 1.24$, 95% CI $[0.73, 1.75]$. A similar pattern was found for learning interest. The experimental group achieved a higher posttest mean ($M = 85.44$) than the control group ($M = 79.72$). The difference of 5.72 points was statistically significant, 95% CI $[3.53, 7.91]$, $t(70) = 5.217$, $p < .001$. The standardized difference was also large, Cohen's $d = 1.23$, 95% CI $[0.72, 1.73]$.

Table 8.

Between-Group Comparisons of Posttest Outcomes

Outcome	Experimental <i>M</i>	Control <i>M</i>	Mean difference	95% CI	<i>t</i> (70)	<i>p</i>	Cohen's <i>d</i> [95% CI]
HOTS-oriented performance	84.08	78.03	6.06	$[3.77, 8.35]$	5.276	$< .001$	1.24 $[0.73, 1.75]$
Learning interest	85.44	79.72	5.72	$[3.53, 7.91]$	5.217	$< .001$	1.23 $[0.72, 1.73]$

Note. Each group included 36 students. Positive differences and effect sizes favor the experimental group.

Students who received Kahoot!-supported descriptive-text instruction demonstrated significantly higher posttest HOTS-oriented performance and learning interest than students who received conventional instruction. The standardized differences were large for both outcomes. These findings represent unadjusted posttest comparisons, however, rather than direct statistical tests of differential improvement over time. Given the use of

intact, nonrandomized classes, they should therefore be interpreted as evidence of an association between the Kahoot!-supported condition and higher posttest outcomes rather than as definitive proof of an isolated causal effect.

4. Discussion

This study examined whether Kahoot!-supported descriptive-text instruction was associated with higher-order thinking skills (HOTS)-oriented performance and learning interest among vocational EFL students. The experimental group achieved significantly higher posttest scores than the control group in both HOTS-oriented performance and learning interest, with mean differences of 6.06 and 5.72 points, respectively. The corresponding effect sizes were large ($d = 1.24$ and $d = 1.23$). These findings are consistent in direction with previous reviews and meta-analyses showing that Kahoot!, gamification, and audience-response systems can support cognitive, motivational, and behavioral learning outcomes (Bai et al., 2020; Hunsu et al., 2016; Sailer & Homner, 2020; Wang & Tahir, 2020). However, the effects observed in this study were larger than the generally small-to-moderate average effects reported in broader syntheses, including the recent Kahoot!-specific meta-analysis by Özdemir (2025). The present effect sizes should therefore be regarded as substantial within this particular sample rather than assumed to represent the effect that would necessarily occur in other classrooms.

The higher HOTS-oriented performance of the experimental group is best explained by the instructional design surrounding Kahoot!, rather than by the platform itself. In the intervention, Kahoot! questions required students to analyze relationships among descriptive details, evaluate the relevance and effectiveness of language choices, justify interpretations, and create contextually appropriate descriptions. These processes were reinforced through explanatory feedback, discussion of alternative answers, and follow-up writing tasks. Research on critical-thinking instruction consistently indicates that higher-order thinking develops more effectively when cognitive processes are explicitly elicited, scaffolded, discussed, and applied to meaningful tasks rather than expected to emerge from content exposure alone (Abrami et al., 2015; Loyens et al., 2023; Tiruneh et al., 2018; Yin et al., 2023). Thus, Kahoot! appears to have functioned as a structured environment in which students repeatedly practiced the forms of reasoning assessed in the posttest.

This interpretation is particularly important in EFL education because the formal inclusion of HOTS in a curriculum or textbook does not guarantee cognitively demanding classroom practice. Previous studies have identified inconsistencies between teachers' recognition of HOTS and their actual instructional practices, as well as limitations in the depth with which HOTS is represented in English-language materials (Assaly & Jabarin, 2024; Erdiana & Panjaitan, 2023; Tyas et al., 2020). The present findings suggest that carefully designed Kahoot! activities can help operationalize abstract HOTS objectives by requiring students to perform observable acts of analysis, evaluation, justification, and creation. Nevertheless, the cognitive level of an activity is determined by what students are required to do with information, not merely by the presence of a C4, C5, or C6 label. Questions must therefore be designed around substantive reasoning rather than classified as higher order solely because they contain verbs such as analyze, evaluate, or create (Larsen et al., 2022; Stanny, 2016).

Immediate feedback may have provided another mechanism underlying the cognitive finding. Kahoot! enabled students and the teacher to see response patterns promptly, allowing misconceptions and competing interpretations to become visible while the questions were still cognitively active. However, feedback is most educationally useful when it is timely, understandable, linked to the task, and followed by an opportunity to act on the information rather than being limited to confirmation of whether an answer is correct (Dawson et al., 2019; Molin et al., 2021; Winstone et al., 2017; Zhang & Hyland, 2022). The teacher-led discussions and written follow-up activities in the present intervention may therefore have been more important than the display of correct answers itself. Processes in which learners respond, explain their reasoning, consider alternative answers, and revise their understanding have also been associated with deeper processing and improved learning in research on peer instruction and interactive response systems (Knight & Brame, 2018; Tullis & Goldstone, 2020; Vickrey et al., 2015).

The reasoning time provided during the Kahoot! activities may also have helped prevent the platform's competitive features from reducing complex questions to tests of response speed. HOTS items generally require students to examine information, compare alternatives, retrieve textual evidence, and formulate a defensible

response. Research on response time in gamified quizzes demonstrates that relationships between speed and accuracy are complex and can vary according to question difficulty, familiarity, motivation, and the type of cognitive processing required (Liang, 2024). Therefore, the use of Kahoot! for HOTS instruction requires sufficient processing time and should not automatically reward rapid answers as evidence of better reasoning. The present results are more plausibly associated with the combination of timed participation and subsequent opportunities for explanation than with speed-based competition alone.

The connection between the Kahoot! activities and written production provides a further explanation for the higher HOTS-oriented performance. Students did not finish the learning sequence after choosing an answer; they subsequently applied their analysis and evaluation when producing descriptive responses. This integration of thinking and writing is consistent with research demonstrating that technology-supported, inquiry-oriented, and explicitly structured instruction can improve critical thinking and EFL writing simultaneously (Ebadi & Rahimi, 2018; Simonovic et al., 2023; Wu et al., 2024; Yin et al., 2023). Writing requires learners to transform reasoning into linguistically explicit meaning, while the need to explain and justify ideas can expose weaknesses that remain hidden in recognition-based tasks. The follow-up writing activities may consequently have helped transfer students' responses from short-term quiz decisions into more elaborated descriptive-text performance.

This instructional alignment is especially relevant because descriptive writing can be demanding for Indonesian EFL learners. Students may struggle to select and organize relevant details, elaborate ideas, use appropriate vocabulary and grammatical features, and create a coherent representation of the object being described (Nuralisaputri & Megawati, 2023; Purnamasari et al., 2021; Ratnaningsih & Azizah, 2019). Moreover, increasing the cognitive demands of an L2 writing task can alter learners' planning, formulation, and linguistic performance, making the calibration of task complexity particularly important (Frear & Bitchener, 2015; Révész et al., 2017; Xu et al., 2022). In the present study, the progression from analyzing model descriptions to evaluating language choices and creating new descriptions may have distributed these demands across stages, making complex writing processes more manageable.

The findings also extend much of the existing Kahoot! literature. Kahoot! has frequently been used as a review, formative-assessment, or knowledge-recall tool, and previous studies have commonly examined general achievement, classroom participation, engagement, or students' perceptions of the platform (Cameron & Bizo, 2019; Owen & Licorish, 2020; Poblaciones et al., 2021; Ranieri et al., 2021). By assessing analysis, evaluation, and creation through a constructed-response test, the present study offers evidence that Kahoot!-supported instruction can be associated with outcomes beyond factual recognition. Nevertheless, this finding should not be interpreted as evidence that conventional Kahoot! quizzes automatically measure or develop HOTS. The questions, teacher mediation, explanatory feedback, and productive follow-up tasks differentiated this intervention from the simple use of Kahoot! for recalling isolated information.

The second major finding was the higher learning-interest score of the experimental group. Kahoot!'s visual presentation, simultaneous responding, immediate feedback, points, progress displays, and structured competition may have created a more attentive and enjoyable learning environment. Previous studies have similarly associated game-based response systems with increased attention, enjoyment, participation, motivation, and classroom interaction (Cameron & Bizo, 2019; Licorish et al., 2018; Morillas Barrio et al., 2016; Turan & Meral, 2018). Specific game-design elements can influence learners' feelings of competence, social connection, and involvement, but their effects depend on how those elements are combined with the learning activity (Sailer et al., 2017). Thus, the experimental group's higher interest was unlikely to have resulted from points and competition in isolation. The relevance of the questions, opportunities to participate, immediate information about performance, teacher explanation, and written application probably contributed collectively to the observed affective outcome.

The learning-interest result must nevertheless be interpreted in relation to the distinction between triggering and maintaining interest. Novelty, visual stimulation, competition, and immediate feedback can trigger situational interest, but this initial response does not necessarily develop into sustained interest unless students also perceive the content as meaningful, manageable, and personally relevant (Harackiewicz et al., 2016; Knogler et al., 2015; Murayama et al., 2019; Renninger et al., 2019). Technology-supported language activities may initially stimulate interest, but maintaining that interest requires continued task value and meaningful interaction rather than

repeated exposure to the same digital feature (Fryer et al., 2017). Because learning interest was measured immediately after the intervention, the present study demonstrates a short-term post-intervention difference but cannot determine whether students' interest remained higher after the novelty of Kahoot! diminished. Delayed measurement would be necessary to distinguish temporary situational interest from more stable interest in English learning.

The simultaneous findings for HOTS-oriented performance and learning interest suggest that the cognitive and affective outcomes may have operated complementarily. Higher-order descriptive tasks require students to maintain attention, invest effort, compare alternatives, and persist through uncertainty. Interest can support these processes by directing attention toward the learning content, promoting continued effort, and increasing willingness to engage deeply with a task (Guo & An, 2025; McIntyre et al., 2021; Nuutila et al., 2021; Rieger et al., 2022). In foreign-language learning, engagement is similarly understood as involving interconnected behavioral, cognitive, emotional, and social processes rather than participation alone (Oga-Baldwin, 2019; Philp & Duchesne, 2016). Kahoot! may therefore have created affective conditions that encouraged students to remain involved in cognitively demanding tasks, while successful participation and visible progress may, in turn, have reinforced their interest. This explanation remains theoretical, however, because the study did not test correlations or mediation between learning interest and HOTS. The results establish parallel between-group differences, not evidence that improvement in one outcome caused improvement in the other.

The vocational-school context further strengthens the pedagogical relevance of the findings. Research in vocational English education indicates that students' engagement is influenced by whether learning activities are connected to their programs, future occupational practices, and perceived communication needs (Avci & Engin-Demir, 2021; Hajar & Triastuti, 2021; Natsir et al., 2022; Skarpaas, 2024). A vocationally oriented approach can make English learning more accessible and purposeful when academic language objectives are connected to recognizable professional situations (Skarpaas & Hellekjær, 2021). This is important because workplace communication involves more than mastery of general linguistic forms; learners must interpret context, select relevant information, adapt messages to particular audiences, and produce texts suited to occupational purposes (Al Hilali & McKinley, 2021; Kamil & Muhammad, 2021; Ramamurthy et al., 2021). Kahoot!-supported descriptive instruction may therefore become more meaningful when prompts require students to describe products, tools, services, workplaces, tourism resources, or procedures connected to their vocational specialization.

Pedagogically, these findings indicate that Kahoot! should be treated as one component of an instructional sequence rather than as an inherently effective intervention. Teachers seeking to develop HOTS should use questions that require students to differentiate relevant details, compare descriptions, identify relationships, evaluate evidence, justify judgments, and construct original responses. Students should receive adequate reasoning time, explanatory feedback, opportunities to discuss alternative interpretations, and follow-up writing tasks through which they apply the reasoning elicited during the quiz. Competition should remain supportive because excessive emphasis on ranking and speed may discourage students who require longer processing time. This qualification is consistent with critical reviews showing that the effectiveness of educational gamification is shaped by instructional design, learner characteristics, content, implementation conditions, and the specific combination of game elements rather than by gamification alone (Dichev & Dicheva, 2017; Ranieri et al., 2021).

Several limitations qualify the interpretation of the results. First, the study involved only two intact classes, with one class representing each condition. Class membership was therefore confounded with instructional condition, and unmeasured differences in classroom climate, peer relationships, teacher–student interaction, or prior experience with digital learning may have affected the posttest scores. Second, the independent-samples *t* tests compared the groups separately at pretest and posttest but did not directly test whether the two groups experienced significantly different rates of change. Future studies should involve multiple classes in each condition and employ ANCOVA or mixed-effects models to adjust for baseline scores and account for the clustering of students within classes.

Third, learning interest was measured through self-report immediately after the intervention. The scores may consequently reflect students' perceptions of novelty or enjoyment rather than sustained behavioral engagement. Future research should combine questionnaires with classroom observations, participation records, interviews,

and delayed measures of interest. Fourth, Cronbach's alpha for the constructed-response test indicated consistency among item scores but did not establish consistency between scorers. The absence of an inter-rater reliability coefficient limits confidence in the reproducibility of the essay scoring. Subsequent studies should employ at least two trained scorers and report an appropriate agreement coefficient. Finally, the relatively narrow score dispersion contributed to the large standardized effects. Replication across schools, vocational specializations, teachers, and longer intervention periods is needed before the magnitude of these differences can be generalized.

Vocational EFL students who received Kahoot!-supported descriptive-text instruction achieved higher posttest HOTS-oriented performance and learning interest than students receiving conventional instruction. The findings support the use of Kahoot! as a structured pedagogical scaffold when it is combined with cognitively demanding questions, sufficient reasoning time, explanatory feedback, discussion, and written application. The study's contribution therefore lies not merely in showing that students responded positively to a gamified platform, but in demonstrating how a game-based response system can be incorporated into an instructional sequence that connects affective engagement with analysis, evaluation, and creation in descriptive-text learning. Given the nonequivalent intact-class design, these findings should be regarded as strong context-specific evidence of an association rather than definitive proof that Kahoot! alone caused the observed outcomes.

4. Conclusion

This study examined HOTS-oriented descriptive-text performance and learning interest among vocational EFL students receiving Kahoot!-supported or conventional instruction. Students in the Kahoot!-supported group achieved significantly higher posttest scores for both HOTS-oriented performance and learning interest, with mean differences of 6.06 and 5.72 points, respectively. The large standardized differences ($d = 1.24$ and $d = 1.23$) indicate that the observed advantages were educationally meaningful within this sample. The findings suggest that Kahoot! can support cognitive and affective learning when it is integrated into a purposeful instructional sequence involving cognitively demanding questions, adequate reasoning time, explanatory feedback, classroom discussion, and follow-up writing. Its pedagogical value therefore extends beyond entertainment or factual review: it can help students analyze, evaluate, and create descriptive texts while sustaining their interest in learning. Teachers should consequently treat Kahoot! as an instructional scaffold rather than as an inherently effective technology. These conclusions remain subject to the study's nonequivalent intact-class design, limited sample, immediate self-reported measurement of learning interest, and absence of reported inter-rater reliability for the constructed-response test. The results therefore provide strong context-specific evidence of an association between Kahoot!-supported instruction and higher posttest outcomes, but not definitive proof that Kahoot! alone caused the differences. Future studies should involve multiple classes and schools, apply analyses that adjust directly for baseline scores and classroom clustering, include delayed assessments and behavioral measures of engagement, and establish scorer agreement to determine whether these outcomes are sustained and generalizable.

References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, 85(2), 275–314. <https://doi.org/10.3102/0034654314551063>
- Adams, N. E. (2015). Bloom's taxonomy of cognitive learning objectives. *Journal of the Medical Library Association*, 103(3), 152–153. <https://doi.org/10.3163/1536-5050.103.3.010>
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959. <https://doi.org/10.1177/001316448004000419>
- Al Hilali, T. S., & McKinley, J. (2021). Exploring the socio-contextual nature of workplace writing: Towards preparing learners for the complexities of English L2 writing in the workplace. *English for Specific Purposes*, 63, 86–97. <https://doi.org/10.1016/j.esp.2021.03.003>
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

- Assaly, I., & Jabarin, A. (2024). Arab Israeli EFL teachers' perceptions and practices vis-à-vis teaching higher-order thinking skills: A complicated relationship. *Language Teaching Research*, 28(4), 1635–1655. <https://doi.org/10.1177/13621688211032426>
- Avci, N., & Engin-Demir, C. (2021). Promoting the vocational English language competences of students in an English for specific purposes context: A needs assessment. *Psycho-Educational Research Reviews*, 10(3), 125–139. https://doi.org/10.52963/PERR_Biruni_V10.N3.08
- Bai, S. T., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, 30, Article 100322. <https://doi.org/10.1016/j.edurev.2020.100322>
- Bernacki, M. L., & Walkington, C. (2018). The role of situational interest in personalized learning. *Journal of Educational Psychology*, 110(6), 864–881. <https://doi.org/10.1037/edu0000250>
- Braun, H. I., Shavelson, R. J., Zlatkin-Troitschanskaia, O., & Borowiec, K. (2020). Performance assessment of critical thinking: Conceptualization, design, and implementation. *Frontiers in Education*, 5, Article 156. <https://doi.org/10.3389/feduc.2020.00156>
- Çal, A., Admiraal, W., & Mearns, T. (2022). The what-how-why of English in the workplace: Perspectives from Turkish engineers. *European Journal of Engineering Education*, 47(2), 333–352. <https://doi.org/10.1080/03043797.2021.1946014>
- Cameron, K. E., & Bizo, L. A. (2019). Use of the game-based learning platform Kahoot! to facilitate learner engagement in animal science students. *Research in Learning Technology*, 27, Article 2225. <https://doi.org/10.25304/rlt.v27.2225>
- Chan, C. S. C. (2019). Long-term workplace communication needs of business professionals: Stories from Hong Kong senior executives and their implications for ESP and higher education. *English for Specific Purposes*, 56, 68–83. <https://doi.org/10.1016/j.esp.2019.07.003>
- Chan, C. S. C. (2021). Helping university students discover their workplace communication needs: An eclectic and interdisciplinary approach to facilitating on-the-job learning of workplace communication. *English for Specific Purposes*, 64, 55–71. <https://doi.org/10.1016/j.esp.2021.07.002>
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7–29. <https://doi.org/10.1177/0956797613504966>
- Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education*, 44(1), 25–36. <https://doi.org/10.1080/02602938.2018.1467877>
- Denny, M., Denieffe, S., & O'Sullivan, K. (2023). Non-equivalent control group pretest–posttest design in social and behavioral research. In A. L. Nichols & J. E. Edlund (Eds.), *The Cambridge handbook of research methods and statistics for the social and behavioral sciences* (Vol. 1, pp. 314–332). Cambridge University Press. <https://doi.org/10.1017/9781009010054.016>
- Derakhshan, A., Kruk, M., & Mehdizadeh, M. (2021). Boredom in online classes in the Iranian EFL context: Sources and solutions. *System*, 101, Article 102556. <https://doi.org/10.1016/j.system.2021.102556>
- Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain—A critical review. *International Journal of Educational Technology in Higher Education*, 14, Article 9. <https://doi.org/10.1186/s41239-017-0042-5>
- Driscoll, D. L., Paszek, J., Gorzelsky, G., Hayes, C. L., & Jones, E. (2020). Genre knowledge and writing development: Results from the Writing Transfer Project. *Written Communication*, 37(1), 69–103. <https://doi.org/10.1177/0741088319882313>
- Ebadi, S., & Rahimi, M. (2018). An exploration into the impact of WebQuest-based classroom on EFL learners' critical thinking and academic writing skills: A mixed-methods study. *Computer Assisted Language Learning*, 31(5–6), 617–651. <https://doi.org/10.1080/09588221.2018.1449757>
- Erdiana, N., & Panjaitan, S. (2023). How is HOTS integrated into the Indonesian high school English textbook? *Studies in English Language and Education*, 10(1), 60–77. <https://doi.org/10.24815/siele.v10i1.26052>

- Frear, M. W., & Bitchener, J. (2015). The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing*, 30, 45–57. <https://doi.org/10.1016/j.jslw.2015.08.009>
- Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, 75, 461–468. <https://doi.org/10.1016/j.chb.2017.05.045>
- Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science*, 24(3), 343–360. <https://doi.org/10.1214/09-STS301>
- Gopalan, M., Rosinger, K., & Ahn, J. B. (2020). Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education*, 44(1), 218–243. <https://doi.org/10.3102/0091732X20903302>
- Guo, J., & An, F. (2025). Exploring the categories of students' interest and their relationships with deep learning in technology-supported environments. *Scientific Reports*, 15, Article 10370. <https://doi.org/10.1038/s41598-025-95294-2>
- Haataja, E. S. H., Tolvanen, A., Vilppu, H., Kallio, M., Peltonen, J., & Metsäpelto, R.-L. (2023). Measuring higher-order cognitive skills with multiple-choice questions: Potentials and pitfalls of Finnish teacher-education entrance. *Teaching and Teacher Education*, 122, Article 103943. <https://doi.org/10.1016/j.tate.2022.103943>
- Hajar, I., & Triastuti, A. (2021). A need analysis of ESP materials for agriculture students at Indonesian vocational school in West Sulawesi. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 6(1), 13–26. <https://doi.org/10.21093/ijeltal.v6i1.840>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy Insights from the Behavioral and Brain Sciences*, 3(2), 220–227. <https://doi.org/10.1177/2372732216655542>
- Henry, A., & Thorsen, C. (2020). Disaffection and agentic engagement: “Redesigning” activities to enable authentic self-expression. *Language Teaching Research*, 24(4), 456–475. <https://doi.org/10.1177/1362168818795976>
- Hoang, T. N., Tran, D. T., Hoang, G. T. H., & Tran, T. T. T. (2024). English communication needs of professional engineers and the relevance of English language training in higher education. *Taiwan Journal of TESOL*, 21(2), 33–64. [https://doi.org/10.30397/TJTESOL.202410_21\(2\).0002](https://doi.org/10.30397/TJTESOL.202410_21(2).0002)
- Huber, C. R., & Kuncel, N. R. (2016). Does college teach critical thinking? A meta-analysis. *Review of Educational Research*, 86(2), 431–468. <https://doi.org/10.3102/0034654315605917>
- Hunsu, N. J., Adesope, O. O., & Bayly, D. J. (2016). A meta-analysis of the effects of audience-response systems on cognition and affect. *Computers & Education*, 94, 102–119. <https://doi.org/10.1016/j.compedu.2015.11.013>
- Ismail, M. A.-A., Ahmad, A., Mohammad, J. A.-M., Fakri, N. M. R. M., Nor, M. Z. M., & Pa, M. N. M. (2019). Using Kahoot! as a formative-assessment tool in medical education: A phenomenological study. *BMC Medical Education*, 19, Article 230. <https://doi.org/10.1186/s12909-019-1658-z>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Kamil, M. A., & Muhammad, A. M. (2021). Professional communication competence in English for occupational purposes courses: A systematic literature review and proposal of a framework. *Journal of Teaching and Learning for Graduate Employability*, 12(2), 260–278. <https://doi.org/10.21153/jtlge2021vol12no2art1075>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>
- Kartika-Ningsih, H., & Gunawan, W. (2019). Recontextualisation of genre-based pedagogy: The case of Indonesian EFL classrooms. *Indonesian Journal of Applied Linguistics*, 9(2), 335–347. <https://doi.org/10.17509/ijal.v9i2.20231>
- Kim, S.-H. (2021). Analyzing workplace English communication use and needs: Insights from Korean corporate employees. *Journal of Pan-Pacific Association of Applied Linguistics*, 25(1), 1–24. <https://doi.org/10.25256/PAAL.25.1.1>

- Knight, J. K., & Brame, C. J. (2018). Peer instruction. *CBE—Life Sciences Education*, 17(2), Article fe5. <https://doi.org/10.1187/cbe.18-02-0025>
- Knogler, M., Harackiewicz, J. M., Gegenfurtner, A., & Lewalter, D. (2015). How situational is situational interest? Investigating the longitudinal structure of situational interest. *Contemporary Educational Psychology*, 43, 39–50. <https://doi.org/10.1016/j.cedpsych.2015.08.004>
- Kruk, M., Pawlak, M., & Zawodniak, J. (2021). Another look at boredom in language instruction: The role of the predictable and the unexpected. *Studies in Second Language Learning and Teaching*, 11(1), 15–40. <https://doi.org/10.14746/ssllt.2021.11.1.2>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t* tests and ANOVAs. *Frontiers in Psychology*, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362. <https://doi.org/10.1177/1948550617697177>
- Larsen, T. M., Endo, B. H., Yee, A. T., Do, T., & Lo, S. M. (2022). Probing internal assumptions of the revised Bloom's taxonomy. *CBE—Life Sciences Education*, 21(4), Article ar66. <https://doi.org/10.1187/cbe.20-08-0170>
- Lee, I., & Mak, P. (2018). Metacognition and metacognitive instruction in second-language writing classrooms. *TESOL Quarterly*, 52(4), 1085–1097. <https://doi.org/10.1002/tesq.436>
- Lenchuk, I., & Ahmed, A. (2021). Tapping into Bloom taxonomy's higher-order cognitive processes: The case for multiple-choice questions as a valid assessment tool in the ESP classroom. *Arab World English Journal, Special Issue on COVID-19 Challenges*, 1, 160–171. <https://doi.org/10.24093/awej/covid.12>
- Liang, Z. (2024). More haste, less speed? Relationship between response time and response accuracy in gamified online quizzes in an undergraduate engineering course. *Frontiers in Education*, 9, Article 1412954. <https://doi.org/10.3389/feduc.2024.1412954>
- Licorish, S. A., Owen, H. E., Daniel, B., & George, J. L. (2018). Students' perceptions of Kahoot!'s influence on teaching and learning. *Research and Practice in Technology Enhanced Learning*, 13, Article 9. <https://doi.org/10.1186/s41039-018-0078-8>
- Liu, C., & Chen, M. (2022). A genre-based approach in the secondary-school English writing class: Voices from student-teachers in the teaching practicum. *Frontiers in Psychology*, 13, 992360. <https://doi.org/10.3389/fpsyg.2022.992360>
- Loyens, S. M. M., van Meerten, J. E., Schaap, L., & Wijnia, L. (2023). Situating higher-order, critical, and critical-analytic thinking in problem- and project-based learning environments: A systematic review. *Educational Psychology Review*, 35(2), Article 39. <https://doi.org/10.1007/s10648-023-09757-x>
- Lüdtke, O., & Robitzsch, A. (2025). ANCOVA versus change score for the analysis of two-wave data. *The Journal of Experimental Education*, 93(2), 363–395. <https://doi.org/10.1080/00220973.2023.2246187>
- Maryani, I., Prasetyo, Z. K., Wilujeng, I., Purwanti, S., & Fitriawanati, M. (2021). HOTS multiple-choice and essay questions: A validated instrument to measure higher-order thinking skills of prospective teachers. *Journal of Turkish Science Education*, 18(4), 674–690. <https://doi.org/10.36681/tused.2021.97>
- McIntyre, M. M., Gundlach, J. L., & Graziano, W. G. (2021). Liking guides learning: The role of interest in memory for STEM topics. *Learning and Individual Differences*, 85, Article 101960. <https://doi.org/10.1016/j.lindif.2020.101960>
- Molin, F., Haelermans, C., Cabus, S. J., & Groot, W. (2021). Do feedback strategies improve students' learning gain? Results of a randomized experiment using polling technology in physics classrooms. *Computers & Education*, 175, Article 104339. <https://doi.org/10.1016/j.compedu.2021.104339>
- Moore, T., & Morton, J. (2017). The myth of job readiness? Written communication, employability, and the “skills gap” in higher education. *Studies in Higher Education*, 42(3), 591–609. <https://doi.org/10.1080/03075079.2015.1067602>
- Morillas Barrio, C., Muñoz-Organero, M., & Sánchez Soriano, J. (2016). Can gamification improve the benefits of student-response systems in learning? An experimental study. *IEEE Transactions on Emerging Topics in Computing*, 4(3), 429–438. <https://doi.org/10.1109/TETC.2015.2497459>

- Murayama, K., FitzGibbon, L., & Sakaki, M. (2019). Process account of curiosity and interest: A reward-learning perspective. *Educational Psychology Review*, 31(4), 875–895. <https://doi.org/10.1007/s10648-019-09499-9>
- Nadjib, N., & Triastuti, A. (2023). Indonesian EFL teachers' portrayal of genre-based teaching practices: Perceptions, implementation, and challenges. *Journal on English as a Foreign Language*, 13(2), 475–498. <https://doi.org/10.23971/jeft.v13i2.6501>
- Natsir, M., Purba, A. S., Ellyana, E., Saragih, A. T., & Amal, B. K. (2022). English teaching in an Indonesian vocational high school majoring in industrial machinery engineering. *AL-ISHLAH: Jurnal Pendidikan*, 14(2), 1743–1754. <https://doi.org/10.35445/alishlah.v14i2.1221>
- Negretti, R., & McGrath, L. (2018). Scaffolding genre knowledge and metacognition: Insights from an L2 doctoral research-writing course. *Journal of Second Language Writing*, 40, 12–31. <https://doi.org/10.1016/j.jslw.2017.12.002>
- Nuralisaputri, V. V., & Megawati, F. (2023). Writing descriptive texts: What makes this difficult for lower-secondary-school learners? *Indonesian TESOL Journal*, 5(2), 120–133. <https://doi.org/10.24256/itj.v5i2.2718>
- Nuutila, K., Tapola, A., Tuominen, H., Molnár, G., & Niemivirta, M. (2021). Mutual relationships between the levels of and changes in interest, self-efficacy, and perceived difficulty during task engagement. *Learning and Individual Differences*, 92, Article 102090. <https://doi.org/10.1016/j.lindif.2021.102090>
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, 78(1), 33–84. <https://doi.org/10.3102/0034654307313793>
- Oga-Baldwin, W. L. Q. (2019). Acting, thinking, feeling, making, collaborating: The engagement process in foreign language learning. *System*, 86, Article 102128. <https://doi.org/10.1016/j.system.2019.102128>
- Oga-Baldwin, W. L. Q., & Nakata, Y. (2017). Engagement, gender, and motivation: A predictive model for Japanese young language learners. *System*, 65, 151–163. <https://doi.org/10.1016/j.system.2017.01.011>
- Oga-Baldwin, W. L. Q., Nakata, Y., Parker, P. D., & Ryan, R. M. (2017). Motivating young language learners: A longitudinal model of self-determined motivation in elementary-school foreign-language classes. *Contemporary Educational Psychology*, 49, 140–150. <https://doi.org/10.1016/j.cedpsych.2017.01.010>
- Owen, H. E., & Licorish, S. A. (2020). Game-based student-response system: The effectiveness of Kahoot! on junior and senior information-science students' learning. *Journal of Information Technology Education: Research*, 19, 511–553. <https://doi.org/10.28945/4608>
- Özdemir, O. (2025). Kahoot! Game-based digital learning platform: A comprehensive meta-analysis. *Journal of Computer Assisted Learning*, 41(1), Article e13084. <https://doi.org/10.1111/jcal.13084>
- Page, L. C., Lenard, M. A., & Keele, L. (2020). The design of clustered observational studies in education. *AERA Open*, 6(3), 1–26. <https://doi.org/10.1177/2332858420954401>
- Pawlak, M., Kruk, M., Zawodniak, J., & Pasikowski, S. (2020). Investigating factors responsible for boredom in English classes: The case of advanced learners. *System*, 91, Article 102259. <https://doi.org/10.1016/j.system.2020.102259>
- Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50–72. <https://doi.org/10.1017/S0267190515000094>
- Poblaciones, M. J., Garcia-White, T., & Marin, C. (2021). Students' perceptions of real-time quiz Kahoot! as a review tool in higher education: A case study. *International Journal of Engineering Pedagogy*, 11(4), 165–174. <https://doi.org/10.3991/ijep.v11i4.21359>
- Purnamasari, D., Hidayat, D. N., & Kurniawati, L. (2021). An analysis of students' writing skill on English descriptive text. *English Education: Jurnal Tadris Bahasa Inggris*, 14(1), 101–114. <https://doi.org/10.24042/ee-jtbi.v14i1.7943>
- Ramamurthy, V., Alias, N., & DeWitt, D. (2021). The need for technical communication for 21st century learning in TVET institutions: Perceptions of industry experts. *Journal of Technical Education and Training*, 13(1), 148–158. <https://doi.org/10.30880/jtet.2021.13.01.017>

- Ranieri, M., Raffaghelli, J. E., & Bruni, I. (2021). Game-based student-response system: Revisiting its potentials and criticalities in large-size classes. *Active Learning in Higher Education*, 22(2), 129–142. <https://doi.org/10.1177/1469787418812667>
- Ratnaningsih, S., & Azizah, A. (2019). Error analysis in the descriptive-text writing of vocational-high-school students. *Dinamika Ilmu*, 19(1), 175–186. <https://doi.org/10.21093/di.v19i1.1364>
- Renninger, K. A., & Hidi, S. E. (2020). To level the playing field, develop interest. *Policy Insights from the Behavioral and Brain Sciences*, 7(1), 10–18. <https://doi.org/10.1177/2372732219864705>
- Renninger, K. A., Bachrach, J. E., & Hidi, S. E. (2019). Triggering and maintaining interest in early phases of interest development. *Learning, Culture and Social Interaction*, 23, Article 100260. <https://doi.org/10.1016/j.lcsi.2018.11.007>
- Révész, A., Kourtali, N.-E., & Mazgutova, D. (2017). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning*, 67(1), 208–241. <https://doi.org/10.1111/lang.12205>
- Rieger, S., Göllner, R., Spengler, M., Trautwein, U., Nagengast, B., & Roberts, B. W. (2022). The persistence of students' academic effort: The unique and combined effects of conscientiousness and individual interest. *Learning and Instruction*, 80, Article 101613. <https://doi.org/10.1016/j.learninstruc.2022.101613>
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32, 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game-design elements on psychological-need satisfaction. *Computers in Human Behavior*, 69, 371–380. <https://doi.org/10.1016/j.chb.2016.12.033>
- Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research, and Evaluation*, 22(1), Article 4. <https://doi.org/10.7275/ca7y-mm27>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Simonovic, B., Correa Vione, K., Stupple, E., & Doherty, A. (2023). It is not what you think it is how you think: A critical thinking intervention enhances argumentation, analytic thinking and metacognitive sensitivity. *Thinking Skills and Creativity*, 49, 101362. <https://doi.org/10.1016/j.tsc.2023.101362>
- Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107. <https://doi.org/10.7334/psicothema2013.256>
- Skarpaas, K. G. (2024). The importance of relevance: Factors influencing upper-secondary vocational students' engagement in L2 English-language teaching. *Scandinavian Journal of Educational Research*, 68(7), 1395–1409. <https://doi.org/10.1080/00313831.2023.2250360>
- Skarpaas, K. G., & Hellekjær, G. O. (2021). Vocational orientation—A supportive approach to teaching L2 English in upper-secondary-school vocational programmes. *International Journal of Educational Research Open*, 2, 100064. <https://doi.org/10.1016/j.ijedro.2021.100064>
- Stanny, C. J. (2016). Reevaluating Bloom's taxonomy: What measurable verbs can and cannot say about student learning. *Education Sciences*, 6(4), Article 37. <https://doi.org/10.3390/educsci6040037>
- Suharyadi, & Basthomi, Y. (2020). Patterns of the teaching and learning cycle of GBA by EFL teachers in Indonesia. *Journal of Education and E-Learning Research*, 7(1), 34–41. <https://doi.org/10.20448/journal.509.2020.71.34.41>
- Sureeyatanapas, P., Srisawasdi, N., & Sureeyatanapas, P. (2024). A need analysis of English proficiency in engineering graduates: Perspectives of the world's leading companies in Thailand. *European Journal of Engineering Education*, 49(6), 1179–1202. <https://doi.org/10.1080/03043797.2024.2377314>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tardy, C. M., Sommer-Farias, B., & Gevers, J. (2020). Teaching and researching genre knowledge: Toward an enhanced theoretical framework. *Written Communication*, 37(3), 287–321. <https://doi.org/10.1177/0741088320916554>

- Tiruneh, D. T., De Cock, M., & Elen, J. (2018). Designing learning environments for critical thinking: Examining effective instructional approaches. *International Journal of Science and Mathematics Education*, 16(6), 1065–1089. <https://doi.org/10.1007/s10763-017-9829-z>
- Tullis, J. G., & Goldstone, R. L. (2020). Why does peer instruction benefit student learning? *Cognitive Research: Principles and Implications*, 5, Article 15. <https://doi.org/10.1186/s41235-020-00218-5>
- Turan, Z., & Meral, E. (2018). Game-based versus non-game-based: The impact of student-response systems on students' achievement, engagement, and test anxiety. *Informatics in Education*, 17(1), 105–116. <https://doi.org/10.15388/infedu.2018.07>
- Tyas, M. A., Nurkamto, J., & Marmanto, S. (2020). Cultivating students' higher-order thinking skills in EFL classes: The role of the teacher and the textbook. *International Online Journal of Education and Teaching*, 7(1), 267–276. <https://iojet.org/index.php/IOJET/article/view/773>
- Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015). Research-based implementation of peer instruction: A literature review. *CBE—Life Sciences Education*, 14(1), Article es3. <https://doi.org/10.1187/cbe.14-11-0198>
- Wang, A. I., & Tahir, R. (2020). The effect of using Kahoot! for learning: A literature review. *Computers & Education*, 149, Article 103818. <https://doi.org/10.1016/j.compedu.2020.103818>
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “ $p < .05$.” *The American Statistician*, 73(sup1), 1–19. <https://doi.org/10.1080/00031305.2019.1583913>
- What Works Clearinghouse. (2017). *WWC standards brief: Confounding factors*. Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_brief_confounds_101117.pdf
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17–37. <https://doi.org/10.1080/00461520.2016.1207538>
- Wu, T.-T., Silitonga, L. M., & Murti, A. T. (2024). Enhancing English writing and higher-order thinking skills through computational thinking. *Computers & Education*, 213, 105012. <https://doi.org/10.1016/j.compedu.2024.105012>
- Xu, T. S., Zhang, L. J., & Gaffney, J. S. (2022). Examining the relative effects of task complexity and cognitive demands on students' writing in a second language. *Studies in Second Language Acquisition*, 44(2), 483–506. <https://doi.org/10.1017/S0272263121000310>
- Yin, X., Saad, M. R. B. M., & Halim, H. B. A. (2023). A systematic review of critical-thinking instructional pedagogies in EFL writing: What do we know from a decade of research? *Thinking Skills and Creativity*, 49, Article 101363. <https://doi.org/10.1016/j.tsc.2023.101363>
- Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51, Article 100586. <https://doi.org/10.1016/j.asw.2021.100586>