https://doi.org/10.17323/jle.2026.29815

Guided Inquiry with Educational Videos: Supporting Retention and Invested Cognitive Effort in EFL Higher Education

Hamza Hassoun, Naima Trimasse

Ibnou Zohr University, Agadir, Souss-Massa, Morocco

ABSTRACT

Background. While videos increasingly support flexible and student-centred learning in higher education, their pedagogical potential is often limited in practice. Transient information effects and insufficient emphasis on generative processing constrain active cognitive engagement. Empirical evidence evaluating how specific pedagogical frameworks, such as the 5E guided inquiry model, impact cognitive load dimensions in EFL higher education remains limited.

Purpose. To investigate how guided inquiry can foster generative processing in video-based learning and help address the transient information effect in terms of learning outcomes and cognitive load.

Method. A pre/post-test quasi-experiment was conducted with 135 English department students, divided into a control group (n=65) and a treatment group (n=70). The participants were first-year EFL students who were novices in the subject domain and reported varied English proficiency, most commonly intermediate or upper-intermediate. Both groups studied the same content across eight sessions. The control group used educational videos and direct instruction, while the treatment group used educational videos combined with the 5E model of guided inquiry. Cognitive load was measured across essential, extraneous, and generative processing, while pre- and post-tests assessed retention and transfer.

Results. Findings demonstrate that the treatment group achieved significantly higher retention ($p < .001$, $r = .414$). Due to a ceiling effect, transfer outcomes were deemed uninterpretable rather than evidence of comparability. While the aggregate cognitive-load subscales did not show between-group significance, results demonstrated a localised effect on invested cognitive effort, with one item remaining significant after Holm-Bonferroni correction ($p < .001$, $r = .326$). Within this context, the findings suggest that the initial increase in extraneous processing was not sustained over time and appeared to diminish as students adapted to the guided inquiry structure.

Conclusion. This study applies multi-dimensional cognitive load measurements to a 5E guided inquiry intervention within an EFL higher education context. The results indicate that the original hypothesis was supported for retention and partially supported, at the item level, for invested cognitive effort. These findings offer a context-bound perspective for curriculum designers, suggesting that structured inquiry prompts can support foundational knowledge consolidation and specific facets of engagement in comparable EFL multimedia learning environments.

KEYWORDS

guided inquiry; educational videos; transient information effect; cognitive load; generative processing; EFL higher education

Citation: Hassoun, H., & Trimasse, N. (2026). Guided Inquiry with educational videos: supporting retention and invested cognitive effort in EFL higher education. *Journal of Language and Education, 12*(1), 50-72. https://doi.org/10.17323/jle.2026.29815

Correspondence:
Hamza Hassoun,
hamza.hassoun.65@edu.uiz.ac.ma

Received: November 20, 2025
Accepted: March 16, 2026
Published: March 31, 2026

INTRODUCTION

Videos have become integral to both formal and informal learning (Pires et al., 2022). With AI enabling the rapid creation of instructional content, videos have transformed how learners access and engage with information (Xu et al., 2024). Modern videos support more flexible and student-centred learning that reflects students' digital engagement practices (Hassoun et al., 2023; Kerimbayev et al., 2023). They offer rich visual and auditory input, allowing complex concepts to be conveyed efficiently and vividly (Mayer, 2024). Importantly, they are easily accessed and attuned to today's learners, who increasingly prefer concise, visually supported, and on-demand access to information (Pellas, 2025). However, realising the pedagogical potential of videos depends not merely on their availability, but on their implementation (Fiorella, 2021). Decades of research on multimedia have clarified that effective instruction must respect human cognitive architecture; it should minimise extraneous processing, manage essential processing, and foster generative processing (Mayer, 2021a, 2024).

A central difficulty with video formats is the transient information effect, which occurs when fleeting audiovisual input overloads working memory by limiting opportunities for adequate processing (Jiang & Sweller, 2021). Because videos continuously replace visual and auditory stimuli, learners must simultaneously process new information while retaining prior information. This often increases extraneous cognitive processing when content is complex or unfamiliar (Castro-Alonso et al., 2018).

Moderating factors such as prior knowledge, expertise, and language proficiency heavily influence how learners handle transient information (Wong et al., 2018). For example, while subtitles can cause split-attention for native speakers, they reduce cognitive load for English as a Foreign Language (EFL) learners by providing dual cues for comprehension (Ritzhaupt et al., 2015; Pannatier & Bétrancourt, 2024). These findings illustrate that cognitive design principles must be contextualised rather than universally applied. This limitation is accentuated by contemporary viewing habits that favour immediacy and passive consumption (Hassoun et al., 2023). Habitual exposure to fast-paced content has been linked to reduced attentional endurance and heightened cognitive strain (Alruwaili, 2025). Consequently, students may struggle to maintain focus or experience diminished motivation when instructional content demands extended concentration or reflective processing.

Research indicates that videos must be pedagogically structured to counteract these attentional tendencies by embedding inquiry, reflection, and elaboration tasks that sustain engagement and promote generative learning (Beege & Ploetzner, 2025; Ruf et al., 2023). Such integration aligns with constructivist and cognitive principles, which view learning as an active process of meaning-making. Generative processing has been redefined as the central mechanism distinguishing surface comprehension from deep learning. Consequently, Fiorella (2023) emphasises that research should advance beyond optimising multimedia design principles to examining the specific pedagogical conditions that actively stimulate generative engagement.

Guided inquiry learning offers a viable structure for fostering such engagement (de Jong, 2021). The efficacy of guided inquiry in fostering generative processing can be theoretically understood through its activation of metacognitive mechanisms, which provide the essential cognitive foundation for meaningful learning. While the Cognitive Theory of Multimedia Learning (CTML) establishes that generative processing involves the learner's effort to organise and integrate new information with prior knowledge, its scope is primarily cognitive (Mayer, 2021a). It does not explicitly account for the cyclical, metacognitive processes of self-regulation strategies that govern the initiation and multilevel execution of generative effort.

The Metacognitive Model of Multimedia Learning complements CTML and explains how generative processing is initiated and sustained (Azevedo & Dever, 2021). Theoretically, guided inquiry is expected to scaffold this self-regulation cycle by providing a framework that requires learners to engage in the model's iterative phases. The guided structure prompts learners to define goals and subgoals (initial phase), dynamically search and filter information based on task relevance (adaptation phase), and compare the constructed mental model with initial objectives (final phase).

This required self-monitoring is thought to mobilise the learner's limited cognitive resources toward organising and integrating the content. Consequently, learning is elevated from superficial processing as the student actively links new information to existing knowledge structures, a process known as elaborative encoding (Jaeger et al., 2025). Elaborative encoding, particularly through generating explanations, requires reasoning abilities and the generation of inferences that go beyond the given content, forcing the revision of mental models (Brod, 2020).

The advantage of these generative strategies is rooted in the explicit generation of semantic mediators, which serve as cues for later recall (Jaeger et al., 2025). This deepened encoding is hypothesised to translate into enhanced generative processing, helping to allocate the learner's limited cognitive resources toward deep schema construction (Brod, 2020). Furthermore, elaborative encoding has been shown to be consistently more beneficial than retrieval practice for associative memory (Jaeger et al., 2025). The integrated mental model is less likely to decay and more structurally flexible, leading to better retention of information and transfer of underlying principles to novel problems. Under this lens, guided inquiry is theorised to serve as the external prompt that initiates and sustains the metacognitive process, potentially enhancing the intensity and quality of the learner's generative processing (Azevedo & Dever, 2021).

Guided inquiry is designed to externalise generative processes by embedding them into the learning task. Meta-analytic evidence supports its benefits for conceptual understanding and critical thinking (Lazonder & Harmsen, 2016; Pedaste et al., 2015). While advocates of direct instruction contend that even guided inquiry can impose excessive cognitive demands on novices (Clark et al., 2012; Kirschner et al., 2006; Sweller et al., 2007), subsequent analyses clarify that scaffolded inquiry, rather than minimal guidance, achieves optimal cognitive results (Chi et al., 2021; Hmelo-Silver et al., 2007; Ouzzine et al., 2022).

Overall, prior video-based learning research has primarily optimised extraneous and essential cognitive processing; consequently, generative processing has rarely been treated as a primary outcome variable, especially in EFL video contexts. While guided inquiry is conceptually indicated to foster deep learning, empirical evidence evaluating the impact of specific pedagogical frameworks, such as the 5E instructional model, on multi-dimensional cognitive load components in EFL higher education remains limited. This gap is crucial, as research on inquiry strategies remains largely centred on STEM disciplines, and relying on non-authentic or single-administration classroom environments limits ecological validity.

Investigating the utility of the 5E guided inquiry model in Morocco represents a relevant case study of a broader international phenomenon. The challenges surrounding the transient information effect, passive media habits, and the insufficient fostering of generative processing are common across many educational contexts. The present study proposes a structured approach to enhancing video-based instruction that requires minimal resources and can be applied in low-tech environments, allowing its core design principles to be adapted across various educational contexts.

Adopting a cognitive approach, the present study investigates the use of the 5E guided inquiry model, conceptually positioned as a generative processing activity, to complement learning from videos in EFL higher education. Specifically, the study measures the impact of this framework on learning outcomes (retention and transfer) and examines its influence on the three dimensions of cognitive processing. Building on this rationale, the study addresses the following research questions:

RQ1: What is the effect of utilising the 5E guided inquiry model with videos on learning outcomes (retention and transfer) in EFL higher education?

RQ2: What is the effect of utilising the 5E guided inquiry model with videos on multi-dimensional cognitive load in EFL higher education?

The study operates under the directional hypotheses that both the 5E treatment group and the direct instruction control group will exhibit positive cognitive processing rates and learning outcomes due to the use of well-designed multimedia materials. However, the treatment group was expected to show stronger retention and higher localised generative engagement, while transfer was explored as a learning outcome**.** These hypotheses are grounded in prior research supporting the efficacy of generative strategies, the current understanding of information processing mechanisms, and the theoretical premise that external scaffolding of the metacognitive cycle is theorised to drive deeper cognitive construction.

METHOD

Research Design

The study adopted a quantitative approach, non-randomised pre/post-test quasi-experimental design, to investigate the impact of guided inquiry with video on learning in EFL higher education (Cohen et al., 2018). The pre-test/post-test non-equivalent group design was chosen as it aligns with the research questions and the contextual constraints. Conducted over five weeks with 135 first-year English department students, two intact classes were used.

Because participants were not randomly assigned at the individual level, the treatment condition was randomly assigned at the class level, resulting in a control group (n=65) and a treatment group (n=70). Both groups studied the British Culture and Society module across eight sessions, with the control group using educational videos and direct instruction, and the treatment group using educational videos and the 5E instructional model of guided inquiry.

Adopting a repeated measures approach, cognitive load was measured at four distinct times over the course of four weeks. This approach allowed for tracking changes in cognitive processing over time, assessing the ongoing impact of the guided inquiry intervention on students' cognitive processing. The weekly measurement focused on essential, extraneous, and generative cognitive processing dimensions.

The Treatment

Video Selection Criteria

The videos used in this study were selected to comply with established design principles and avoid overwhelming cognitive resources (Fiorella & Mayer, 2021; Mayer & Fiorella, 2021b). Video selection and implementation followed research-based guidelines, focusing on principles that reduce extraneous cognitive processing and manage essential cognitive processing.

Educational videos, ranging from 3-5 minutes, were selected to present and introduce key concepts and events related to lessons' objectives (Brame, 2016). The videos incorporated conversational and comprehensible input, as well as light background music and subtitles. To optimise learning, some videos were edited to remove irrelevant content to the lesson's objectives or potentially inappropriate content, supporting content personalisation. This approach intended to use videos in alignment with human cognitive architecture, minimising cognitive overload.

The 5E Instructional Model

The study employed two distinct instructional approaches. The control group followed direct instruction, utilising educational videos alongside conventional lecturing techniques. In contrast, the treatment group implemented the 5E instructional model of guided inquiry (Bybee, 2015; de Jong, 2021; de Jong & Lazonder, 2014; Ruiz-Martín & Bybee, 2022).

To support treatment fidelity, the instructor continuously monitored phase transitions based on formative learner outputs (e.g., responses to heuristic prompts, comprehension checks) rather than strict chronological adherence. To maintain consistency, the instructor utilized standardized lesson plans and detailed instructional checklists. Both groups were exposed to the exact same educational videos, targeted learning objectives, and comparable instructional materials. A standard 90-minute session was structurally operationalised as detailed in Table 1.

Participants

The target population consists of EFL students in higher education in Agadir, Morocco. Due to time, resource, and accessibility constraints, the accessible population was limited to EFL students at the Higher School of Education and Training (ESEF). Convenience sampling was employed to select the two groups for the study. Intact classes were used to maintain natural classroom dynamics, as recruiting participants outside their regular courses could reduce motivation and interfere with authentic learning behaviours.

The treatment group consisted of 70 participants. The majority were females (85.7%), with males comprising 14.3% of the group. Most participants were 18 years old (74.3%), while others were 17 (7.1%), 19 (15.7%), and 20 (2.9%) years old. In terms of English proficiency, nearly half of the participants (47.1%) rated themselves as intermediate, followed by upper-intermediate (25.7%). Fewer participants identified as elementary (14.3%), advanced (7.1%), beginner (4.3%), and proficient/native-like (1.4%).

The control group consisted of 65 participants. The majority were also females (86.2%), with males comprising 13.8% of the group. Most participants were 18 years old (64.6%), followed by 19 (21.5%), 17 (10.8%), and both 20 and 21 years old (1.5% each). Regarding English proficiency, most participants (43.1%) identified as intermediate, followed by upper-intermediate (21.5%) and elementary (20.0%). Fewer participants rated themselves as advanced (7.7%) and beginner (7.7%).

The sample was mostly female and drawn from a single institution, which limits how far the findings can be generalised to other student populations or settings. However, there was no significant difference in age, gender, and language proficiency between the treatment group and the control group. Pre-test results and demographic data indicate that both groups had no prior knowledge of the subject and shared similar demographic characteristics, suggesting their comparability.

While the use of intact groups presents limitations for generalisability, they constitute the complete accessible population within the English department at this institution. Informed consent was obtained from all participants for the use of their learning data and responses for research purposes. All participants were from non-vulnerable populations, and confidentiality and anonymity were maintained throughout the study.

Table 1

Operationalisation of the 5E Guided Inquiry Intervention

Phase	Est. Duration	Instructional Activity	Target Cognitive Function
Engage	~10 min	Heuristic prompts used prior to video viewing.	Activate prior knowledge schemas; prime working memory for new integration.
Explore	~25 min	Segmented viewing of educational videos accompanied by guiding questions and performance tracking.	Manage essential cognitive processing; minimise extraneous processing.
Explain	~15 min	Instructor-led structured elaboration to address knowledge gaps, paired with prompts requiring students to synthesise concepts.	Facilitate active schema construction and link new information to prior knowledge.
Elaborate	~20 min	Application to novel case studies using gradually fading temporary scaffolds.	Promote schema automation and support elaborative encoding.
Evaluate	~10 min	Scenario-based discussions and structured tracking logs (e.g., student notebooks and instructor-led whiteboard summaries of key terms and concepts).	Elicit metacognitive reflection; consolidate generative cognitive engagement.

Instruments

Pre/Post-Test

The achievement test was a researcher-developed instrument designed to evaluate the intervention's impact on learning. Administered as an identical pre-test and post-test, the tool measured retention and transfer outcomes in alignment with the CTML (Mayer, 2021a; Mayer & Fiorella, 2021a).

To support content validity, test items were mapped to the British Culture and Society module objectives. A test blueprint was developed to map each assessment item against the specific learning objectives and cognitive domains (retention vs. transfer) of the module. The test was reviewed by a panel of experts, including the study supervisor, a teacher trainer, and two university professors. This panel evaluated the instrument for subject-matter accuracy and face validity, leading to revisions for item clarity and time efficiency.

The test was scored out of 20 points, with 8 marks allocated to retention and 12 to transfer tasks. Retention was measured through three item types: matching historical events (e.g., the Magna Carta) to descriptions, identifying political institutions and their functions within given scenarios, and multiple-choice questions. The transfer section consisted of four open-ended tasks requiring students to apply conceptual knowledge to novel contexts. These included explaining the historical evolution of British social classes, analysing the impacts of British colonialism (e.g., the global spread of English), and applying sociological reasoning to the influence of the British socioeconomic background and educational access.

Scoring for open-ended items utilised a point-based analytic rubric. Points were awarded based on the presence of predetermined key concepts and the logical coherence of reasoning. Linguistic fluidity and grammatical accuracy were not primary criteria. Responses were required to be coherent and demonstrate a clear understanding to be considered correct. This method prioritised content quality and aimed to fairly capture students' capacity to transfer learning. While blind scoring was not feasible due to the researcher's role as the course instructor and exam invigilator, intra-rater consistency was addressed by independently re-checking all test scores and data entries on a separate occasion to maintain consistency and minimise errors.

Cognitive Load Measure: Self-Report Questionnaire

The cognitive load measure used in this study was a self-report questionnaire developed by Krieglstein et al. (2023) to assess perceived cognitive load within Cognitive Load Theory (CLT) (Paas & Sweller, 2021). The questionnaire differentiates cognitive load into three types: essential, extraneous, and generative, with five items for each type. The items were rated using a 9-point Likert scale (1 = not at all applicable; 9 = fully applicable), chosen for its high reliability (Krieglstein et al., 2022). Despite concerns about subjective self-ratings, research shows respondents can reliably assess their perceived mental effort (Ayres et al., 2021; Skulmowski, 2022).

Research Quality

Multiple measures were incorporated to support the validity and reliability of the findings. Both groups were taught using identical content and videos, with the only difference being the form of instruction. The treatment condition was randomly assigned to one of the existing groups to reduce bias. The groups consisted of first-year students from the same major, creating comparable participant pools. The groups studied for identical periods, and the same instructor delivered the lessons in a consistent structure to both groups to minimise variability in teaching style and quality. Although the classroom environment was not fully lab-controlled, this strengthened ecological validity by resembling real-life learning scenarios.

While Cronbach's Alpha was not measured for the achievement test due to the psychometric challenge of item heterogeneity across the retention and transfer subscales, the study's overall reliability was maintained through the design implemented to manage both participant and researcher error. Learning outcomes were assessed using tests that included both retention and transfer tasks, capturing different aspects of learning. All instruments were piloted and assessed for face validity and potential bias. Pre-tests were administered one week before the treatment to minimise testing effects and verify classroom equilibrium. Data indicated that both groups had minimal prior knowledge of the target content, providing suitable conditions to assess the intervention's impact.

As each group studied two consecutive sessions each week, cognitive load measurements were strategically timed at the end of the first session to minimise potential fatigue impacts on the data. Krieglstein et al. (2023) suggest conducting repeated assessments throughout the learning process to minimise the influence of the subjective nature of self-reports. In agreement with this, cognitive load was measured at the end of the first session each week over the four-week period. This enabled the quasi-experiment to better understand cognitive load dynamics over time, providing a more accurate assessment by minimising extraneous variables that could bias the findings.

Although the measure was previously validated by Krieglstein et al. (2023), internal consistency, Cronbach's alpha, was calculated after the first administration to account for potential contextual effects. The Cronbach's alpha values for both groups from this initial administration are presented in Table 2.

Table 2

Cronbach's Alpha values for Cognitive Load Measure in the First Administration

Cognitive Processing Dimension	Group	Cronbach's Alpha
Essential	Treatment	.695
	Control	.788
Extraneous	Treatment	.837
	Control	.748
Generative	Treatment	.733
	Control	.706

Data integrity was addressed by carefully checking for errors and mistakes. The instructor reviewed the self-reports upon collection for missing data, returning them immediately to students if items were skipped. The self-reports and tests were corrected, entered into SPSS, and double-checked on different days to support intra-rater consistency. Construct validity was maintained by defining and operationalising variables, while statistical validity was supported by verifying the assumptions for parametric tests and using non-parametric alternatives when necessary.

Data Collection

Due to the intact sampling method, the preassigned schedule had classes for the groups at different times. The first group had classes on Friday afternoon from 2:00 PM to 5:30 PM, with a 30-minute break at 3:30 PM. The second group had classes on Saturday morning from 8:30 AM to 12:00 PM, with a 30-minute break at 10:00 AM. Each week, both groups had two sessions. The assignment of the treatment group was determined randomly. A third party selected a wrapped paper to designate the treatment group, which was the first group with classes on Friday. The potential impact of this scheduling variance on student fatigue and attention is formally acknowledged as a possible confound and is accounted for when interpreting the results.

The quasi-experiment lasted five weeks. In the first session of the first week, students were given a pre-test at the beginning of the lesson. They were instructed to answer based on what they already knew, with a clear emphasis that this was a pre-course survey and that the answers would not affect their grades. The students completed the test within approximately 20 minutes, after which they submitted their responses. The remainder of the first session was dedicated to an ice-breaker activity to help the instructor get to know the students. After the 30-minute break, the second session consisted of an introduction to the course. The lessons commenced in the second week. The form of instruction followed the model and treatment outlined earlier. After four weeks, or eight sessions of instruction, the post-test was administered on a Saturday morning, with all groups taking the test on the same day.

A baseline cognitive load measurement was intentionally omitted during the Week 1 pre-treatment phase to minimise survey fatigue immediately following the administration of the pre-test. The implications of this design choice are formally acknowledged and addressed in the Limitations section. The cognitive load measure was administered starting in week two. It was consistently issued at the end of each first session to minimise the influence of tiredness or boredom on the data. In the first administration, items were explained in both English and Arabic as needed, to facilitate students' comprehension of each question. Special attention was given to distinguishing between extraneous and essential load. Each item was explained individually, allowing students to answer at their own pace. Students were encouraged to be honest, as their feedback was crucial for improving instruction. By week three, students were familiar with the process and responded independently, with the instructor moderating and providing assistance as needed. The same process was followed for weeks four and five.

Data Analysis

The analysis of the pre-post test data included checking parametric assumptions. Normality was violated across all datasets, requiring the use of non-parametric statistical tests. Mann-Whitney U tests were employed to assess the differences in learning outcomes between groups, while Wilcoxon Signed-Rank tests were utilised for within-group comparisons.

The cognitive load measure contained 15 items with a 9-point Likert scale. The analysis of the scale was configured so that higher values consistently indicated favourable cognitive processing, while lower values indicated harmful cognitive load. Items 1-10, pertaining to essential and extraneous processing, were reverse-coded to maintain consistent interpretation. Consequently, for all subscales, higher scores indicate more favourable learning conditions: better managed essential processing, lower extraneous processing, and higher generative engagement.

Both descriptive and inferential statistics were applied to the cognitive load questionnaire for within-subject and between-groups analyses. The analysis was conducted at two distinct tiers. First, a macro-level analysis was performed on the aggregate subscale dimensions, essential, extraneous, and generative processing, to evaluate cognitive demand. Second, a micro-level, item-by-item analysis was executed to provide an in-depth understanding of the specific cognitive mechanisms activated by the 5E model, yielding insights to inform pedagogical practice.

For within-subjects analysis, medians and interquartile ranges (IQRs) were calculated, and the Friedman test was applied to each subscale dimension and question within each group to determine if cognitive load changed over the 4-week period. Since the Friedman test indicated significant differences across the four time points, post hoc pairwise comparisons were conducted to determine which specific weeks differed from each other. To reduce the risk of Type I error due to multiple comparisons, the significance level was adjusted using the Bonferroni correction, setting the alpha threshold at $p < .0083$.

For between-subjects analysis, the ordinal data of cognitive load were analysed for significant differences between the two groups using the Mann-Whitney U test, significant at 0.05. For the item-level cognitive load analysis, which involved 60 distinct comparisons (15 items across 4 timepoints), a Holm-Bonferroni sequential correction was implemented. Unlike the standard Bonferroni method, the Holm-Bonferroni approach evaluates ranked *p*-values against a dynamic, step-down threshold. This sequential method was selected because it controls the family-wise error rate across a large volume of comparisons without overly inflating Type II error, preserving statistical power while meeting assumptions for multiple testing.

RESULTS

The Effect of Using Guided Inquiry with Videos on Learning Outcomes

The data for both the treatment and control groups was found to be not normally distributed according to the Kolmogorov-Smirnov and Shapiro-Wilk tests ($p < .05$). Due to this violation, the Mann–Whitney U test was employed as the nonparametric alternative.

A Mann-Whitney U test was conducted to investigate the difference in the pre-test scores between the control and treatment groups. The results indicated that there was no significant difference between the two groups in terms of retention, $Z = -0.419$, $p = .675$, and transfer, $Z = 0.000$, $p = 1.000$, at the 0.05 significance level.

For the post-test, the data was also not normally distributed. Hence, the Mann–Whitney U test was employed as the nonparametric alternative.

A Mann–Whitney U test was conducted to examine differences in post-test scores between the treatment and control groups. The results showed a significant difference in retention scores, $Z = –4.806$, $p < .001$, indicating that the groups differed at the 0.05 significance level with a medium effect size, $r = .414$. However, no significant difference was found

Table 3

Mann-Whitney U Test Results of The Pre-test for The Treatment and Control Groups

	Groups	n	Median	IQR	Mean Rank	Sum of Ranks	*U*	*Z*	*p*-values	*r*
Retention	Treatment	70	.00	0-1	66.79	4675.00	2190.00	-.419	.675	.036
	Control	65	.00	0-1	69.31	4505.00				
Transfer	Treatment	70	.00	.00	68.00	4760.00	2275.00	.000	1.000	-
	Control	65	.00	.00	68.00	4420.00				

Note. Significant at 0.05 level. n = sample size, U = Mann-Whitney U statistic, Z = standardised test statistic, P = p-value, r = effect size.

Table 4

Mann-Whitney U test Results of the Post-test for the Treatment and Control Groups

	Groups	n	Median	IQR	Mean Rank	Sum of Ranks	*U*	*Z*	*p*-values	*r*
Retention	Treatment	70	6	5-7	83.28	5829.50	1205.50	-4.806	.000	.414
	Control	65	5	3-6	51.55	3350.50				
Transfer	Treatment	70	10.62	9.25-11.5	73.09	5116.50	1918.50	-1.578	.115	.136
	Control	65	10.25	8-11	62.52	4063.50				

Note. Significant at 0.05 level. n = sample size, U = Mann-Whitney U statistic, Z = standardised test statistic, P = p-value, r = effect size.

for transfer scores, $Z = -1.58$, $p = .115$. This non-significant result is important because the high post-test scores, approaching the maximum possible score of 12, indicate a potential ceiling effect.

Descriptive analysis of the post-transfer scores revealed a pronounced negative skewness (-1.152) and artificially compressed variance at the upper bound (Treatment Variance = 2.85 vs. Control Variance = 4.57), indicating a significant ceiling effect. A frequency distribution showed that 32.6% of the total sample scored between 11.00 and 12.00 out of a maximum of 12, with 16.3% achieving a perfect score. This concentration of data at the upper bound indicates that the transfer instrument lacked sufficient headroom to capture the full range of participant variance, resulting in mathematical compression of the between-group data.

Concerning the within-subject analysis, the pre-to-post gains provided strong evidence of overall instructional efficacy for both conditions. The treatment group's median retention score increased from 0 to 6, while the control group's median increased from 0 to 5. Wilcoxon Signed-Rank tests demonstrated statistically significant gains for both groups (treatment: $Z = -7.32$, $p < .001$, $r = .87$; control: $Z = -6.82$, $p < .001$, $r = .85$).

It should be noted that the interpretation of the absolute magnitude of these pre-to-post gains is constrained by an observed floor effect in the baseline measurements, which is discussed further in the Limitations. Similarly, a significant improvement was observed in transfer scores between pre-test and post-test (treatment: $Z = -7.282$, $p < .001$, $r = .87$; control: $Z = -7.020$, $p < .001$, $r = .87$). The median gain was 10.62 for the treatment group and 10.25 for the control group, both with large effect sizes.

The directional hypothesis that the treatment group would outperform the control group was supported for retention outcomes. However, the hypothesis regarding superior transfer outcomes remains statistically uninterpretable and thus limited due to the measurement limitations.

The Effect of Using Guided Inquiry with Videos on Cognitive Load

Within-Subjects Analysis of Cognitive Load

Friedman tests were conducted to examine whether there were significant changes in self-reported scores across four time points (Week 2 to Week 5) for both the treatment and control groups (see Appendix A for the complete statistical table). Following the Friedman tests for change over time, Wilcoxon signed-rank tests were used to conduct pairwise comparisons (see Appendix B for the pairwise-values and Bonferroni corrections).

The essential cognitive processing subscale was reverse-coded; higher scores indicate greater ease of understanding and manageability of essential content. Friedman tests showed significant improvements over the four weeks in both groups. Median scores increased from moderate levels (around 4) at Week 2 to high levels (around 7) by Week 5. The effect size, as measured by Kendall's W, was .205 for the treatment group and .163 for the control group, indicating a small proportion of the overall variance in cognitive load.

Pairwise Wilcoxon tests showed that gains were generally sustained from Week 2 to Week 5. While both groups exhibited significant early development in essential processing, the control group demonstrated more consistent significant growth in the latter half of the intervention (W4–W5) compared to the treatment group, which primarily stabilised after the initial increases.

The extraneous cognitive processing subscale was also reverse-coded so that higher scores indicate low mental effort spent on irrelevant or distracting elements. The Friedman test indicated that the treatment group experienced systematically significant positive changes in extraneous processing, though the effect size was very small (Kendall's $W = .092$). Post hoc pairwise comparisons clarified that this variance was not evenly distributed across all timepoints; rather, significant improvements were isolated to the early stages of the intervention. In contrast, the control group remained statistically stable, with no significant changes detected. Item-level median scores generally rose from moderate levels (around 5–6) at Week 2 to higher levels (around 7) by Week 5 in the treatment group. Kendall's W was .092, indicating small but consistent changes over time. Hence, by the end of the study, both groups maintained favourable, efficient overall scores on extraneous cognitive processing.

For the final subscale, generative cognitive processing was not reverse-coded, so higher scores indicate greater investment in organising and integrating knowledge. The Friedman test revealed a clear difference in changes of cognitive processing between the groups. The treatment group experienced a systematic and significant improvement for four items. In contrast, the control group's generative processing remained statistically stable throughout the intervention. Pairwise analysis further supported this finding, showing that the treatment group experienced multiple periods of significant growth. Across all significant Friedman test results, the effect sizes (Kendall's W) ranged from .066 to .235, indicating a small effect of time on the variance in cognitive load.

Between-Groups Analysis of Cognitive Load

To evaluate the overall impact of the instructional conditions on cognitive processing throughout the intervention, participant responses across the four timepoints were aggregated into three primary subscales: essential, extraneous, and gen-

Table 5

Mann-Whitney U Test Results Comparing Cognitive Load Subscales Between Groups

	Week	*U*	*Z*	*p*-values	*r*
Essential Processing	2	2673.00	1.776	.076	.153
	3	2359.50	0.375	.707	.032
	4	2283.50	0.038	.970	.003
	5	2435.00	0.717	.474	.062
Extraneous Processing	2	3143.00	3.869	.000	.333
	3	2389.00	0.512	.609	.044
	4	2341.00	0.294	.768	.025
	5	2251.00	-.107	.914	.009
Generative Processing	2	2221.00	-.241	.809	.021
	3	1989.00	-1.276	.202	.110
	4	2199.50	-.337	.736	.030
	5	1854.50	-1.872	.061	.161

Note. Significant at 0.05 level. U = Mann-Whitney U statistic, Z = standardised test statistic, P = p-value, r = effect size.

erative processing. A series of Mann-Whitney U tests were conducted to compare these aggregate scores between the treatment and control groups at each week.

As presented in Table 5, the Mann-Whitney U tests indicated that the two groups did not differ significantly in their levels of essential processing at any point during the study. Regarding extraneous processing, a significant difference was observed during the first measurement, where the control group (*Mdn* = 7; *Iqr* = 2) showed significantly favourable ratings compared to the treatment group (*Mdn* = 6; *Iqr* = 3), *U* = 3143.00, *Z* = 3.87, $p < .001$. However, this difference did not persist, as no significant variations were found in the subsequent measurements. Finally, the analysis of generative processing revealed no statistically significant differences between the groups throughout the intervention. While the results for the last measurement showed a slight trend toward significance, $p = .061$, they remained above the alpha threshold of .05.

For the item-level cognitive load analysis, line graphs were used to illustrate the progressive development and change of cognitive load over the four-week intervention period for each group. To address the risk of Type I error associated with multiple comparisons across 60 data points, a Holm-Bonferroni sequential correction was applied to all item-level Mann-Whitney U test results.

For essential processing, Figure C1 in Appendix C shows the initial uncorrected analysis of the Mann-Whitney U tests. It indicates significant differences between the groups for Item 1 in Week 2 ($Z = -2.202$, $p = .028$, $r = .189$), Item 5 in Week 2 ($Z = -2.334$, $p = .020$, $r = .201$), Item 5 in Week 4 ($Z = -2.094$, $p = .036$, $r = .180$), and Item 5 in Week 5 ($Z = -2.586$, $p = .010$, $r = .222$). However, following the Holm-Bonferroni adjustment, none of these differences retained statistical significance. This item-level parity supports the aggregate subscale findings, suggesting that the baseline cognitive demand of the multimedia content was adequately managed and perceived equally by both groups across all sessions.

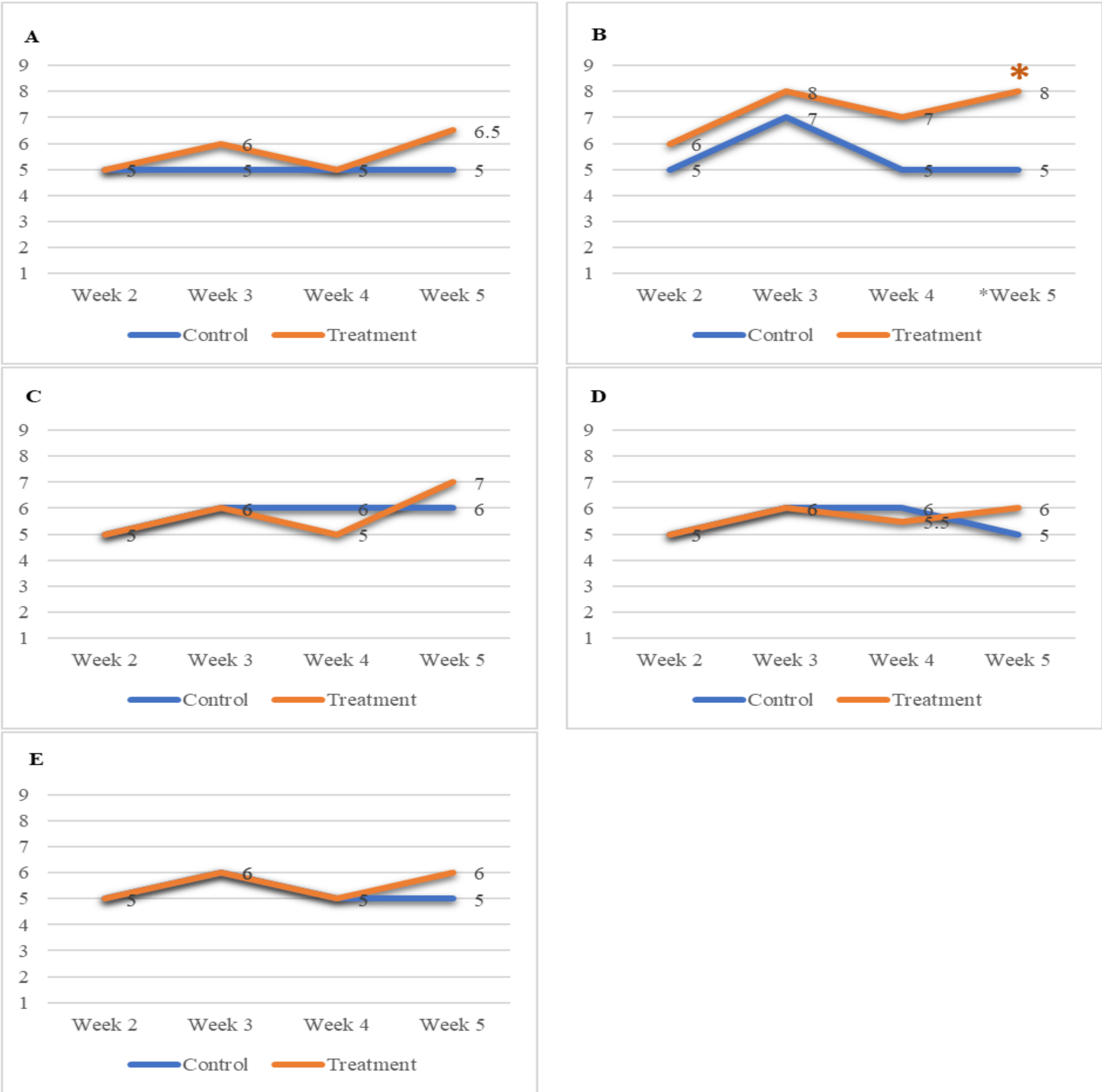
Figure D1 in Appendix D presents the initial uncorrected analysis of the Mann-Whitney U tests. It indicates significant differences between the groups for all extraneous processing items, but only during the first measurement: Item 6 ($Z = -2.679$, $p = .007$, $r = .230$), Item 7 ($Z = -2.783$, $p = .005$, $r = .239$), Item 8 ($Z = -3.950$, $p < .001$, $r = .340$), Item 9 ($Z = -2.682$, $p = .007$, $r = .231$), and Item 10 ($Z = -2.265$, $p = .024$, $r = .195$). However, only Item 8 retained statistical significance post-correction.

Concerning generative processing, the uncorrected analysis of the Mann-Whitney U tests indicated significant differences between the groups for three items: Item 11 in Week 5 ($Z = -2.699$, $p = .007$, $r = .232$), Item 12 in Week 3 ($Z = -2.920$, $p = .004$, $r = .251$) and Week 5 ($Z = -3.790$, $p < .001$, $r = .326$), and Item 14 in Week 5 ($Z = -2.136$, $p = .033$, $r = .184$). However, as demonstrated in Figure 1 below, the Holm-Bonferroni correction revealed that the differences for Items 11 and 14 no longer reached the adjusted significance threshold. Conversely, Item 12 remained significant at the conclusion of the intervention in Week 5.

Overall, the within-subjects analysis demonstrated that the treatment group experienced measurable improvements across all cognitive load dimensions over the four weeks, whereas the control group's cognitive processing remained largely stable. Nevertheless, the between-groups analysis revealed that these internal gains did not reach statistical significance on the aggregate subscales.

Figure 1

Median Generative Cognitive Load Ratings Over Four Weeks for Control and Treatment Groups Across Five Items



Note. Panels A–E display generative cognitive processing items 11–15: (A) active reflection; (B) invested cognitive effort; (C) comprehensive understanding; (D) expanding prior knowledge; (E) knowledge application. Asterisks (*) indicate statistically significant differences between the groups following the Holm-Bonferroni sequential correction.

While the groups demonstrated statistical parity on overall generative processing, the item-level analysis revealed a localised advantage for the treatment group in invested effort. It must be explicitly stated that this item-level finding should be interpreted as a localised effect on invested cognitive effort, not as evidence of a broad improvement in generative processing. Therefore, while the original hypothesis regarding systematic differences in overall generative processing was not fully supported, the findings indicate that the treatment was associated with specific aspects of generative engagement, namely invested cognitive effort.

DISCUSSION

The aim of this study was to investigate the effect of guided inquiry with videos on learning in EFL higher education. Specifically, it aimed at enhancing generative processing and mitigating the transient effect of educational videos. The study targeted learning outcomes, retention and transfer, and cognitive processing as dependent variables to examine the impact of the treatment.

Retention and Transfer

As hypothesised, the findings revealed that both groups achieved significant learning gains. However, the treatment group demonstrated significantly higher retention, yielding a moderate-to-large effect size. This magnitude indicates substantial practical significance, suggesting that the 5E model may have facilitated active schema construction, which enhanced retention of core cultural concepts more effectively than direct instruction. According to CTML, multimedia materials that combine words and visuals reduce extraneous processing and optimise essential processing, thereby promoting meaningful learning (Mayer, 2021a). Yet, even well-designed videos often leave generative processing insufficiently addressed. Guided inquiry is theoretically expected to compensate for this gap by explicitly stimulating generative engagement. Learners in the treatment condition were required to question, hypothesise, and reflect while interacting with videos, a process that can be interpreted as transforming potentially passive observation into active cognitive construction (Azevedo & Dever, 2021).

This process aligns with recent evidence showing that inquiry prompts significantly enhance encoding and long-term retention by promoting elaboration and retrieval-based processing (Fiorella, 2023). The findings are also consistent with the benefits of elaborative encoding retrieval practice (Jaeger et al., 2025). Therefore, guided inquiry can be interpreted as an instructional mechanism that may support schema construction and semantic integration in long-term memory by encouraging more focused cognitive engagement during the learning process (Brod, 2020; Schneider et al., 2021).

While these outcomes are consistent with theoretical frameworks emphasising the benefits of elaborative encoding and metacognitive regulation (Azevedo & Dever, 2021; Jaeger et al., 2025), these specific underlying mechanisms were not directly measured in this study. Therefore, the activation of the metacognitive cycle must be treated as an inferential interpretation rather than a confirmed empirical mechanism.

The findings not only support the global relevance of established multimedia principles and the necessity of fostering active engagement in learning but may also suggest the generalisable value of self-regulation strategies and elaborative encoding for significantly enhancing retention across diverse instructional contexts. The results also contribute to the long-standing pedagogical debate between direct instruction and inquiry-based learning (Clark et al., 2012; Kirschner et al., 2006; Sweller et al., 2007). While direct instruction remains efficient for declarative knowledge acquisition, recent meta-analyses and classroom-based studies demonstrate that guided inquiry, when structured with cognitive scaffolds, supports deeper conceptual retention without sacrificing efficiency (de Jong et al., 2023). The current findings reinforce this perspective, showing that guided inquiry effectively complements well-designed multimedia to support the consolidation and retention of learning.

Concerning transfer, the between-groups null result cannot be interpreted as evidence of comparability between the conditions. While both groups scored relatively high compared to the pre-test, this finding is confounded by an observed ceiling effect on the transfer instrument. A frequency distribution analysis reveals that 32.6% of the total sample scored at or near the maximum threshold. This compression of variance invalidates the measure as a discriminating instrument for between-group comparisons, rendering the transfer results uninterpretable and leaving the research question unresolved.

Regarding the potential cause of this measurement limitation, the significant within-subjects gains across both groups suggest that the baseline multimedia materials, designed according to established multimedia principles, were effective at promoting foundational knowledge abstraction (Mayer, 2021b). Because these core materials provided such an adequate baseline of understanding across all participants, the transfer assessment lacked the necessary difficulty threshold to differentiate performance between the groups. Future research must utilise calibrated, unconstrained transfer instruments to accurately assess any potential effects of the 5E model on knowledge transfer.

Essential Cognitive Processing

Across the four-week intervention, both groups demonstrated significant favourable improvements in essential cognitive processing, reflecting increased ease and manageability of the learning content. This pattern suggests that the videos effectively managed essential processing, allowing learners to allocate cognitive resources to deeper generative and schema construction rather than surface decoding. However, no significant differences were observed between the treatment and control groups at either the cognitive processing subscales or item level.

Interpreted within the measurement constraints outlined in the limitations, the findings suggest that the cognitive demands were controlled and perceived equally across both conditions. Consequently, the addition of the 5E guided inquiry scaffolds did not appear to overwhelm the treatment group's perceived difficulty or artificially inflate their essential cognitive load. From a pedagogical standpoint, this demonstrates that the effective management of essential cognitive processing does not entail minimising difficulty by handing learners information, but rather involves structuring cognitive challenge through guidance.

In traditional video-based environments, essential processing often remains a passive act of comprehension. However, the guided inquiry cycle transforms this process by intentionally placing learners in active epistemic roles that require hypothesis generation, questioning, and meaning construction (Azevedo & Dever, 2021). Theoretically, because the foundational multimedia materials effectively supported initial decoding, learners had the available cognitive capacity to allocate toward complex self-monitoring, regulatory behaviours, and active reflection (Baker et al., 2020; Behrendt et al., 2024; Kaldaras et al., 2024). This suggests that the 5E model can facilitate a crucial shift from passive reception to generative engagement while maintaining a balanced cognitive load.

Extraneous Cognitive Processing

In the first measurement, the control group showed significantly more favourable extraneous processing scores than the treatment group, yielding a medium effect size. However, this difference disappeared in subsequent sessions, with both groups maintaining consistently efficient, focused cognitive engagement.

Accounting for the scheduling limitation, the transient spike in the treatment group indicates an adaptation phase as learners adjusted to the procedural demands of the guided inquiry structure. This initial elevation aligns with findings from Beege and Ploetzner (2025), who reported that prompts and guiding questions can initially inflate perceived extraneous processing. However, because their study utilised a short-term, single-session laboratory design, it likely captured only this early adaptation phase rather than a stable cognitive pattern. The present longitudinal findings extend this line of research by illustrating that what appears as transient overload can, over longer exposure, resolve into efficient regulation of cognitive effort. These results underscore that early, temporary increases in extraneous processing can be a common feature when introducing novel inquiry-based strategies.

This observed pattern highlights the importance of distinguishing between poor multimedia design and temporary procedural unfamiliarity. When learners first navigate structured inquiry, the unfamiliar mechanics of task management can temporarily act as an extraneous source of cognitive load. However, the convergence of extraneous processing scores after the first week reflects the learners' development of procedural fluency. As the mechanics of the 5E inquiry cycle became familiar and automated, this extraneous noise disappeared.

Finally, the overall favourable ratings of extraneous processing across both groups suggest that transient information effects may be mitigated under specific conditions, including short, segmented videos, subtitles, and guided inquiry prompts in this EFL context. Theoretically, well-structured videos can mitigate transient load through signalling and segmentation, while inquiry prompts further support cognitive pacing by encouraging elaboration and self-explanation (Mayer, 2021a). Consistent with previous research, the use of subtitles, typically considered a source of split-attention in CTML, did not appear to have a detrimental effect on cognitive load, providing supporting evidence for their benefits among EFL learners (Pannatier & Bétrancourt, 2024).

Generative Cognitive Processing

The results revealed no significant differences between the groups at the aggregate subscale level for generative processing. While the treatment group consistently exhibited significant within-subject improvements over time, the aggregate between-group difference at Week 5 fell short of statistical significance ($p = .061$). Following the Holm-Bonferroni correction, only one specific item pertaining to invested cognitive effort demonstrated a statistically significant advantage for the treatment group, yielding a moderate effect size ($r = .326$).

Accounting for these limitations and the strict statistical data, the findings describe a localised effect on invested cognitive effort, not evidence of a general improvement in generative processing. The 5E model did not comprehensively transform the students' overall generative processing. The moderate magnitude of the single item-level effect suggests that the 5E model acts as a scaffold for stimulating invested cognitive effort, rather than a broad remedy for generative underperformance. Therefore, claims of broad instructional superiority regarding generative processing are not empirically supported.

This localised increase in invested cognitive effort conceptually aligns with meta-analytic evidence supporting inquiry-based learning structures (Lazonder & Harmsen, 2016; Pedaste et al., 2015). Furthermore, it counters the argument that inquiry approaches inherently impose an excessive cognitive load on novices (Kirschner et al., 2006; Sweller et al., 2007), reinforcing the consensus that scaffolding is the critical moderating factor (Chi et al., 2021; Hmelo-Silver et al., 2007). However, the precise psychological mechanisms driving this effort remain unmeasured. While the guided approach is theorised to externalise generative processes and prompt learners to reflect on constructed mental models (Azevedo & Dever, 2021), this study did not explicitly measure metacognitive regulation or elaborative encoding. Therefore, the theoretical link between the observed increase in invested effort and the activation of a metacognitive self-regulation cycle must be treated as inferential rather than empirically confirmed.

While the within-subjects shifts in cognitive load were statistically significant, the effect sizes representing this temporal change were small (0.066 to 0.235). This indicates that repeated exposure over the four weeks accounted for only a modest proportion of the variance in cognitive processing. In authentic educational settings, cognitive load is multi-determined by factors such as prior knowledge, motivation, and working memory capacity. Therefore, a small time-based effect size is expected, further underscoring that time alone does not substantially alter cognitive load without targeted instructional intervention, and that claims of broad instructional superiority must be tempered.

The original hypothesis was supported only for retention and partially, at the item level, for invested cognitive effort. The study presents a data-driven contribution by systematically measuring all three components of cognitive processing within an EFL higher education context. The findings provide preliminary classroom-based evidence that guided inquiry may support selected aspects of generative engagement, particularly invested cognitive effort. Hence, guided inquiry offers a practical, evidence-based strategy to adapt video-based learning for contemporary digital habits, supporting the transition of multimedia instruction from passive information delivery to active knowledge construction.

Limitations

The current study's findings, while offering meaningful theoretical and empirical contributions, should be interpreted with consideration of specific methodological and contextual limitations. One key limitation of the study is the use of convenience sampling and the lack of random assignment. This limits the internal validity of the study and the broad generalisability of the findings, making it difficult to attribute differences solely to the intervention. Furthermore, the sample was predominantly female and drawn from a single institution. Future research should consider participant characteristics, as well as the gender of the instructor or video presenter, as explicit variables.

Regarding measurement constraints, the absence of a baseline cognitive load measurement and the structural confounds associated with class scheduling limit causal inference. Time-of-day effects and end-of-week fatigue introduce uncontrolled variance that limits causal attribution regarding this initial spike in extraneous load (Anderson et al., 2014). Additionally, a constraint was observed in the pre-test knowledge assessment. Both the treatment and control groups exhibited a median pre-test retention score of 0 (IQR = 0–1). While this reflects a genuine absence of prior knowledge regarding the novel foreign-language content, this compressed distribution mathematically constitutes a floor effect. This lack of instrument sensitivity at the lower bound restricts the interpretability of the absolute magnitude of the pre-to-post learning gains. Furthermore, the transfer subscale produced a significant ceiling effect, which rendered the between-group transfer findings uninterpretable. The study design also lacked a delayed post-test, preventing conclusions regarding the long-term durability of the observed retention gains.

Regarding the measurement of cognitive load, Krieglstein et al. (2023) acknowledge that the theoretical separation of load types may not always be evident in practice. The self-assessment approach relies on learners' metacognitive abilities to accurately reflect on their experiences (Klepsch et al., 2017). Consequently, the mental effort measured via Likert scales provides only a partial picture of the actual cognitive demands (Paas et al., 2008). This reliance on self-reporting was necessitated by the constraints of the large sample size, making physiological or detailed behavioural indicators logistically infeasible. Therefore, future studies should enrich the interpretation of cognitive load by incorporating objective, direct measures alongside qualitative data to offer more precise insights.

While the study's theoretical framework relies on metacognitive engagement, the research lacked direct measures of metacognitive regulation, meaning the mechanisms driving the observed effort remain inferential. Future research should acknowledge that guided inquiry may concurrently influence the development of critical thinking, collaborative skills, and self-regulated learning.

CONCLUSION

The purpose of this study was to investigate the integration of the 5E instructional model within video-based EFL multimedia learning. The findings indicate that 5E-based guided inquiry can support retention of foundational knowledge and localised invested cognitive effort. These results suggest that instructional design in multimedia environments benefits from activities that scaffold active effort alongside the traditional optimisation of essential and extraneous processing.

Regarding cognitive demands, the study indicates that these outcomes can be achieved without creating a sustained increase in extraneous cognitive load. As novice EFL learners in the treatment group did not experience detrimental extraneous load compared to the control group, guided inquiry appears to be capable of prompting engagement without imposing additional cognitive strain. These results highlight the utility of an initial adaptation period when implementing inquiry strategies, during which instructors should provide temporary scaffolds and monitor cognitive load to adjust pacing as learners develop procedural fluency.

Furthermore, the findings suggest that the effective management of cognitive processing may not entail minimising difficulty by providing information; rather, it may involve balancing information delivery with structured engagement activities. The findings suggest that guided inquiry may also have relevance for declarative learning tasks in EFL multimedia contexts. Because the 5E model is adaptable and does not require extensive technological resources, it remains a practical alternative for educational settings where learners may rely heavily on passive video viewing.

These findings offer a context-bound framework for integrating segmented videos with structured inquiry prompts in EFL higher education. The approach provides an evidence-based alternative to passive multimedia consumption, indicating that instructional scaffolding can support the consolidation of learning in specific academic settings. However, the pedagogical implications regarding knowledge transfer remain unresolved. Practitioners should therefore consider these strategies primarily for supporting retention and foundational understanding until further evidence regarding transfer is established. Future research should utilise more sensitive transfer metrics and direct metacognitive measures to delineate the boundary conditions and specific mechanisms of this instructional approach in comparable multimedia learning contexts.

ACKNOWLEDGEMENTS

The École Supérieure de l'Éducation et de la Formation (ESEF) is gratefully acknowledged for granting permission to conduct the study on its premises. Sincere appreciation is also extended to the Faculty of Letters and Human Sciences, Ibn Zohr University, for its continued academic support and encouragement throughout this research. Heartfelt thanks are expressed to the first-year students who participated in the study for their time and engagement.

FUNDING DECLARATION

This work was funded by the CNRST (Centre National pour la Recherche Scientifique et Technique) of Morocco as part of *the PhD-Associate Scholarship-PASS* program. The scholarship provided financial support for the doctoral studies but was not involved in the study design, data collection, analysis, interpretation, writing of the report, or decision to submit the article for publication.

HUMAN ETHICS AND CONSENT TO PARTICIPATE DECLARATIONS

The study was conducted in accordance with internationally recognised ethical standards for research involving human participants, including the ethical guidelines of the American Psychological Association (APA). Formal ethics approval was not required under institutional regulations, while permission to conduct the study was granted by ESEF. Informed consent was obtained from all participants for the use of their non-sensitive learning data and responses for research purposes. All participants were from non-vulnerable populations, and confidentiality and anonymity were maintained throughout the study.

DATA AVAILABILITY STATEMENTS

The data supporting this study are available from the corresponding author upon reasonable request. Access will be granted to qualified researchers for non-commercial purposes, provided that ethical guidelines are followed and participant confidentiality is maintained.

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS' CONTRIBUTIONS

Hamza Hassoun: conceptualization; methodology; formal analysis; investigation; data curation; writing – original draft; writing – review and editing.

Naima Trimasse: conceptualization; methodology; validation; resources; writing – review and editing; supervision.

REFERENCES

Alruwaili, R. F. (2025). Scroll immersion and short-form video use: Predictors of attention, memory, and fatigue among Saudi social media users. *Acta Psychologica, 260*, 1-16. https://doi.org/10.1016/j.actpsy.2025.105674

Anderson, J. A. E., Campbell, K. L., Amer, T., Grady, C. L., & Hasher, L. (2014). Timing is everything: Age differences in the cognitive control network are modulated by time of day. *Psychology and Aging, 29*(3), 648–657. https://doi.org/10.1037/a0037243

Ayres, P., Lee, J. Y., Paas, F., & van Merriënboer, J. J. G. (2021). The Validity of Physiological Measures to Identify Differences in Intrinsic Cognitive Load. *Frontiers in Psychology, 12*, 1-16. https://doi.org/10.3389/fpsyg.2021.702538

Azevedo, R., & Dever, D. (2021). Metacognition in Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 132–142). Cambridge University Press. https://doi.org/10.1017/9781108894333.013

Baker, K., Jessup, N. A., Jacobs, V. R., Empson, S. B., & Case, J. (2020). Productive struggle in action. *Mathematics Teacher: Learning and Teaching PK-12 MTLT, 113*(5), 361–367. https://doi.org/10.5951/mtlt.2019.0060

Beege, M., & Ploetzner, R. (2025). Learning from interactive video: the influence of self-explanations, navigation, and cognitive load. *Instructional Science, 53*, 99–119. https://doi.org/10.1007/s11251-024-09693-5

Behrendt, M. G., Clark, C., Elliott, M., & Dauer, J. (2024). Relation of life sciences students' metacognitive monitoring to neural activity during biology error detection**.** *npj Science of Learning, 9*, 1-12. https://doi.org/10.1038/s41539-024-00231-z

Brame, C. J. (2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE—Life Sciences Education, 15*(4), 1-6. https://doi.org/10.1187/cbe.16-03-0125

Brod, G. (2020). Generative Learning: Which strategies for what age? *Educational Psychology Review, 33*, 1295–1318. https://doi.org/10.1007/s10648-020-09571-9

Bybee, R. W. (2015). *The BSCS 5E instructional model: Creating teachable moments*. NSTA Press.

Castro-Alonso, J. C., Ayres, P., Wong, M., & Paas, F. (2018). Learning symbols from permanent and transient visual presentations: Don't overplay the hand. *Computers & Education, 116*, 1–13. https://doi.org/10.1016/j.compedu.2017.08.011

Chi, S., Wang, Z., & Liu, X. (2021). Moderating effects of teacher feedback on the associations among inquiry-based science practices and students' science-related attitudes and beliefs. *International Journal of Science Education, 43*(14), 2426–2456. https://doi.org/10.1080/09500693.2021.1968532

Clark, R. E., Kirschner, P. A., & Sweller, J. (2012). Putting students on the path to learning: The case for fully guided instruction. *American Educator, 36*(1), 5–11.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge. https://doi.org/10.4324/9781315456539

de Jong, T. (2021). The guided inquiry principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 394–402). Cambridge University Press. https://doi.org/10.1017/9781108894333.041

de Jong, T., & Lazonder, A. W. (2014). The guided discovery learning principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 371–390). Cambridge University Press. https://doi.org/10.1017/cbo9781139547369.019

de Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C. E., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K., & Zacharia, Z. C. (2023). Let's talk evidence – The case for combining inquiry-based and direct instruction. *Educational Research Review, 39*, 100536. https://doi.org/10.1016/j.edurev.2023.100536

Fiorella, L. (2021). Multimedia Learning with Instructional Video. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (pp. 487–497). Cambridge University Press. https://doi.org/10.1017/9781108894333.050

Fiorella, L. (2023). Making sense of generative learning. *Educational Psychology Review, 35*(2). https://doi.org/10.1007/s10648-023-09769-7

Fiorella, L., & Mayer, R. E. (2021). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 185–198). Cambridge University Press. https://doi.org/10.1017/9781108894333.019

Hassoun, A., Beacock, I., Consolvo, S., Goldberg, B., Kelley, P. G., & Russell, D. M. (2023). Practicing information sensibility: How gen Z engages with online information. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–17). Association for Computing Machinery. https://doi.org/10.1145/3544548.3581328

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107. https://doi.org/10.1080/00461520701263368

Jaeger, A., Martins, T. H. P. G., Rodrigues, J. P. P., Muniz, B. F. B., Fonseca, A. L. S. da S., & Gonçalves, A. de O. (2025). The benefits of elaborative encoding over retrieval practice for associative learning. *Memory & Cognition, 53*, 1592–1607. https://doi.org/10.3758/s13421-024-01671-z

Jiang, D., & Sweller, J. (2021). The transient information principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 268–274). Cambridge University Press. https://doi.org/10.1017/9781108894333.027

Kaldaras, L., Wang, K. D., Nardo, J. E., Price, A., Perkins, K., Wieman, C., & Salehi, S. (2024). Employing technology-enhanced feedback and scaffolding to support the development of deep science understanding using computer simulations. *International Journal of STEM Education, 11*, 1-17. https://doi.org/10.1186/s40594-024-00490-7

Kerimbayev, N., Umirzakova, Z., Shadiev, R., & Jotsov, V. (2023). A student-centered approach using modern technologies in distance learning: a systematic review of the literature. *Smart Learning Environments, 10*, 1-28. https://doi.org/10.1186/s40561-023-00280-8

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*, 1-18. https://doi.org/10.3389/fpsyg.2017.01997

Krieglstein, F., Beege, M., Rey, G. D., Ginns, P., Krell, M., & Schneider, S. (2022). A Systematic Meta-Analysis Of The Reliability And Validity Of Subjective Cognitive Load Questionnaires In Experimental Multimedia Learning Research. *Educational Psychology Review, 34*, 2485–2541. https://doi.org/10.1007/s10648-022-09683-4

Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review, 35*, 1-37. https://doi.org/10.1007/s10648-023-09738-0

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 1-38. https://doi.org/10.3102/0034654315627366

Mayer, R. E. (2021a). Cognitive Theory of Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 57–72). Cambridge University Press. https://doi.org/10.1017/9781108894333.008

Mayer, R. E. (2021b). The Multimedia Principle. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 145–157). Cambridge University Press. https://doi.org/10.1017/9781108894333.015

Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review, 36*, 1–25. https://doi.org/10.1007/s10648-023-09842-1

Mayer, R. E., & Fiorella, L. (2021a). Introduction to Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 3–16). Cambridge University Press. https://doi.org/10.1017/9781108894333.003

Mayer, R. E., & Fiorella, L. (2021b). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 243–260). Cambridge University Press. https://doi.org/10.1017/9781108894333.025

Ouzzine, A., Erguig, R., & Boudlal, A. (2022). Discovery-based teaching methodology: A framework for quality teaching and learning. *Journal of Applied Language and Culture Studies, 5*, 9–27.

Paas, F., & Sweller, J. (2021). Implications of Cognitive Load Theory for Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 73–81). Cambridge University Press. https://doi.org/10.1017/9781108894333.009

Paas, F., Ayres, P., & Pachman, M. (2008). Assessment of cognitive load in multimedia learning: Theory, methods and applications. In D. H. Robinson & G. Schraw (Eds.), *Recent Innovations in Educational Technology That Facilitate Student Learning* (pp. 11–35). Information Age Publishing. https://doi.org/10.1108/978-1-60752-942-220251003

Pannatier, M., & Bétrancourt, M. (2024). Learning from academic video with subtitles: When foreign language proficiency matters. *Learning and Instruction, 90*, 1-11. https://doi.org/10.1016/j.learninstruc.2023.101863

Pedaste, M., Mäeots, M., Siiman, L., de Jong, T., van Riesen, S., Kamp, E., Manoli, C., Zacharia, Z., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Pellas, N. (2025). The impact of AI-generated instructional videos on problem-based learning in science teacher education. *Education Sciences, 15*(1), 102. https://doi.org/10.3390/educsci15010102

Pires, F., Masanet, M.-J., Tomasena, J. M., & Scolari, C. A. (2022). Learning with YouTube: Beyond formal and informal through new actors, strategies and affordances. *Convergence: The International Journal of Research into New Media Technologies, 28*(3), 838–853. https://doi.org/10.1177/13548565211020545

Ritzhaupt, A. D., Pastore, R., & Davis, R. (2015). Effects of captions and time-compressed video on learner performance and satisfaction. *Computers in Human Behavior, 45*, 222–227. https://doi.org/10.1016/j.chb.2014.12.020

Ruf, A., Zahn, C., Roos, A.-L., & Opwis, K. (2023). How do enhanced videos support generative learning and conceptual understanding in individuals and groups?. *Educational Technology Research and Development, 71*, 2243–2269. https://doi.org/10.1007/s11423-023-10275-4

Ruiz-Martín, H., & Bybee, R. W. (2022). The cognitive principles of learning underlying the 5E Model of Instruction. *International Journal of STEM Education, 9*, 1-9. https://doi.org/10.1186/s40594-022-00337-z

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2021). The Cognitive-Affective-Social Theory of Learning in digital environments (CASTLE). *Educational Psychology Review, 34*, 1-38. https://doi.org/10.1007/s10648-021-09626-5

Skulmowski, A. (2022). Guidelines for choosing cognitive load measures in perceptually rich environments. *Mind, Brain, and Education, 17*, 20–28. https://doi.org/10.1111/mbe.12342

Sweller, J., Kirschner, P. A., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A Reply to commentaries. *Educational Psychologist, 42*(2), 115–121. https://doi.org/10.1080/00461520701263426

Wong, M., Castro-Alonso, J. C., Ayres, P., & Paas, F. (2018). Investigating gender and spatial measurements in instructional animation research. *Computers in Human Behavior, 89*, 446–456. https://doi.org/10.1016/j.chb.2018.02.017

Xu, T., Liu, Y., Jin, Y., Qu, Y., Bai, J., Zhang, W., & Zhou, Y. (2024). From recorded to AI-generated instructional videos: A comparison of learning performance and experience. *British Journal of Educational Technology, 56*, 1463–1487. https://doi.org/10.1111/bjet.13530

APPENDIX A

DETAILED FRIEDMAN TEST RESULTS FOR COGNITIVE LOAD ITEMS

Table A1

Median Scores and Friedman Test Results for Cognitive Load Items Across Four Weeks

Item	Groups	*n*	Week2		Week 3		Week 4		Week 5		Chi-Square (χ^2)	*df*	*p*-values	Kendall's W
			Mdn	(Iqr)	*Mdn*	(Iqr)	*Mdn*	(Iqr)	*Mdn*	(Iqr)				
Essential Processing	Treatment	70	4	(3)	6	(4)	5	(2.5)	7	(3)	43.11	3	.000	.205
	Control	65	5	(3)	6	(4)	6	(4)	8	(3)	31.7	3	.000	.163
1	Treatment	70	4	(2.5)	6	(4.25)	6	(4.25)	7	(4)	38.52	3	.000	.183
	Control	65	5	(2.5)	6	(4.5)	6	(4)	8	(4)	45.85	3	.000	.235
2	Treatment	70	5	(3)	7	(4.25)	6.5	(3)	8	(4)	25.41	3	.000	.121
	Control	65	7	(4)	7	(5)	7	(3)	7	(4)	9.7	3	.021	.050
3	Treatment	70	4	(5)	5	(4)	5	(4)	7.5	(3)	39.53	3	.000	.188
	Control	65	4	(3)	6	(4)	6	(4)	6	(3)	20.73	3	.000	.106
4	Treatment	70	4	(3)	5	(4)	5	(4)	7	(4)	24.33	3	.000	.116
	Control	65	4	(3)	6	(3)	5	(4)	7	(3)	25.1	3	.000	.129
5	Treatment	70	4	(4.25)	5	(6)	5	(4)	5	(4)	7.44	3	.059	.035
	Control	65	5	(4)	5	(5)	6	(4)	7	(4.5)	7.51	3	.057	.039
Extraneous Processing	Treatment	70	6	(3)	7	(4)	7	(3)	7	(4)	19.35	3	.000	.092
	Control	65	7	(2)	7	(4)	7	(4)	7	(3.5)	3.54	3	.315	.018
6	Treatment	70	5	(3)	5	(5)	5	(3)	7	(4)	17.71	3	.001	.084
	Control	65	6	(3.5)	6	(4)	7	(3)	7	(3)	12.04	3	.007	.062
7	Treatment	70	6	(3)	7.5	(4)	7	(4)	7	(4)	16.22	3	.001	.077
	Control	65	7	(3)	7	(4)	8	(4.5)	7	(4)	2.20	3	.532	.011
8	Treatment	70	5	(5)	7	(4)	6	(3)	7	(4)	20.5	3	.000	.097
	Control	65	8	(3.5)	7	(3.5)	7	(4)	7	(3)	4.57	3	.206	.023
9	Treatment	70	6	(4.25)	7	(4.25)	7	(4)	7	(4)	15.6	3	.001	.074
	Control	65	8	(3.5)	7	(4.5)	7	(5)	7	(4)	1.7	3	.640	.009
10	Treatment	70	6	(4)	7.5	(4)	6.5	(3.25)	7	(4)	13.9	3	.003	.066
	Control	65	7	(3.5)	7	(4)	6	(5.5)	7	(4)	8.5	3	.037	.043
Generative Processing	Treatment	70	5	(3)	7	(3)	6	(2)	7	(3)	24.87	3	.000	.118
	Control	65	5	(3)	6	(3)	5	(3)	6	(4)	3.18	3	.365	.016
11	Treatment	70	5	(3)	6	(4)	5	(2)	6.5	(3)	9.8	3	.020	.047
	Control	65	5	(3)	5	(3)	5	(4.5)	5	(4)	3.95	3	.267	.020
12	Treatment	70	6	(5)	8	(4)	7	(3)	8	(4)	13.89	3	.003	.066
	Control	65	5	(5.5)	7	(4)	5	(5)	5	(4)	4.89	3	.180	.025
13	Treatment	70	5	(2)	6	(4.25)	5	(3)	7	(2)	27.67	3	.000	.132
	Control	65	5	(3)	6	(3)	6	(4.5)	6	(4)	1.45	3	.693	.007

Item	Groups	*n*	Week2		Week 3		Week 4		Week 5		Chi-Square (χ²)	*df*	*p*-values	Kendall's W
			Mdn	**(Iqr)**	***Mdn***	**(Iqr)**	***Mdn***	**(Iqr)**	***Mdn***	**(Iqr)**				
14	Treatment	70	5	(4)	6	(3.25)	5.5	(3)	6	(3)	17.83	3	.000	.085
	Control	65	5	(4)	6	(3)	6	(4)	5	(4)	5.16	3	.161	.026
15	Treatment	70	5	(4)	6	(4)	5	(3)	6	(2)	3.56	3	.312	.017
	Control	65	5	(4)	6	(2)	5	(3)	5	(3)	4.63	3	.201	.024

Note. n = sample size, Mdn = Median, P= p-value, W= effect size. Significant at 0.05 level.

APPENDIX B

DETAILED PAIRWISE COMPARISONS OF COGNITIVE LOAD ACROSS WEEKS

Table B1

Significant Pairwise Comparisons from Wilcoxon Signed-Rank Tests Across Four Weeks

Item	Group	W2–W3	W2–W4	W2–W5	W3–W4	W3–W5	W4–W5
Essential Processing	Treatment	.000*	.000*	.000*	.698	.014	.019
	Control	.018	.006*	.000*	.837	.004*	.001*
1	Treatment	.000*	.000*	.000*	.606	.015	.004*
	Control	.001*	.002*	.000*	.672	.003*	.000*
2	Treatment	.010	.038	.001*	.605	.209	.023
	Control	.628	.114	.007*	.193	.048	.270
3	Treatment	.006*	.003*	.000*	.642	.002*	.001*
	Control	.001*	.002*	.000*	.893	.140	.029
4	Treatment	.005*	.001*	.000*	.642	.018	.100
	Control	.003*	.007*	.000*	.458	.032	.004*
5	Treatment	.244	.099	.018	.765	.303	.500
	Control	.717	.127	.004*	.181	.018	.219
Extraneous Processing	Treatment	.003*	.002*	.008*	.941	.571	.495
	Control	.251	.073	.183	.484	.789	.228
6	Treatment	.186	.038	.001*	.453	.012	.069
	Control	.439	.123	.003*	.307	.032	.083
7	Treatment	.010	.010	.020	.888	.810	.746
	Control	.293	.388	.170	.805	.679	.887
8	Treatment	.003*	.036	.002*	.174	.389	.039
	Control	.616	.232	.097	.345	.362	.801
9	Treatment	.037	.010	.022	.628	.366	.762
	Control	.425	.336	.631	.924	.722	.719
10	Treatment	.003*	.008*	.016	.630	.578	.627
	Control	.755	.091	.938	.035	.606	.005*
Generative Processing	Treatment	.000*	.160	.000*	.004*	.632	.000*
	Control	.025	.298	.070	.084	.353	.343
11	Treatment	.062	.232	.012	.433	.283	.061
	Control	.451	.920	.455	.414	.343	.272
12	Treatment	.001*	.042	.010	.051	.692	.109
	Control	.531	.654	.239	.177	.023	.378
13	Treatment	.003*	.135	.000*	.102	.197	.000*
	Control	.102	.390	.316	.220	.466	.623
14	Treatment	.001*	.086	.000*	.046	.324	.001*
	Control	.004*	.045	.122	.208	.070	.870

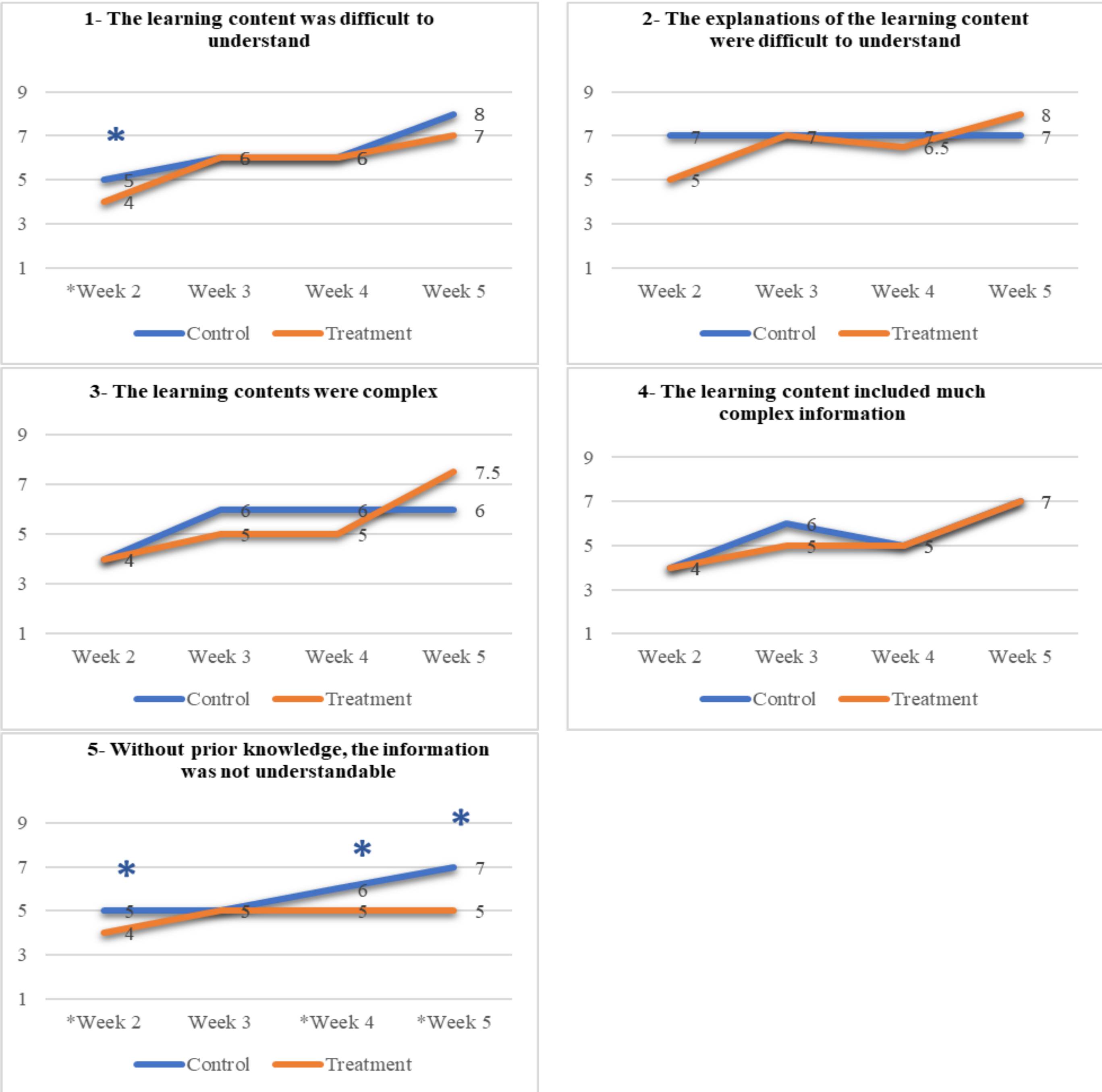
Item	Group	W2–W3	W2–W4	W2–W5	W3–W4	W3–W5	W4–W5
15	Treatment	.123	.448	.132	.266	.932	.155
	Control	.011	.074	.010	.380	.936	.233

Note. Asterisks (*) indicate significance at Bonferroni-corrected alpha level (.0083).

APPENDIX C

UNCORRECTED BETWEEN-GROUP ITEM-LEVEL TRENDS IN ESSENTIAL COGNITIVE LOAD

Figure C1

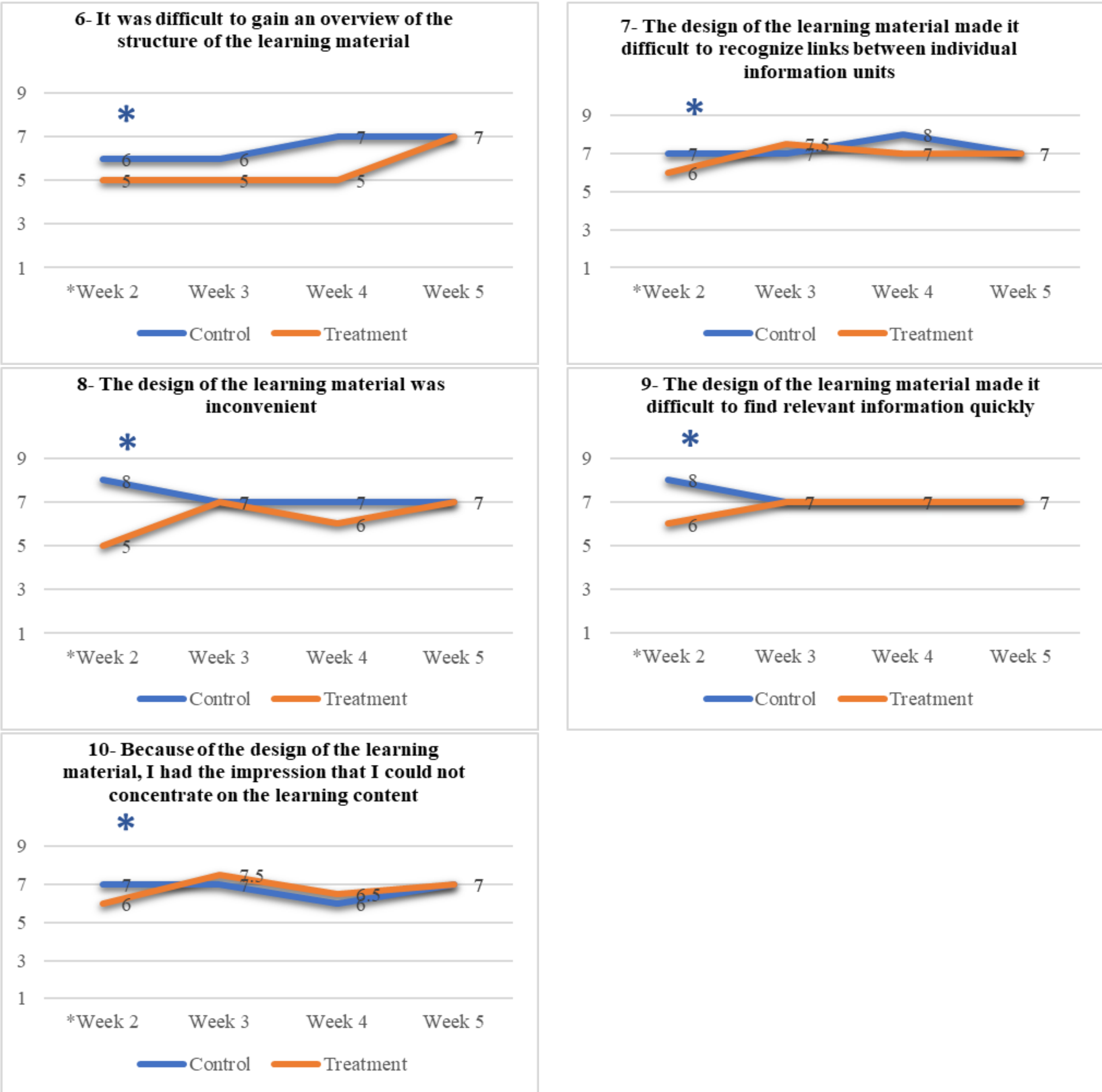


Note. Asterisks (*) indicate raw, uncorrected statistical significance (p < .05). These figures are provided for exploratory visual reference only and should not be interpreted as confirmatory after Holm-Bonferroni correction.

APPENDIX D

UNCORRECTED BETWEEN-GROUP ITEM-LEVEL TRENDS IN EXTRANEOUS COGNITIVE LOAD

Figure D1



Note. Asterisks (*) indicate raw, uncorrected statistical significance (p < .05). These figures are provided for exploratory visual reference only and should not be interpreted as confirmatory after Holm-Bonferroni correction.