Volume 5, Issue 1, June 2026, ISSN: 2948-4316
DOI: https://doi.org/10.5281/zenodo.20823291
Published on: 30 June 2026

Journal of Social Sciences and Business
Quest International University
https://qiu.edu.my/journal/index.php/jssb/index

THE KNOWLEDGE RETENTION AND WRITING PERFORMANCE OF ESL UNDERGRADUATES IN A PRIVATE UNIVERSITY: A COMPARATIVE STUDY OF FINAL EXAMINATIONS AND CONTINUOUS ASSESSMENTS

Angella, Faridah Hanum Ghazali

Quest International University, Perak, Malaysia

**Corresponding author: faridah.hanum@qiu.edu.my*

Abstract: Assessment plays a crucial role in higher education by guiding teaching and learning processes. Two primary forms of assessment are Final Examination (FE) and Continuous Assessment (CA). This study compares their effectiveness in fostering knowledge retention and writing achievement among ESL undergraduates at a private university. A mixed-method, convergent parallel design was employed with 14 purposively selected participants divided into two groups. Group A (N=8) completed a timed, invigilated FE, while Group B (N=6) undertook a take-home CA, both based on the topic of Alternate Ethical Paradigms. Data were collected through written assessments and follow-up semi-structured interviews. Grounded in the Multi-Store Memory Model (Atkinson & Shiffrin, 1968) and the Yerkes-Dodson Law (1908), the study examined differences in knowledge retention and writing performance. Results showed that FE participants experienced a 56.1% decline in retained knowledge after one week, compared to a 50.3% decline in the CA group, indicating better retention under CA. Writing performance also differed significantly, with Group A scoring an average of 56.88% (C+) and Group B achieving 77.5% (A-). Thematic analysis further revealed that participants perceived CA as more effective in assessing understanding and less stressful than FE. Overall, the findings suggest that CA provides a more supportive and effective approach for enhancing both retention and writing performance, offering implications for assessment practices in higher education.

Keywords: continuous assessment, final examination, knowledge retention, multi-store memory model, writing performance

INTRODUCTION

An assessment is a form of evaluation deployed to measure learners' knowledge based on the learning outcomes set by the Ministry of Education (French et al., 2024). Learners will be given a means to certify their levels of achievement according to their results in the assessments. It is no different in higher education, where assessments are used to lay a foundation for learners to modify their learning strategies accordingly and for teachers to adjust their teaching strategies to educate the

learners more efficiently (French et al., 2024; Rezigalla et al., 2017). Assessment is thus very important as it guides the teaching and learning process. It can aid teachers and learners to scaffold knowledge and is a determinant of learners' knowledge in the subject they are studying. Hence, it sets an expectancy for their ability to work in their respective field later. Assessments have undergone multiple shifts over many decades, changing according to the demands of society and educational development. Assessments can generally be divided into two different components – final examination (FE) and continuous assessment (CA). FE refers to a one-shot assessment conducted in a controlled environment within a given time. It is a popular means of assessment of learners' performance and is recognised globally. It is very high-stakes due to its summative nature, whereby the learners have one chance to perform well and are evaluated as a whole (French et al., 2024). CA refers to one or more assignments deployed to learners within the duration of the course for them to complete over an extended period (Playfoot et al., 2023). It is also known as frequent assessment because it is formative, meaning learners can receive feedback on their performance and make necessary improvements towards their work. There are split views regarding the topic of which is a better assessment, efficiency in measuring knowledge or student-friendly testing. Some learners prefer CA because of its flexible assessment, ranging from assignments, quizzes and presentations, allowing them to perform better without posing as an academic load. On the other hand, FE is preferred because it is a fixed and familiar format for learners, making them reluctant to embrace other forms of assessment with confidence.

PROBLEM STATEMENTS

Learners struggle to retain much of the information they have studied when assessed using FE than CA (French et al., 2024). Learners strive to achieve good grades because of parental pressure and societal expectations. Hence, they would take the necessary steps to ensure achievements. Learners typically opt to memorise the contents of a subject if they are to be assessed through FE, as the answers are derived solely from textbooks. The information memorised will only be retained in short-term memory for the sake of the examination and will soon be forgotten afterwards. However, CA can help learners retain knowledge much better because they are required to conduct external research and reading on the topic to complete the assessment. The experience will make the information more memorable because the learners have a better understanding beyond what is taught in the classroom.

Learners face challenges in performing writing skills well through FE compared to CA (Gebremariam et al., 2025). Writing skills are an important foundation of assessments, in which all learners must write in most forms of testing to achieve a grade. Grades in education are significant in society, for it is the stepping stones for learners to further their education in the field they are eligible

for based on their performance. FE is a high-stakes assessment, intended as a one-time summative evaluation that determines learners' overall knowledge and ability. This poses a problem when learners are unable to perform at their best due to external issues such as illnesses or unfavourable weather. CA, on the other hand, is more forgiving in comparison. Learners will be continuously assessed through various methods throughout the semester, allowing them to improve as they go while focusing on specific topics. This gives them many opportunities to bring up their grades and improve their writing skills for future evaluations.

RESEARCH QUESTIONS

RQ1: What are the differences in knowledge retention in FE and CA?

RQ2: What are the differences in writing achievements in FE and CA?

RQ3: What is the effectiveness of FE and CA in promoting knowledge retention and writing performance?

RESEARCH OBJECTIVES

RO1: Investigate differences in knowledge retention in FE and CA.

RO2: Investigate differences in writing achievements in FE and CA.

RO3: Compare the effectiveness of FE and CA in promoting knowledge retention and writing achievements.

SIGNIFICANCE OF THE STUDY

The study is significant in offering insights to academic and educational institutions on a more appropriate assessment method that optimises learners' learning experience. A significant motivation for the research is to determine which method effectively assesses learners' knowledge of the subject matter, rather than their ability to memorise information. By understanding a better method for learners to be assessed properly, lecturers in a private institution can reevaluate existing courses and revamp them accordingly. A few studies have been conducted to compare the efficacy of final examinations (FE) and continuous assessments (CA) in measuring learners' performance. There is also research on how CA contribute to improving FE performance. Many current studies focus on the positive impact of shifting from FE to CA. Although previous studies have examined the effectiveness of continuous assessment and final examinations separately, limited research has comparatively investigated their simultaneous effects on both knowledge retention and writing performance among

ESL undergraduates. Existing studies also tend to focus primarily on academic achievement or student perceptions without integrating cognitive and psychological frameworks such as the Multi-Store Memory Model and the Yerkes-Dodson Law. Furthermore, few studies within the Malaysian ESL higher education context have employed a mixed-method approach to examine how assessment methods influence both long-term retention and writing quality. Therefore, this study addresses these gaps by comparatively examining the effectiveness of final examinations and continuous assessments in promoting knowledge retention and writing performance among ESL undergraduates in a private university.

SCOPE AND LIMITATIONS

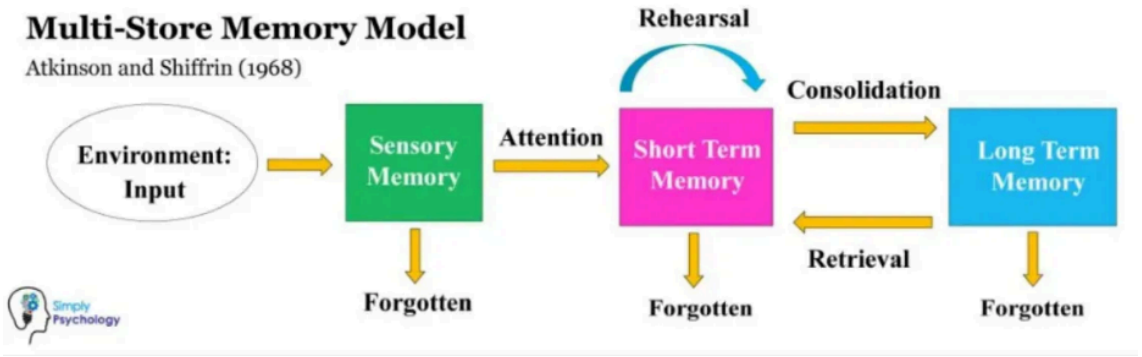
The possible limitations that the researcher recognises are sampling bias and self-report bias. Data collected from a single programme may not represent broader populations, and interview responses may be influenced by social desirability.

THEORETICAL FRAMEWORKS

Multi-Store Memory Model

Figure 1

The cognitive journey of information.

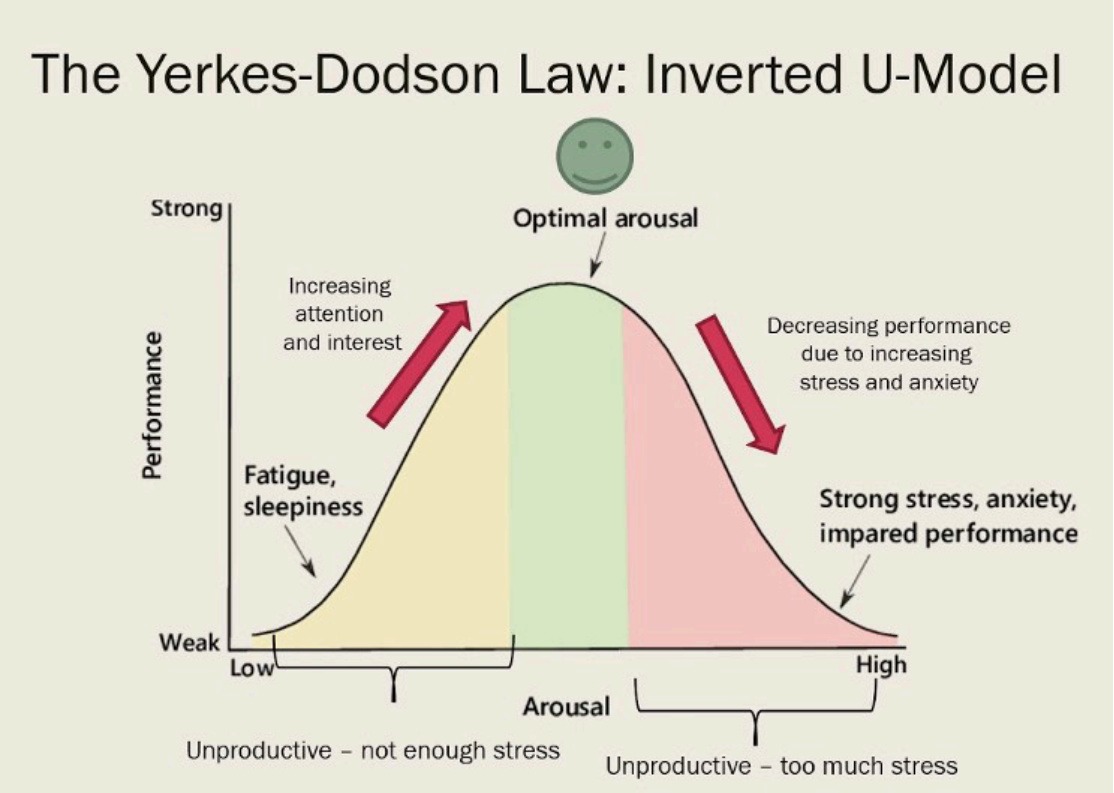


The Multi-Store Memory Model, introduced by Atkinson and Shiffrin (1968), explains how information moves through sensory, short-term, and long-term memory (Plancher & Barrouillet, 2020; Wixted, 2024). Sensory memory briefly retains input, which enters short-term memory with limited capacity unless rehearsed. Through repetition, information transfers into long-term memory, which has unlimited capacity (Finley, 2025). This model explains how assessment practices influence retention.

Yerkes-Dodson Law

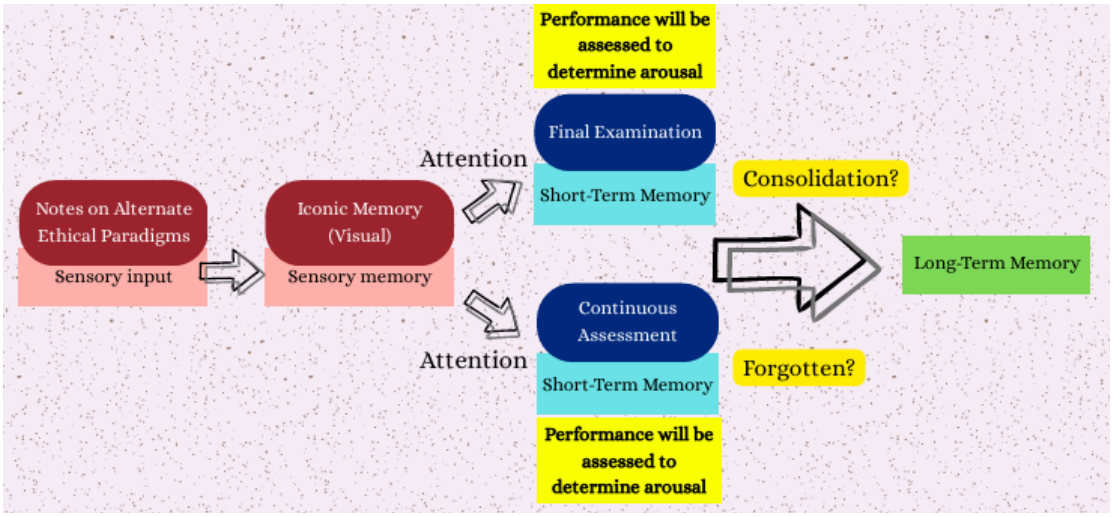
Figure 2

The Inverted U-Model showcasing the empirical relationship between arousal and performance.



The Yerkes-Dodson Law (1908) describes the relationship between arousal and performance, often represented as an inverted U-shaped curve (Kuhn et al., 2025; Nieuwenhuis, 2024). Low arousal reduces motivation, while high arousal causes anxiety and impairs performance (Sörensen et al., 2022). Optimal arousal produces the best outcomes (Kaur et al., 2024; Al-Zoubi et al., 2024). This explains performance differences under varying assessment conditions.

CONCEPTUAL FRAMEWORK



This study examines FE and CA as independent variables, with knowledge retention and writing performance as dependent variables. Knowledge retention relates to long-term memory, while writing performance reflects arousal levels during assessment. Participants receive identical input, processed through memory systems, and are assessed under different conditions reflecting varying arousal levels.

LITERATURE REVIEWS

Knowledge retention involves storing and retrieving information from long-term memory (Azaki & Rivett, 2022) and is essential for improving performance (Enakrire & Smuts, 2023). CA promotes retention through application and feedback, while FE sparks concerns because it often relies on memorisation-based learning, leading to short-term retention (French et al., 2024; Gratchev, 2023). Feedback and testing effects enhance retention (Guran et al., 2020; de Lima & Jaeger, 2020; Yang et al., 2019).

Writing requires organisation, development, and language proficiency (Wright et al., 2021). Many learners, especially non-native speakers, face challenges due to limited practice (Swathi et al., 2025). CA improves writing through iterative feedback and revision opportunities, while timed exams limit development (Cinkara & Au, 2019; Obeiah & Bataineh, 2016; Zaabalawi & Zaabalawi, 2024). Teacher guidance also enhances writing performance (Do, 2023). Digital tools further support improvement in CA settings (Farhan, 2025).

Learners' preferences vary based on individual strengths (Neto et al., 2022). Many prefer assessments that promote critical thinking and practical skills, such as case-based tasks (Melo et al., 2022). Learners generally prefer CA due to flexibility and reduced stress (Rezigalla et al., 2017). CA

supports engagement and application, while FE is often perceived as stressful (Graf et al., 2023; Jopp et al., 2023).

METHODOLOGY

This study adopts a mixed-method design that integrates quantitative and qualitative approaches to provide a more comprehensive understanding of retention and writing performance under final examination (FE) and continuous assessment (CA) conditions. Quantitative data address "to what extent" questions by comparing participants' writing performance and retention outcomes across assessment methods (Lim, 2024; Ghanad, 2023), while qualitative data explore students' experiences, perceptions and learning strategies to explain the "how" and "why" behind the quantitative findings (Oranga & Matere, 2023; Tenny et al., 2022). The integration of both data types enables triangulation, where interview findings support and explain patterns identified in the essay results.

14 ESL undergraduates from a private university in Perak participated in the study. Due to the study's small-scale nature, it is positioned as an exploratory study intended to provide preliminary insights rather than broad generalisations. The sample size was considered appropriate because the study involved detailed textual assessment and in-depth qualitative interviews. From the 14 participants, the top 5 performers from each group are selected for semi-structured interviews, as qualitative saturation commonly occurs between 9–17 interviews (Hennink & Kaiser, 2022). Purposive sampling was employed to ensure participants were BA TESL students who had not previously taken Educational Management (Memon et al., 2025). BA TESL students were selected because they possess similar English proficiency and academic writing experience to meaningfully demonstrate differences in retention, writing performance and learning strategies under FE and CA conditions. Excluding students who had previously taken the course also helped minimise prior content familiarity that could influence performance.

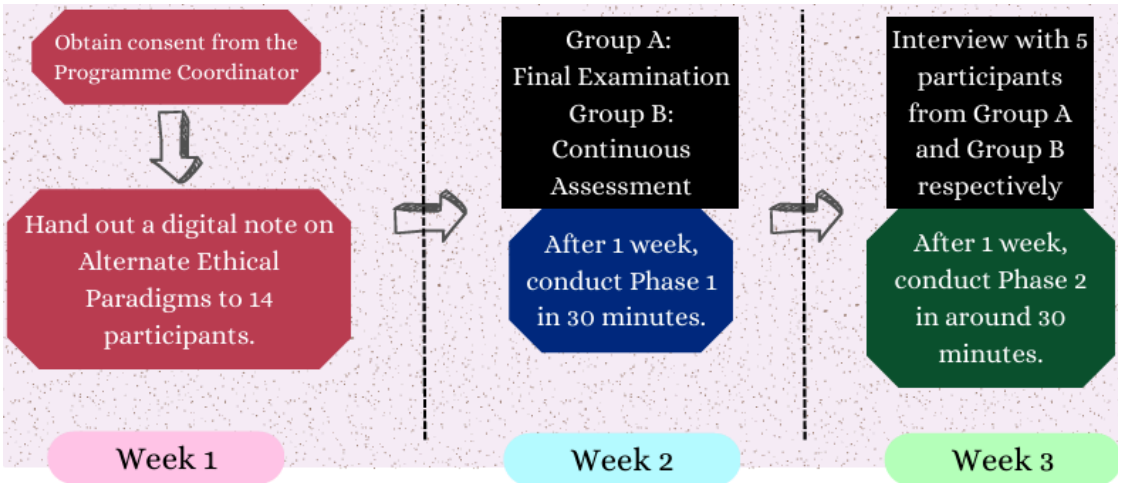
Participants were divided into Group A and Group B to ensure comparability and reduce assessment bias. Both groups completed the same learning content and essay tasks, but experienced different assessment sequences. Group A completed the task under FE conditions first, followed by CA conditions; while Group B completed the assessments in the reverse order. This counterbalancing approach helped reduce order effects, practice effects and familiarity bias. Essays completed under FE were written within a timed setting, whereas CA essays were completed as take-home assignments.

To enhance scoring consistency, the rubric criteria were adapted from established ESL writing assessment frameworks and applied uniformly across all scripts (Abdi, 2023; Lee, 2020). The assessment focused on clearly operationalised criteria, including content, organisation, language use,

analysis and contextualisation. Although formal inter-rater reliability was not conducted due to the small-scale exploratory nature of the study, consistent scoring procedures were maintained throughout the marking process.

Although the small sample size limits the generalisability of the findings, several measures were taken to ensure credibility and trustworthiness. The study employed triangulation through the integration of essay scores and interview data, maintained consistent assessment procedures, and used purposive sampling to ensure participants were relevant to the study objectives. These steps strengthen the reliability of the findings while positioning the study as an exploratory contribution to research on assessment methods in ESL contexts.

DATA COLLECTION



In the pre-collection phase, materials are prepared and approval is obtained. Participants are recruited via WhatsApp and provided with an indemnity form outlining ethical considerations.

During data collection, participants receive digital notes. Group A studies for one week before completing the test under supervision, while Group B completes and submits the assignment within the same period. After one week, follow-up interviews are conducted. This interval allows measurement of retention decline (Esposito, 2025; Opitz & Kubik, 2024). Interviews are recorded with consent.

In the post-collection phase, essays are marked and analysed using JASP. Group A receives grades only, while Group B receives feedback. Interview data are transcribed and analysed thematically.

DATA ANALYSIS

Research Questions	Type of data	How will it be analysed?
What are the differences in knowledge retention between students' performances in the FE and the CA components?	Qualitative	Document Analysis + Thematic Analysis
What are the differences in writing achievements between students' performances in the FE and the CA components?	Quantitative	Jeffreys's Amazing Statistics Program (JASP) (Ott et al., 2025). (Love et al., 2019).
What is the effectiveness of FE and CA in promoting knowledge retention and writing performance among students?	Qualitative	Thematic Analysis

Data are analysed using document analysis, JASP and thematic analysis. Document analysis evaluates essays systematically (Armstrong, 2022). Quantitative data are processed using JASP, which supports statistical analysis and visualisation (Ott et al., 2025; Love et al., 2019). Qualitative data are analysed using thematic analysis to identify patterns in participants' responses (Ahmed et al., 2025), supporting conclusions on knowledge retention and writing performance.

ANALYSIS AND FINDINGS

Participants' Demographic Profile

The demographic profile includes participants' age, year of study and gender. No additional personal data was collected.

Table 1

Participants' age

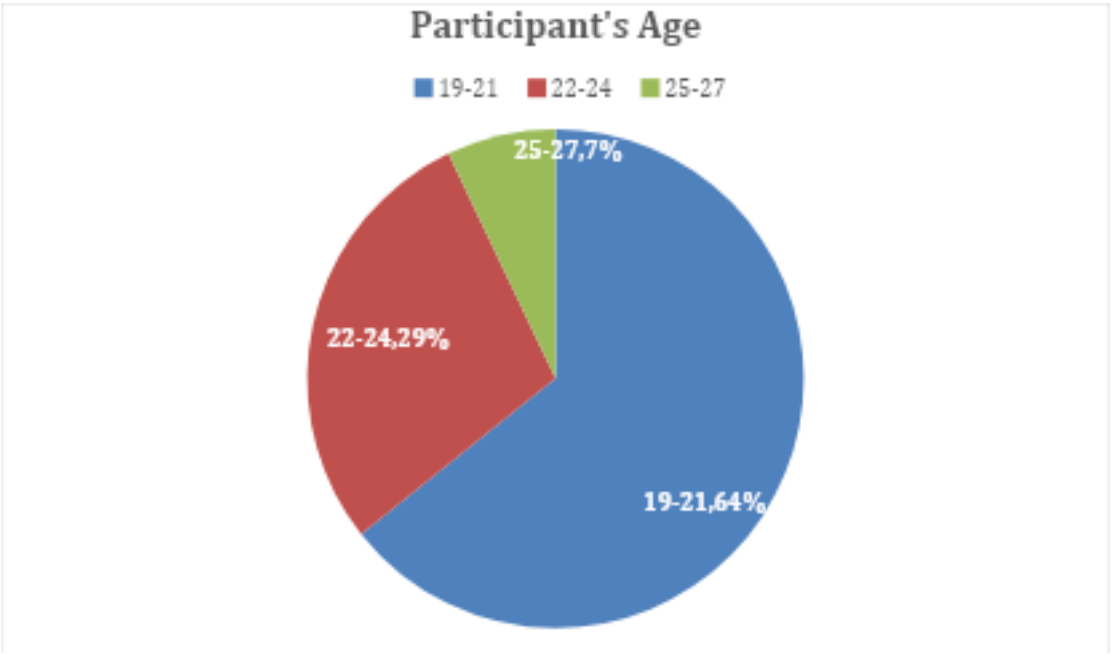


Table 2

Participants' year of study

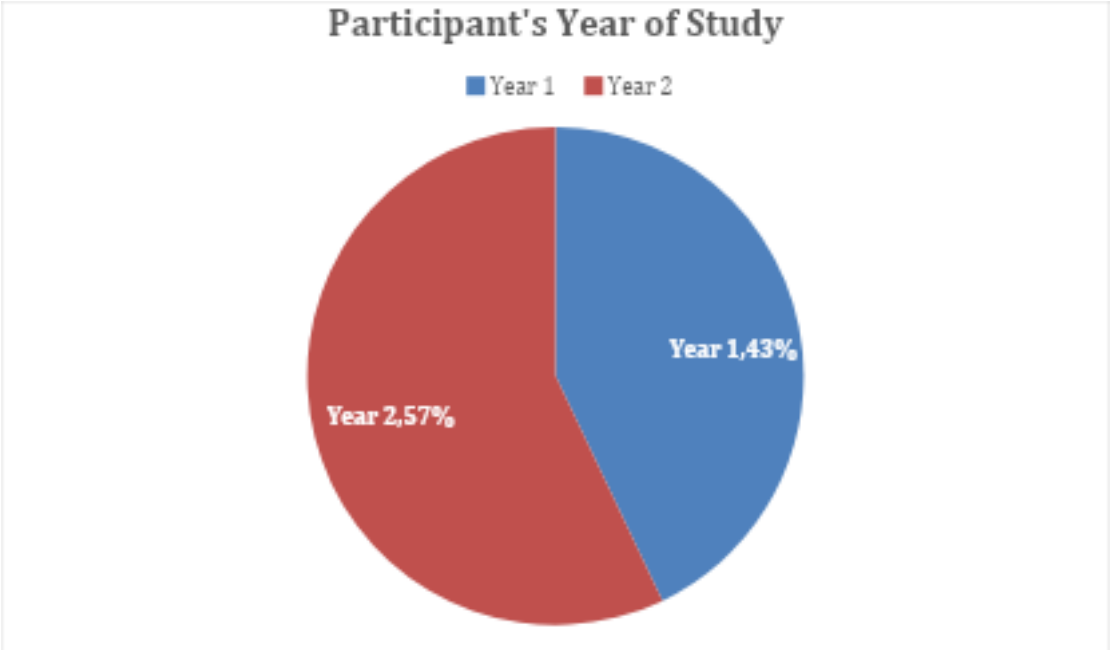
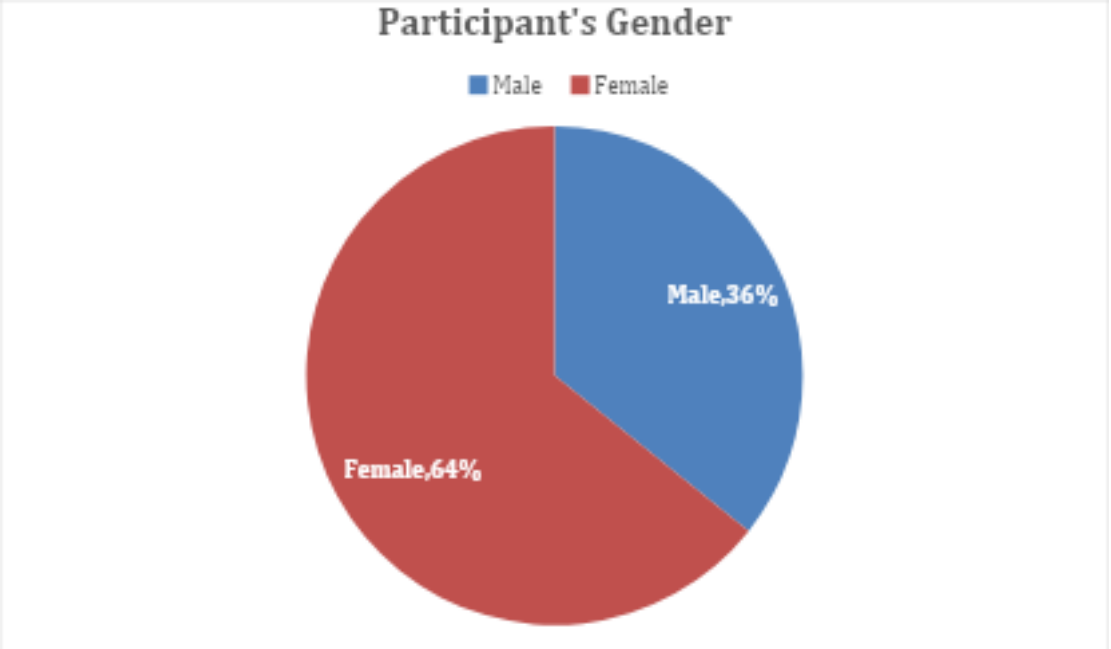


Table 3

Participants' gender



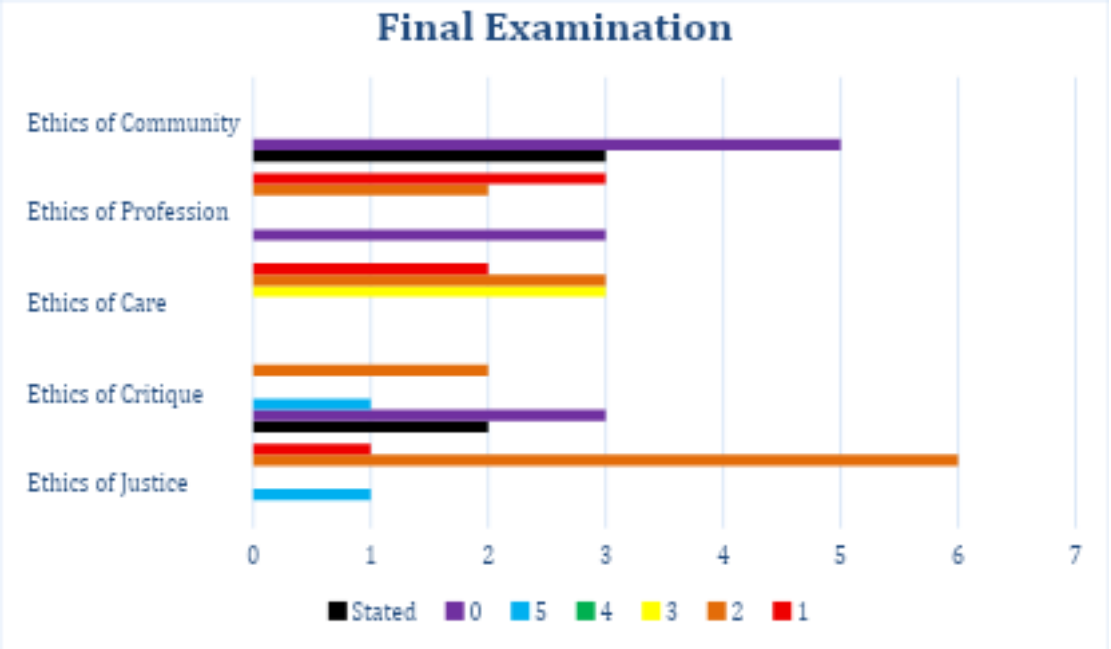
According to Tables 1, 2 and 3, most participants were aged 19–21 (64.3%), followed by 22–24 (28.6%) and 25–27 (7.1%). Year 2 students (57.1%) slightly outnumbered Year 1 students (42.9%), while female participants (64.3%) exceeded male participants (35.7%). Although distributions are acceptable, a more balanced representation is recommended for future studies.

Knowledge Retention

The first research question is addressed through document analysis (Phase 1) and thematic analysis (Phase 2). Written responses measure initial retention, while interview data (Questions 7–12) assess recall after a delay.

Table 4

Final Examination (N=8) Categorisation



Note: Number of participants based on the number of achieved criteria.

Table 5

FE Content Breakdown

Criteria	**Answered (N=8)**	**Total Points (n=40)**	**Percentage (%)**	**Key Word/s**
Justice	8	18	45.0	- Reward & punishment - Rules - Laws
Critique	6	9	22.5	- Voice out/heard
Care	8	17	42.5	- Communication - Empathy - Good relationship

Profession	8	7	17.5	- Safe learning environment
Community	5	0	0.0	

Note: Average percentages of participants' answers and achievements using formulas A = (Answered/N=8)% and TP = (Total/n=40)%.

Table 4 displayed that the participants most frequently recalled Ethics of Justice and Care. The former may be attributed to its position as the first topic in the notes, while the latter is conceptually easier to understand. Ethics of Community was least recalled, likely due to its placement at the end of the material. All participants mentioned Ethics of Justice, Care and Profession, while fewer referenced Ethics of Critique and Community. Table 5 further indicates the key points. From the table, it is revealed that all of the participants mentioned the Ethics of Justice (A=100.0%, TP=45.0%), Care (A=100.0%, TP=42.5%) Profession (A=100.0%, TP=17.5%) in their exam paper, but not every participant mentioned Ethics of Critique (A=75.0%, TP=22.5%) or Community (A=62.5%, TP=0.0%).

Table 6

FE Interview Answers Q7-Q12 Content Breakdown

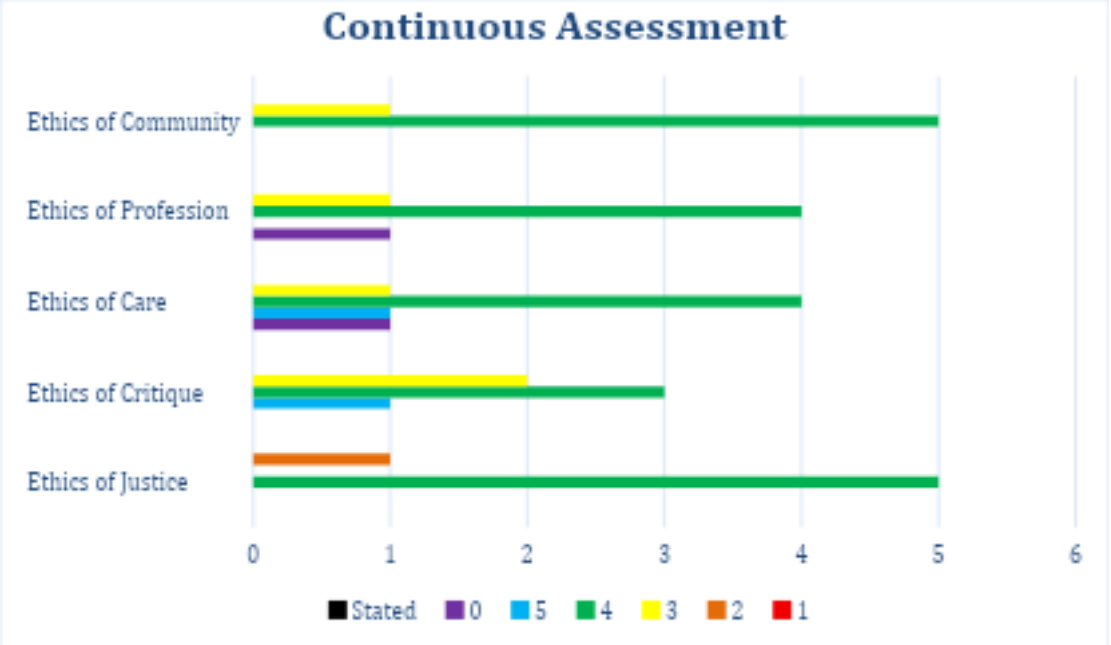
Criteria	**Answered (N=5)**	**Total Points (n=25)**	**Percentage (%)**	**Key Word/s**
Justice	4	3	12	- Fairness - Laws
Critique	3	3	12	- Empowerment - Voice out/heard
Care	3	5	20	- Empathy - Responsiveness
Profession	3	3	12	- Safe learning environment - Professional
Community	0	0	0	

Note: Average percentages of participants' answers and achievements using formulas A = (Answered/N=5)% and TP = (Total/n=25)%.

From Table 6, interview results show a decline in retention. Ethics of Justice (A=50.0%) remained the most recalled, followed by Ethics of Critique (A=37.5%), Care (A=37.5%) and Profession (A=37.5%), while Ethics of Community was not recalled at all (A=0.0%). Although responses varied, Ethics of Care showed slightly stronger elaboration (TP=20.0%). However, in terms of elaboration, students remember more points in the Ethics of Care (TP=20.0%) than in the Ethics of Justice (TP=12.0%), unlike in Phase 1. The findings suggest that the overall retention decreased significantly from Phase 1 to Phase 2. The mean of the information the participants remembered during the exam in Phase 1 is 25.5%, but it dropped to 11.2% during the one-on-one interview session in Phase 2. By calculating the difference, which is 14.3%, over the initial mean of 25.5% and multiplying by 100, it can be mathematically proven that the participants have forgotten 56.1% of what they initially remember.

Table 7

Continuous Assessment (N=6) Categorisation



Note: Number of participants based on the number of achieved criteria.

Table 8

CA Content Breakdown

Criteria	Answered (N=5)	Total Points (n=30)	Percentage (%)	Key Word/s
Justice	5	22	73.3	- Law - Rule
Critique	5	23	76.7	- Injustice - Inequality - Underrepresent - Amplify voices
Care	5	24	80.0	- Empathy - Compassion - Relationship - Communication
Profession	5	19	63.3	- Student-centered - Professional code - Personal value
Community	5	23	76.7	- Shared responsibility - Democratic value - Collaboration - Student

Note: Average percentages of participants' answers and achievements using formulas A = (Answered/N=5)% and TP = (Total/n=30)%.

Table 8 shows that the participants demonstrated stronger performance across all ethics, particularly Ethics of Care (TP=80.0%), Critique (TP=76.7%), Community (TP=76.7%) and Justice (TP=73.3%). Ethics of Profession (TP=63.3%) showed the weakest performance, likely due to limited familiarity with the concept. All participants addressed all five ethics (A=100.0%), with the highest understanding observed in Ethics of Care, followed by Critique, Community and Justice. Ethics of Profession remained the least developed.

Table 9

CA Interview Answers Q7-Q12 Content Breakdown

Criteria	Answered (N=5)	Total Points (n=25)	Percentage (%)	Key Word/s
Justice	5	9	36	- Fairness - Law
Critique	5	9	36	- Discrimination - Challenge system
Care	5	10	40	- Empathy - Relationship - Well-being
Profession	5	7	28	- Professional - Personal
Community	5	11	44	- Collaboration - Communication - Students - Teachers

Note: Average percentages of participants' answers and achievements using formulas A = (Answered/N=5)% and TP = (Total/n=25)%.

From Table 9, the interview findings indicate that all participants retained all five paradigms, with clearer patterns in responses. Ethics of Community (TP=44.0%) and Care (TP=40.0%) were most elaborated, while Ethics of Profession (TP=38.0%) remained weakest. The performance for the Ethics of Justice (TP=36.0%) and Critique (TP=36.0%) is the same. Consistent keywords suggest stronger conceptual understanding. The mean of the information that the participants provided during the assignment in Phase 1 is 74%, but it dropped to 37.2% during the one-on-one interview session in Phase 2. By calculating the difference, which is 37.2%, over the initial mean of 74% and multiplying by 100, the mathematical calculation proves that the participants have forgotten 50.3% of what they initially used during their write-up.

Writing Achievement

The second research question is addressed using quantitative analysis. Writing performance is evaluated using rubric criteria, with mean scores interpreted based on Imsa-ard et al. (2021).

Table 10

Mean Score Interpretation

Level	Mean Score Interpretation Scale
Very Poor	1.00 – 1.80
Poor	1.81 – 2.60
Average	2.61 – 3.40
Good	3.41 – 4.20
Excellent	4.21 – 5.00

Source: Imsa-ard et al. (2021)

Table 11

FE Rubric (N=8, M=3.025)

Criteria	Mean	Standard Deviation	Level
Content	2.625	0.518	Average
Language and Grammar	3.000	1.195	Average
Organisation and Structure	3.625	0.518	Good
Analysis	3.000	0.756	Average
Contextualisation	2.875	0.641	Average
Total Marks	56.88		

From Table 11, the writing performance of the participants in Group A who were assessed using FE can be looked into in terms of the average total marks among the 8 participants and the level of achievements for each of the 5 criteria in the assessment rubrics. The average total mark for Group A is 56.88, putting it in the C+ grade. For FE, the average score (M=3.025) indicates overall average performance, with organisation and structure as the strongest areas. Language, analysis, contextualisation and content remained at average levels. The structured essay format taught in earlier education likely contributed to performance in organisation.

Table 12

CA Rubric (N=6, M=3.556)

Criteria	Mean	Standard Deviation	Level
Content	4.167	0.408	Good
Language and Grammar	3.500	0.548	Good
Organisation and Structure	4.833	0.408	Excellent
Analysis	3.333	0.816	Average
Contextualisation	3.000	0.894	Average
Referencing	2.500	1.049	Poor
Total Marks	77.50		

From Table 12, the writing performance of the participants in Group B who were assessed using CA can be looked into in terms of the average total marks among the 6 participants and the level of achievements for each of the 6 criteria in the assessment rubrics. The average total mark for Group B is 77.5, putting it in the A- grade. For CA, the average score (M=3.556) reflects good performance, with organisation and structure achieving an excellent level. Content and language were also strong, while analysis and contextualisation remained average. Referencing was the weakest area due to difficulties in applying the APA 7th edition format. Overall, CA participants demonstrated higher writing performance than FE participants, likely due to extended time, access to resources and opportunities for revision.

Assessment Effectiveness

The third research question is addressed through thematic analysis of interview responses (Questions 1–6).

Table 13

Breakdown of Interview Answers Q7-Q12

Students	Q1	Q2	Q3	Q4	Q5	Q6
FE1	Memorise	No	No	Forgot	3	Yes. Invigilator=No
FE2	Read PPT	No	No	Forgot	4	Moderate Invigilator=Yes

FE3	Memorise	Yes	Yes	Partial	4	Moderate Invigilator=Yes
FE4	Memorise	Yes	Yes	Remember	10	Yes Invigilator=Yes
FE5	Memorise	Yes	Yes	Forgot	7	Yes Invigilator=Yes
CA1	N/A	Yes	Yes	Remember	8	No
CA2	Make outline Notes	Yes	No	Remember	8	No
CA3	Notes	Yes	Yes	Remember	9	No
CA4	Notes	Yes	Yes	Remember	9	No
CA5	NotesReading	Yes	Yes	Remember	7	Moderate

From Table 13, the findings show clear differences between groups. For question 1, study strategies differed between the two groups. In Group A, 80% relied on memorisation to complete the exam, while only Student FE2 used PowerPoint slides without memorising. Traditional examinations encourage memorisation rather than understanding. In contrast, 80% of Group B referred to notes, including PowerPoint slides and the textbook, while Student CA2 outlined a draft and Student CA5 read additional sources, suggesting that CA required less stressful cognitive work than FE.

Question 2 examined the appropriateness of the assessment in evaluating knowledge of Alternate Ethical Paradigms. In Group A, 40% disagreed that the test was appropriate because it focused on memory, while 60% agreed. In contrast, all Group B interviewees agreed the assignment was appropriate as it required reading and understanding rather than memorisation.

Similarly, question 3 investigated general knowledge assessment. Again, 40% of Group A disagreed due to its focus on memory, while 60% agreed, with Student FE5 noting that tests require examples from different perspectives. Meanwhile, 80% of Group B agreed that the assignment broadened their knowledge, although Student CA2 disagreed due to its narrow focus.

Question 4 explored information retention. In Group A, 60% reported forgetting most of the content, while others recalled limited information. During question 7, FE1, FE2, and FE4 recalled three ethics each, FE5 recalled one, and FE3 recalled none. In contrast, all Group B interviewees remembered most of the content and could name all five paradigms.

Question 5 showed Group A's average usefulness rating was 5.6/10, while Group B averaged 8.2/10, indicating a preference for assignments. Finally, question 6 showed higher stress in Group A, where 60% reported stress and 80% cited invigilator presence, while most Group B participants reported no stress.

DISCUSSION

The findings of this study support previous research suggesting that continuous assessment promotes deeper learning and stronger retention compared to high-stakes examinations (French et al., 2024; Gratchev, 2023). Similar to Rezigalla et al. (2017), participants in the present study perceived continuous assessment as less stressful and more supportive of meaningful understanding. However, this study extends previous literature by comparatively examining both knowledge retention and writing performance within the same research design while integrating cognitive and psychological theoretical frameworks. The findings, therefore, contribute additional insight into how assessment structure influences not only academic outcomes but also long-term cognitive processing among ESL learners.

Findings indicate that CA is more effective in promoting long-term knowledge retention than FE. This conclusion is supported by both quantitative and qualitative data, reflecting differences in performance and cognitive processing. The results align with the Multi-Store Memory Model (Atkinson & Shiffrin, 1968), which explains how information is transferred from sensory input to long-term memory. Both groups received identical materials, ensuring similar initial encoding. However, differences emerged in how information was processed and retained. FE participants relied primarily on memorisation strategies, consistent with maintenance rehearsal, which supports short-term retention but weak long-term consolidation (Wang et al., 2022). This approach allowed temporary recall during the exam but resulted in rapid forgetting. Interview responses confirmed this, as participants struggled to recall key concepts after one week, demonstrating a clear decline in retention. In contrast, CA participants engaged in elaborative rehearsal through activities such as note-taking and linking concepts. These strategies promote deeper processing and stronger connections in memory, increasing retention (Wang et al., 2022). The requirement to research, organise and apply knowledge encouraged integration with existing understanding, resulting in more durable memory traces. Additionally, feedback provided strengthened retention through reinforcement and error correction, aligning with the testing effect (Guran et al., 2020; Yang et al., 2019). Unlike FE participants, who received only grades, CA participants benefited from detailed feedback that supported consolidation. Although both groups experienced some decline over time, CA participants retained substantially more information and demonstrated better recall during interviews.

The findings also show clear differences in writing achievement, with CA participants significantly outperforming FE participants. This difference is attributed to both structural constraints and psychological factors. FE conditions limited participants to a short timeframe without access to resources, restricting their ability to plan, revise and refine their writing. As a result, responses reflected immediate recall rather than developed arguments, leading to average performance levels. In contrast, CA allowed participants to engage in the full writing process, including planning, drafting and revising. Access to resources and extended time enabled more accurate language use, better organisation and stronger content development, resulting in higher overall performance. These findings align with the Yerkes-Dodson Law, which explains the relationship between arousal and performance (Kuhn et al., 2025). FE conditions created high levels of stress, which impaired performance, while CA maintained lower to optimal stress levels, supporting better outcomes. Interview data further support this interpretation. FE participants reported stress related to memorisation, time constraints and invigilation, which affected their ability to focus and perform. High arousal likely reduced working memory efficiency, limiting idea generation and coherence (Knezevic et al., 2023). In contrast, CA participants reported minimal stress, allowing them to focus on writing tasks more effectively. The lower stress environment supported cognitive processes such as planning and revision, leading to improved writing quality. Overall, writing performance differences reflect the combined influence of assessment conditions and psychological factors, with CA providing a more supportive environment for academic writing.

The findings further indicate that CA is perceived as more effective than FE in promoting learning, retention and writing performance. FE participants primarily relied on memorisation strategies and reported limited understanding of the content, often viewing the assessment as focusing on recall rather than comprehension. In contrast, CA participants used a range of strategies, including outlining, note-taking and research, indicating deeper engagement with the material. Perceptions of appropriateness also differed. While FE responses were mixed, CA participants consistently agreed that the assessment effectively measured understanding and application of knowledge. Retention perceptions further highlight this difference: FE participants reported forgetting most of the content shortly after the exam, while CA participants indicated better retention due to repeated interaction with the material. Usefulness ratings followed a similar pattern. FE responses varied widely, often reflecting dissatisfaction with limited retention, whereas CA participants consistently rated the assessment highly, citing improved understanding, critical thinking and real-world application. Stress levels were another key factor. FE participants reported moderate to high stress, influenced by time pressure and invigilation, while CA participants experienced minimal or manageable stress due to flexibility and control over the task. Overall, CA demonstrates greater effectiveness as an assessment method, as it supports deeper learning, sustained retention and improved performance while reducing negative psychological effects.

RECOMMENDATIONS

Future research should expand the sample population to include students from diverse academic backgrounds to examine how different disciplines influence performance under FE and CA. A larger sample size would also improve generalisability. Further studies could analyse differences in cultural or linguistic background to explore how upbringing, language and environment affect learning strategies and assessment performance. Examining varied topics may also reveal how different groups respond to specific subject matter, providing deeper insights into assessment effectiveness.

CONCLUSION

This study compared the effectiveness of final examinations (FE) and continuous assessments (CA) in promoting knowledge retention and writing performance among ESL undergraduates. The findings revealed that CA was more effective in supporting long-term retention, improving writing achievement and encouraging deeper engagement with learning. In contrast, FE participants relied more on memorisation and experienced higher stress levels, which negatively affected retention and writing performance. Guided by the Multi-Store Memory Model and the Yerkes-Dodson Law, the study highlights how assessment conditions influence both cognitive processing and academic outcomes. Overall, the findings suggest that CA provides a more supportive and meaningful approach to assessment in ESL higher education contexts. These findings support prioritising assessment approaches that emphasise understanding rather than memorisation (Vlachopoulos & Makri, 2024). As higher education continues to evolve towards more learner-centred approaches, the findings suggest that continuous assessment may serve as a more meaningful and effective strategy for supporting long-term learning and writing development among ESL undergraduates.

ACKNOWLEDGEMENTS

The authors would like to express their sincere appreciation to all individuals who contributed, directly or indirectly, to the completion of this study. The authors also acknowledge the support and encouragement provided by Quest International University and City University Malaysia, which facilitated the conduct of this research.

DISCLOSURE STATEMENT

The authors report no potential conflict of interest with respect to the research, authorship or publication of this article.

REFERENCES

Abdi, H. (2023). Perception of students about continuous assessment methods and its relation to students' achievement and motivation. *Research Square.* https://doi.org/10.21203/rs.3.rs-3696986/v1

Ahmed, S. K., Mohammed, R. A., Nashwan, A. J., Ibrahim, R. H., Abdalla, A. Q., Ameen, B. M. M., & Khdhir, R. M. (2025). Using thematic analysis in qualitative research. *Journal of Medicine, Surgery, and Public Health, 6*, 100198. https://doi.org/10.1016/j.glmedi.2025.100198

Al-Zoubi, Z., AlKaabi, A., Qablan, A., Bataineh, O., & Issa, H. B. (2024). The impact of work pressure on decision-making effectiveness among department heads in faculties of educational sciences. *PLoS ONE, 19*(8), e0304584. https://doi.org/10.1371/journal.pone.0304584

Armstrong, C. (2022). Key methods used in qualitative document analysis. *SSRN.* https://doi.org/10.2139/ssrn.3996213

Azaki, J. I., & Rivett, U. (2022). Organisational factors affecting knowledge retention in a public organisation. *Proceedings of the 23rd European Conference on Knowledge Management, 23*(2), 1304-1311. https://doi.org/10.34190/eckm.23.2.758

Bernique, A. S. (2020). Lights, camera, anxiety: The spotlight effect, social anxiety and the perception of gaze direction. *Undergraduate Review, 15*(7), 21-37. https://vc.bridgew.edu/undergrad_rev/vol15/iss1/7

Cinkaraa, E., & Au, H. Y. C. (2019). Effects of writing portfolio assessment at tertiary level intensive English program: An action research. *Focus on ELT Journal, 1*(1), 53-69. https://doi.org/10.14744/felt.2019.00006

de Lima, N. K., & Jaeger, A. (2020). The effects of prequestions versus postquestions on memory retention in children. *Journal of Applied Research in Memory and Cognition, 9*(4), 555-563. https://doi.org/10.1016/j.jarmac.2020.08.005

Do, H. M. (2023). Pedagogical benefits and practical concerns of writing portfolio assessment: Suggestions for teaching L2 writing. *Teaching English as a Second or Foreign Language, 27*(1), 1-20. https://doi.org/10.55593/ej.27105a4

Enakrire, R. T., & Smuts, H. (2023). Knowledge retention for enhanced organisational growth in higher education institutions. *Journal of Information & Knowledge Management, 22*(01), 2250054. https://doi.org/10.1142/S021964922250054X

Esposito, A. G. (2025). Learning as retention; an examination of retention of self-derived knowledge through memory integration in a diverse sample of elementary students. *Learning and Individual Differences, 119*, 102655. https://doi.org/10.1016/j.lindif.2025.102655

Farhan, H. (2025). The impact of AI-powered writing tools on students' writing performance: A content analysis and future prospects. *ResearchGate*. https://doi.org/10.13140/RG.2.2.22893.09445

Finley, J. R. (2025). Expanded taxonomies of human memory. *Frontiers in Cognition, 3*, Article 1505549. https://doi.org/10.3389/fcogn.2024.1505549

French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education, 88*, 893-918. https://doi.org/10.1007/s10734-023-01148-z

Gebremariam, H. T., Tessema, M. A., Kahsay, M. T., Mulugeta, Z. A., Ambaye, T. W., & Mulatu, A. T. (2025). Using integrated alternative assessment to explore essay writing skills: An insight into authentic instructional dashboards. *Social Sciences and Humanities Open, 11*, 101571. https://doi.org/10.1016/j.ssaho.2025.101571

Ghanad, A. (2023). An overview of quantitative research methods. *International Journal of Multidisciplinary Research and Analysis, 6*(8), 3794-3803. https://doi.org/10.47191/ijmra/v6-i8-52

Graf, A., Adama, E., Afrifa-Yamoah, E., & Adusei-Asante, K. (2023). Perceived nexus between non-invigilated summative assessment and mental health difficulties: A cross sectional studies. *Journal of Academic Ethics, 21*(4), 609-623. https://doi.org/10.1007/s10805-023-09472-w

Gratchev, I. (2023). Replacing exams with project-based assessment: Analysis of students' performance and experience. *Education Sciences, 13*, 408. https://doi.org/10.3390/educsci13040408

Guran, C. A., Lehmann-Grube, J., & Bunzeck, N. (2020). Retrieval practice improves recollection-based memory over a seven-day period in younger and older adults. *Frontiers in Psychology, 10*, 2997 https://doi.org/10.3389/fpsyg.2019.02997

Handayanie, Y., Rahmat, A., & Priyandoko, D. (2021). Brain dump activities to overcome student's intrinsic cognitive load in reproductive systems online learning. *Jurnal Pengajaran Matematika dan Ilmu Pengetahuan Alam, 26*(2), 8-13. https://doi.org/10.18269/jpmipa.v26i2.37217

Hennink, M., & Kaiser, B. N. (2022). Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Social Science and Medicine, 292*, 114523. https://doi.org/10.1016/j.socscimed.2021.114523

Hoare, A., Osuntade, O. B., & Patel, R. (2025). *Ethical educational leadership*. Thompson Rivers University Open Press. https://openlibrary-repo.ecampusontario.ca/jspui/bitstream/123456789/2359/1/Ethical-Educational-Leadership-1724774937.pdf

Imsa-ard, P., Wichamuk, P., & Chuanchom, C. (2021). Muffled voices from Thai pre-service teachers: Challenges and difficulties during teaching practicum. *Shanlax International Journal of Education, 9*(3), 246-260. https://doi.org/10.34293/education.v9i3.3989

Jopp, R., Pallant, J. L., & Russell, H. (2023). Choose your own adventure: Understanding why students prefer certain types of assessment. *Journal of University Teaching and Learning Practice, 20*(7), 10-14. https://doi.org/10.53761/1.20.7.11

Kaur, J., Maggo, M., Phoolka, S., & Singh, S. (2024). A triad relationship among stress, turnover intention, and performance: A mediation perspective. *The Open Psychology Journal, 17*, e18742501290750. https://doi.org/10.2174/0118743501290750240516053331

Knezevic, E., Nenic, K., Milanovic, V., & Knezevic, N. N. (2023). The role of cortisol in chronic stress, neurodegenerative diseases, and psychological disorders. *Cells, 12*(23), 2726. https://doi.org/10.3390/cells12232726

Kuhn, T., Heisz, J. J., & Middleton, L. E. (2025). The impact of sleep, mental health, and gender on academic performance in Canadian university students. *Frontiers in Education, 10,* 1565920. https://doi.org/10.3389/feduc.2025.1565920

Lee, E. (2020). Evaluating test consequences based on ESL students' perception: An appraisal analysis. *Studies in Applied Linguistics & TESOL at Teachers College, Columbia University, 20*(1), 1-22. https://doi.org/10.7916/salt.v20i1.3394

Lim, W. M. (2024). What is quantitative research? An overview and guidelines. *Australasian Marketing Journal, 33*(2), 199-229. https://doi.org/10.1177/14413582241264622

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J. Ly, A., Gronau, Q. F., Šmíra, M., M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., & Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software, 88*(2), 1-17. https://doi.org/10.18637/jss.v088.i02

Melo, G., Monteza, D., Colson, G., & Zhang, Y. Y. (2022). How to assess? Student preferences for methods to assess experiential learning: A best-worst scaling approach. *PLoS One, 17*(10), e0276745. https://doi.org/10.1371/journal.pone.0276745

Memon, M. A., Thurasamy, R., Ting, H., & Cheah, J. G. (2025). Purposive sampling: a review and guidelines for quantitative research. Journal of Applied Structural Equation Modeling, 9(1), 1-23. https://doi.org/10.47263/JASEM.9(1)01

Neto, J., Neto, F., & Furnham, A. (2022). Predictors of students' preferences for assessment methods. *Assessment & Evaluation in Higher Education, 48*(4), 556-565. https://doi.org/10.1080/02602938.2022.2087860

Nieuwenhuis, S. (2024). Arousal and performance: Revisiting the famous inverted-U-shaped curve. *Trends in Cognitive Sciences, 28*(5), 394-396. https://doi.org/10.1016/j.tics.2024.03.011

Obeiah, S. F., & Bataineh, R. F. (2016). The effect of portfolio-based assessment on Jordanian EFL learners' writing performance. *Bellaterra Journal of Teaching & Learning Language & Literature, 9*(1), 32-46. https://doi.org/10.5565/rev/jtl3.629

Opitz, B., & Kubik, V. (2024). Far transfer of retrieval-practice benefits: Rule-based learning as the underlying mechanism. *Cognitive Research: Principles and Implications*, 9, 65. https://doi.org/10.1186/s41235-024-00598-y

Oranga, J., & Matere, A. (2023). Qualitative research: Essence, types and advantages. *Open Access Library Journal, 10*(12), e11001. https://doi.org/10.4236/oalib.1111001

Ott, V. L., van den Bergh, D., Boutin, B., van Doorn, J., Bartoš, F., Judd, N., van Langen, J., Korthals, L., Kievit, R., Groot, L., & Wagenmakers, E. J. (2025). Information data visualization with raincloud plots in JASP. *Behavior Research Methods, 57,* Article 265, 1-11. https://doi.org/10.3758/s13428-025-02773-5

Plancher, G., & Barrouillet, P. (2020). On some of the main criticisms of the modal model: Reappraisal from a TBRS perspective. *Memory and Cognition, 48*(3), 455-468. https://doi.org/10.3758/s13421-019-00982-w

Playfoot, D., Wilkinson, L. L., & Mead, J. (2023). Is continuous assessment inclusive? An analysis of factors influencing student grades. *Assessment & Evaluation in Higher Education, 48*(7), 938-950. https://doi.org/10.1080/02602938.2022.2150834

Rezigalla, A. A., Abdalla, A. M., Mohammed, H. M., Alhassen, M. M., & Mohammed, M. A. (2017). Students' perceptions toward continuous assessment in anatomy courses. *Journal of Medical Sciences and Health, 3*(2), 5-8. https://doi.org/10.46347/jmsh.2017.v03i02.002

Sörensen, L. K. A., Bohté, S. M., Slagter, H. A., & Scholte, H. S. (2022). Arousal state affects perceptual decision-making by modulating hierarchical sensory processing in a large-scale

visual system model. *PLOS Computational Biology, 18*(4), e1009976. https://doi.org/10.1371/journal.pcbi.1009976

Swathi, M., Ramesh, M., Sathyaseelan, T., Muralikrishnan, S., Pettela, R., & Arun, G. (2025). Factors influencing students' performance in second language writing skills. *World Journal of English Language, 15*(5), 189-207. https://doi.org/10.5430/wjel.v15n5p189

Tenny, S., Brannan, J. M., & Brannan, G. D. (2022). *Qualitative study.* StatPearls. https://www.ncbi.nlm.nih.gov/books/NBK470395/

Wang, Q., Zhang, Y., Kolletar-Zhu, K., Xu, X., & Zhao, L. (2022). Effect of focus on form on vocabulary acquisition in computer-assisted learning. *Mobile Information Systems*, 2022(1), 1-7. https://doi.org/10.1155/2022/4602998

Wixted, J. T. (2024). Atkinson and Shiffrin's (1968) influential model overshadowed their contemporary theory of human memory. *Journal of Memory and Language, 136*, 104471. https://doi.org/10.1016/j.jml.2023.104471

Wright, K. L., Hodges, T. S., Enright, E., & Abbott, J. (2021). The relationship between middle and high school students' motivation to write, value of writing, writer self-beliefs, and writing outcomes. *Journal of Writing Research, 12*(3), 601-623. https://doi.org/10.17239/jowr-2021.12.03.03

Yang, B. W., Razo, J., & Persky, A. M. (2019). Using testing as a learning tool. *American Journal of Pharmaceutical Education, 83*(9), 7324. https://doi.org/10.5688/ajpe7324

Zaabalawi, R. S., & Zaabalawi, J. (2024). Portfolios versus exams: A study to gauge the better student assessment tool. *Language Testing in Asia, 14*, 28. https://doi.org/10.1186/s40468-024-00296-y