Research Journal in Advanced Humanities

RJAH

https://doi.org/10.58256/43mc8f40

Open Access

Check for updates

RESEARCH ARTICLE **Section(s):** *Literature, Linguistics & Criticism*

# AI-assisted academic writing in l2 medical education: Effects on undergraduate medical students' cognitive load and linguistic performance

**Shrowg Alhomaidhi**
Department of English, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia
Email: s.alhomaidhi@psau.edu.sa

## ABSTRACT

This study examines the effects of artificial intelligence (AI) writing tools, specifically ChatGPT, on cognitive load and language skills among undergraduate medical students in Saudi Arabia. Utilizing a counterbalanced within-subject experimental design, 52 students produced two academic papers—one with AI support and one without. Cognitive load was assessed using the NASA task load index (NASA-TLX), while text quality was evaluated through grammatical accuracy, syntactic complexity, and lexical diversity. Results revealed a notable reduction in cognitive load with AI assistance ($p < .001$), particularly in mental demand and effort. Texts generated with AI showed significant improvements in grammatical accuracy (all $p < .01$) but experienced declines in syntactic complexity and lexical diversity (all $p < .01$). A negative correlation between cognitive load and linguistic accuracy was found in the non-AI condition ($r = -.58$, $p < .001$), which disappeared with AI assistance. Overall, while AI tools effectively decreased cognitive load and enhanced grammatical accuracy, they negatively impacted syntactic complexity and lexical diversity. The findings highlight the need for integrating AI writing tools with teaching methodologies that foster deeper language engagement and advanced writing skills.

**KEYWORDS:** AI-assisted writing, cognitive load, linguistic performances, academic writing, medical education, Saudi Arabia, syntactic complexity

**Research Journal in Advanced Humanities**

Volume 7, Issue 3, 2026
ISSN: 2708-5945 (Print)
ISSN: 2708-5953 (Online)

**ARTICLE HISTORY**

Submitted: **03 June 2026**
Accepted: 11 **July 2026**
Published: 13 **August 2026**

**HOW TO CITE**

Alhomaidhi, S. (2026). AI-assisted academic writing in L2 medical education: Effects on undergraduate medical students' cognitive load and linguistic performance. *Research Journal in Advanced Humanities, 7*(3). https://doi.org/10.58256/43mc8f40

**Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited**

## I. INTRODUCTION

The emergence of generative artificial intelligence (AI) writing programmes has resulted in dramatic changes to academic writing in higher education by transforming how students generate, revise, and evaluate written texts (Pereira et al., 2023; Abouelnour et al., 2024). These tools can generate fluent, contextually coherent text across a wide range of genres and disciplines, resulting in their rapid adoption by students who are enrolled in higher education all over the world (Cardon et al., 2023; Nguyen et al., 2024). Although AI tools have the potential to help language learners develop their writing skills, it is well established, there are still concerns about their implications for the cognitive and linguistic processes underlying academic competence, particularly amongst second language (L2) learners. In the case of L2 writers, scholarly writing is characterised by both content and linguistic encoding, which is a dual processing load that places unrealistic demands on working memory resources (Hashemian & Heidari, 2013; Yoon, 2011; Adel, 2026).

In recent years, research on AI-supported writing has grown markedly, in line with the fast uptake of automated text-generation tools within educational contexts (Khater, 2023). Existing studies consistently suggest that AI-assisted writing enhances grammatical accuracy, writing efficiency, and learner motivation. In recent years, scholarly work on writing supported by artificial intelligence has expanded significantly, mirroring the rapid adoption of automated text-generation tools in educational settings. Notwithstanding these purported advantages, an expanding body of research has raised doubts about whether writing supported by artificial intelligence fosters sustainable language development (Adel, 2023). Overreliance on AI tools can dampen the motivation and interest of language learners in engaging with the cognitive and linguistic processes essential for the development of long-term language learning (Barrot, 2023; Kohnke et al., 2023; Su et al., 2023; Warschauer et al., 2023). Medical students form an important population because their academic writing tasks (case reports, clinical summaries, research papers and reflective writing) involve more than just the production of language Adel et al., 2024a). They form the basis for the cognitive skills that support the language precision and accuracy required in professional medical communication, which directly impact on patient safety and the quality of medical care (Ellahham, 2021; Hull, 2016). Other studies show the importance of cognitive, individual factors and contextual in shaping psychological experiences and linguistic (Elhalafaway et al., 2025; Abou Adel et al., 2025; Alhourani et al. 2025; Alenzi et al., 2026a). These perspectives provide a useful basis for examining the effects of AI-assisted writing on cognitive load and linguistic performance in medical education (Khalifa et al., 2025; Alenezi et al., 2026b; Abou Adel et al., 2026; Alrefai et al., 2026). Therefore, assessing the impact of using AI for writing is crucial to determine if these tools are beneficial or detrimental to cognitive and linguistic skills in medical writing (Chami et al., 2026; Adel et al., 2024b).

Cognitive Load Theory (CLT) (Sweller, 1988) serves as a suitable theoretical model to describe how learners allocate limited cognitive resources when facing complex tasks like writing in a classroom (Asif et al., 2025). In CLT, learning is influenced by intrinsic, extraneous and germane cognitive load (Adel, 2022; Adel et al., 2025). While AI-powered writing may reduce cognitive load that is associated with the language production process, it could also reduce the level of active cognitive load learners experience when composing. Previous studies have consistently shown that L2 writers have significantly higher cognitive demands than L1 writers, including the dual demands of language processing and content production. In this regard, Manchón and Polio (2022) have determined that producing L2 writing requires linguistic encoding that requires more cognitive resources than writing L1. Despite these theoretical views, there is no clear indication yet that using AI to write reduces L2 medical students' cognitive load, nor whether it enhances or hinders their language development (Adel, 2019).

Prior studies have indicated that AI-enabled writing may improve writing quality, but there are still considerable drawbacks. There is limited evidence on the impact of using AI for writing on cognitive load in L2 medical writing contexts, and whether cognitive load reduction corresponds to changes in linguistic performance. To fill these research gaps, this study focused on the impact of AI-supported writing on cognitive load and linguistic outcomes of L2 undergraduate medical students in Saudi Arabia. The study sought the following three research questions (RQs):

RQ1. Does AI-assisted writing (ChatGPT) reduce L2 medical students perceived cognitive load compared to non-AI-assisted writing?

RQ2. Does AI-assisted (ChatGPT) writing alter the syntactic complexity and lexical diversity in the medical students' academic writing?

RQ3. What is the relationship between cognitive load and linguistic accuracy across AI-assisted and non-AI conditions?

### *A. Significance of the Study*

There are four major contributions of this study. First, it expands the application of the Cognitive Load Theory by explicitly looking at how the use of AI for writing impacts cognitive load and language performance in L2 medical writing, thereby providing new evidence in a specific educational context. Second, the study uses a multi-dimensional methodological approach, which combines a subjective assessment of the cognitive load in the form of NASA-TLX with linguistic analysis. The analysis includes syntactic complexity, lexical diversity, and linguistic accuracy. Thirdly, the findings have implications for medical educators, highlighting the need to ensure a proper balance between AI-supported writing and independent language production and critical revision. Lastly, the findings of this study can be used to support institutional and educational policies to make responsible use of generative AI in medical education, especially in English-medium education where language accuracy and professional communication are essential.

## II. RESEARCH DESIGN AND METHODOLOGY

### *A. Design*

To enhance the statistical power and minimize the confounds of individual language proficiency, prior AI experience, and background medical knowledge, a within-subjects, counterbalanced experimental design was used for this study. Each participant completed two comparable academic writing tasks under counterbalanced conditions, one with AI assistance and one without, with the order of the writing conditions randomly assigned to minimise potential order and practice effects. Participants were randomly assigned to one of two writing sequences: (a) AI-assisted writing followed by non-AI-assisted writing or (b) non-AI-assisted writing followed by AI-assisted writing. The independent variable was the writing condition (AI-assisted versus non-AI-assisted), whereas the dependent variables were perceived cognitive load, syntactic complexity, lexical diversity, and linguistic accuracy.

### *B. Setting and Participants*

The research recruited 52 Saudi undergraduate medical students who were enrolled in English-medium medical programmes at a tertiary institution in Saudi Arabia. A sample of 52 participants was deemed suitable for the within-subject experimental design, since each participant completed both experimental conditions, which strengthened statistical power and lowered variability between participants. A convenience sampling approach was adopted because the participants were easily reachable to the researchers and satisfied the predetermined inclusion criteria. These criteria were as follows: (1) current enrolment in an undergraduate medical programme, (2) being a non-native English speaker, and (3) having no self-reported learning disabilities that might influence academic writing performance. Participants were recruited from medical students in Years 2 to 4 by means of departmental announcements. Participation was completely voluntary and had no effect on students' academic performance. All subjects gave informed written consent before the beginning of data collection.

### *C. Data Collection Instruments*

The study employed a demographic questionnaire, two paired writing tasks, a standardized AI Assistance Protocol, the NASA Task Load Index (NASA-TLX), a standardized rubric for linguistic accuracy and error coding, and an editing behaviour log to collect data related to participant characteristics, cognitive load, linguistic performance, and AI interaction patterns. The demographic questionnaire included information on age, gender, academic year, first language (L1), self-assessment of English proficiency and previous experience with AI (AI-assisted writing tools). With respect to the writing tasks, the researchers created two academic writing tasks and pre-tested them before the study. Participants had to create 300 - 400 words analytical responses to a clinical scenario for each assignment, which were related to the year of study. The tasks were chosen to conform to the familiarity of the topics, genre, cognitive demands and difficulty, based on ratings collected from an independent pilot sample of five medical students who were not part of the main study. Task A was on the pathophysiology of type 2 diabetes mellitus and Task B was on evidence-based management of community acquired pneumonia. In both tasks, participants were required to make use of their clinical knowledge and formulate a coherent and convincing academic argument. AI Assistance Protocol: ChatGPT was the specified generative AI writing tool for the AI-assisted condition. Each participant was provided with the same written guidelines for responsible use of AI and asked to employ the tool to draft, edit and organize their writing while maintaining their own medical knowledge in the final product. To capture how AI text was handled, an editing behavior log was used to record the frequency and characteristics of AI-generated text that was accepted, modified, or rejected.

**(a) Cognitive Load Measure**

The NASA-TLX was used to assess perceived cognitive load, which is a validated six-dimensional instrument that measures mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart and Staveland, 1988). Participants gave a rating for each dimension from 0-100, and an overall cognitive-load composite score was calculated after each writing task. The NASA-TLX has exhibited high internal consistency and reported Cronbach's α of 80 (Devos et al., 2020).

**(b) Linguistic Analysis**

The written texts were analysed in terms of syntactic complexity, lexical variety and linguistic accuracy. Mean clause length (MCL) and the clausal subordination ratio (CSR) (dependent clauses/total number of T-units) were used as indicators of syntactic complexity. Lexical diversity was calculated with the Moving Average Type-Token Ratio (MATTR) with a window size of 50 tokens in TAALES. To assess linguistic accuracy, a rubric for error coding was developed by the researcher, which counted grammatical, lexical and orthographic errors per 100 words (E/100 W). Interrater reliability was determined by having a second rater independently rate a subsample of 20% of the writing samples, which yielded a Cohen's kappa of .87, indicating good agreement.

*D. Data Collection Procedure*

All data were collected in a controlled laboratory setting for 2 weeks to ensure the same test conditions throughout the study. All participants were asked to participate in one 90-minute data collection session that began with the completion of a demographic questionnaire. The participants then were asked to do two identical academic writing exercises, with one using AI assistance and one without, under counterbalanced conditions. To reduce recall bias, NASA-TLX was completed right after each writing task to measure the perceived cognitive load of the participants. The AI-assisted condition participants also were asked to fill out an editing behaviour log directly after the writing task, documenting the frequency of their accepting, modifying, and rejecting of the AI-generated text. The experimental design and procedures for both experimental conditions were the same to ensure procedural consistency, including testing setting, instructions, and time allotted.

*E. Data Analysis*

Since the within-subject design was based on repeated measurements from the same participants, paired-sample t-tests were used to compare the composite NASA-TLX scores between the AI-assisted and non-AI-assisted writing conditions. Similarly, paired-sample t-tests were used for analyzing linguistic outcomes: Mean Clause Length (MCL), Clausal Subordination Ratio (CSR), Moving Average Type-Token Ratio (MATTR), and Errors per 100 words (E/100W). A Bonferroni correction was used to reduce the risk of the familywise error rate due to multiple comparisons, and the significance level was adjusted to $\alpha = .0125$. Effect sizes were calculated using Cohen's d and the magnitude of the difference in writing conditions between the two conditions quantified. Pearson product-moment correlation analyses were performed to evaluate the relationships between PCL and LA for the separate conditions of writing. Descriptive data such as means and standard deviations were computed for all study variables before the inferential tests. Normality assumptions were tested prior to the principal analyses using the Shapiro-Wilk test, and, when normality assumptions were not met, the Wilcoxon signed-rank test was used as a non-parametric test. IBM SPSS Statistics Version 28 (IBM Corp., Armonk, NY, USA) was used for all analyses. All inferential tests were carried out as two-tailed tests and significance was defined as $\alpha = .05$, unless otherwise stated.

### *F. Ethical Considerations*

Before data collection, this study was approved by the Institutional Research Ethics Committee, Prince Sattam bin Abdulaziz University (Approval No. SCBR-663/2026). The study involved minimal intervention of physical and psychological nature based on the standard academic writing activities and therefore posed minimal risk to participants. All participants gave informed consent in writing and informed of the right to withdraw from the study at any time without adverse academic or personal consequences. Collected data did not contain any personally identifiable information or names. All datasets were anonymised, protected by passwords, and stored securely, with access restricted to the research team. All data were collected and used solely for research and academic publication purposes and handled in accordance with institutional ethical guidelines.

## III. RESULTS

### *A. Participants' Demographics*

In total, 52 Saudi undergraduate medical students participated in the study. The age range was from 19 to 25 years (M = 21.4, SD = 1.8). The sample consisted of students in different years of study; that is, Year 2 (n = 14, 26.9%), Year 3 (n = 22, 42.3%), and Year 4 (n = 16, 30.8%). Arabic was the L1 of most of the participants (n = 49), with three participants reporting having a different L1. The self-rated English proficiency had a mean of 3.4 (SD = 0.8), suggesting intermediate proficiency across the sample. Eighteen respondents (34.6%) rated themselves at or below 3. This was the lower-proficiency subgroup, and 34 (65.4%) rated themselves at 4 or above, forming the high-proficiency subgroup. Previous experience of using AI writing tools was reported by 38 participants (73.1%); 24 participants (46.2%) reported occasional use and 14 participants (26.9%) reported frequent use. Fourteen respondents (26.9%) had no previous experience of using AI writing tools, for more demographic data, see Table 1.

**Table 1**

Mean, Standard Deviation (in Parentheses), and Percentage of Participant Characteristics for AI assisted Group and Non-AI assisted Group (n = 52).

| **Characteristic** | | ***n*** **(%)** | ***M*** **(*****SD*****) or Range** |
|---|---|---|---|
| **Age (years)** | | | 21.4 (1.8); 19-25 |
| **Gender** | Female | 31 (59.6%) | |
| | Male | 21 (40.4%) | |
| **Year of Study** | Year 2 | 14 (26.9%) | |
| | Year 3 | 22 (42.3%) | |
| | Year 4 | 16 (30.8%) | |
| **L1** | Arabic | 49 (94.2%) | |
| | Other | 3 (5.8%) | |
| **Self-Rated English Proficiency** [a] | | | 3.4 (0.8) |
| | Lower proficiency (≤3) | 18 (34.6%) | |
| | Higher proficiency (≥4) | 34 (65.4%) | |
| **Prior Experience of AI Writing Tools** | No prior experience | 14 (26.9%) | |
| | Occasional use | 24 (46.2%) | |
| | Frequent use | 14 (26.9%) | |

### *B. Preliminary Analyses*

The Shapiro-Wilk tests revealed that the NASA-TLX scores and linguistic outcome variables were distributed normally in both conditions (all $p > .05$), supporting the application of parametric tests. Order effects were examined through the comparison of NASA-TLX scores and linguistic outcomes amongst the respondents who completed the AI-assisted task first ($n = 26$) and those who completed it second ($n = 26$). There were no significant order effects (all $p > .20$), indicating the effectiveness of the counterbalancing. Table 2 below presents the descriptive statistics for the primary outcome variables.

**Cognitive load:** A paired-sample t-test established that the participants in the AI-assisted writing condition reported a significantly lower cognitive load ($M$ = 38.4, $SD$ = 9.1) compared to those in the non-AI-assisted condition ($M$ = 61.7, $SD$ = 10.3), with $t$ (51) = 14.82, $p$ < .001, and d = 2.36. This represented a large effect, suggesting that AI assistance had reduced the perceived mental load associated with the writing tasks. Lower scores were reported across all six NASA-TLX subscales in the AI-assisted condition (all $p$ < .001); the largest reductions were reported in the categories of mental demand (Δ = 26.3 points) and effort (Δ = 24.8 points). The frustration subscale also indicated a significant reduction (Δ = 19.7 points), suggesting that AI assistance reduced the affective burden linked to writing in an L2 medical context.

**Table 2**
NASA-TLX Subscale Scores by Writing Condition ($n$ = 52).

| **NASA-TLX Subscale** | **AI-assisted *M* (*SD*)** | **Non-AI *M* (*SD*)** | **Δ** | ***t*(51)** | ***p*** |
|---|---|---|---|---|---|
| Mental Demand | 34.2 (8.6) | 60.5 (9.8) | -26.3 | 18.64 | < .001 |
| Effort | 35.8 (9.2) | 60.6 (10.1) | -24.8 | 17.21 | < .001 |
| Frustration | 31.4 (8.9) | 51.1 (10.4) | -19.7 | 12.88 | < .001 |
| Temporal Demand | 40.1 (9.7) | 58.3 (10.6) | -18.2 | 11.93 | < .001 |
| Performance | 42.3 (10.1) | 62.8 (11.2) | -20.5 | 12.47 | < .001 |
| Physical Demand | 18.6 (7.4) | 22.1 (8.1) | -3.5 | 3.02 | < .001 |
| **Composite Score** | **38.4 (9.1)** | **61.7 (10.3)** | **-23.3** | **14.82** | **< .001** |

*Note.* Δ = Non-AI minus AI-assisted mean. All subscales rated 0-100. All $p$ < .001.

The subgroup analyses revealed that the most pronounced cognitive load reduction was amongst the lower-proficiency participants, with a reported mean reduction of 28.6 NASA-TLX points compared to a reduction of 20.9 points amongst the higher-proficiency participants (see Table 3).

**Table 3**
NASA-TLX Cognitive Load by English Proficiency Subgroup.

| **Proficiency Group** | ***n*** | **AI-assisted *M* (*SD*)** | **Non-AI *M* (*SD*)** | **Δ** | ***t*** | ***p*** |
|---|---|---|---|---|---|---|
| Lower proficiency (≤3) | 18 | 37.1 (8.4) | 65.7 (10.1) | -28.6 | 12.34 | < .001 |
| Higher proficiency (≥4) | 34 | 39.2 (9.5) | 60.1 (10.8) | -20.9 | 9.84 | < .001 |
| **Between-group difference (Δ)** | - | - | - | **-7.7** | **2.14** | **.037** |

*Note.* Lower proficiency n = 18 (self-rated ≤3); Higher proficiency n = 34 (self-rated ≥4). Δ = Non-AI minus AI-assisted mean. A between-group comparison tests whether the magnitude of Δ differs by proficiency group.

**Syntactic complexity and lexical diversity:** Following the application of the Bonferroni correction (adjusted $\alpha$ = .0125), AI-assisted texts revealed significantly lower syntactic complexity compared to non-AI-assisted texts on both measures (see Table 4). MCL was identified as being significantly shorter in AI-assisted texts (M = 9.2, SD = 1.4) compared to non-AI texts ($M$ = 11.8, $SD$ = 2.1), $t$ (51) = 7.34, $p$ < .001, d = 1.46. Similarly, the CSR was less in the AI-assisted condition ($M$ = 0.38, $SD$ = 0.09) than it was in the non-AI condition ($M$ = 0.54, $SD$ = 0.11), $t$ (51) = 8.21, $p$ = .003, d = 1.63. These findings showed that AI-generated writing had shorter, less syntactically complex sentence structures compared to the texts that the students produced independently. Lexical diversity, which was examined using the MATTR, was significantly lower in the AI-assisted texts (M = 0.71, SD = 0.06) than it was in the non-AI texts ($M$ = 0.79, $SD$ = 0.07), $t$ (51) = 6.43, p = .009, d = 1.21. However, AI-assisted texts had significantly lower error rates ($M$ = 3.1 E/100W, $SD$ = 1.2) compared to non-AI texts ($M$ = 8.7 E/100W, $SD$ = 2.4), $t$ (51) = -16.09, $p$ < .001, d = -3.20, suggesting high surface-level linguistic accuracy.

**Table 4**
Cognitive Load and Linguistic Outcome Comparisons Across Writing Conditions ($n$=52).

| Variable | AI-assisted $M$ ($SD$) | Non-AI $M$ ($SD$) | $t$ (51) | $p$ | $d$ | 95% CI ($d$) |
|---|---|---|---|---|---|---|
| NASA-TLX Composite | 38.4 (9.1) | 61.7 (10.3) | 14.82 | < .001*** | 2.36 | [1.85, 2.87] |
| MCL | 9.2 (1.4) | 11.8 (2.1) | 7.34 | < .001*** | 1.46 | [1.01, 1.91] |
| CSR | 0.38 (0.09) | 0.54 (0.11) | 8.21 | .003** | 1.63 | [1.16, 2.10] |
| MATTR | 0.71 (0.06) | 0.79 (0.07) | 6.43 | .009** | 1.21 | [0.79, 1.63] |
| E/100W | 3.1 (1.2) | 8.7 (2.4) | -16.09 | < .001*** | -3.20 | [-3.84, -2.56] |

*Note.* MCL = Mean Clause Length; CSR = Clausal Subordination Ratio; MATTR = Moving Average Type-Token Ratio; E/100W = Errors per 100 Words. Bonferroni-corrected $\alpha = .0125$ applied to linguistic outcomes. d = Cohen's d; 95% CI reported for d. ** $p < .01$. *** $p < .001$.

The editing behaviour logs revealed that the respondents accepted AI-generated text without conducting modifications on 61.4% of occasions, modified AI text on 28.3% of occasions, and rejected AI text on 10.3% of occasions (see Table 5).

**Table 5**
AI Editing Behaviour Log: Frequency of Interaction Types (AI-Assisted Condition).

| Behaviour | Frequency (%) | $n$ (instances) | Description |
|---|---|---|---|
| Accepted without modification | 61.4% | 638 | AI text retained |
| Modified before use | 28.3% | 294 | AI text edited by participant |
| Rejected | 10.3% | 107 | AI output rejected entirely |

*Note.* Total instances = 1,039 AI interactions logged across all participants. Percentages may not sum to 100% due to rounding.

**Cognitive load and linguistic accuracy:** Pearson correlation analyses were conducted to examine the cognitive load and linguistic accuracy across the AI-assisted and non-AI conditions (see Table 6). In the non-AI-assisted condition, the results revealed a significant negative correlation between cognitive load and linguistic accuracy ($r = -.58$, $p < .001$), suggesting that high perceived cognitive load was associated with increased error rates. This relationship was not statistically significant in the AI-assisted condition ($r = -.21$, $p = .14$). A Fisher's z-transformation test confirmed the differences in these correlations ($z = 2.81$, $p = .005$), highlighting meaningful changes in the relationship between cognitive load and accuracy in the conditions.

**Table 6**
Pearson Correlations Between Cognitive Load (NASA-TLX) and Linguistic Accuracy (E/100W) by Condition ($n$=52).

| Condition | $r$ | 95% CI | $p$ | Fisher's $z$ | $p$ ($z$) |
|---|---|---|---|---|---|
| Non-AI-assisted | -.58 | [-.72, -.40] | < .001 | - | - |
| AI-Assisted | -.21 | [-.47, .07] | .14 | - | - |
| **Between-condition comparison** | - | - | - | **2.81** | **.005** |

*Note.* r = Pearson correlation coefficient between NASA-TLX composite and E/100W. Fisher's z tests the significance of the between-condition difference in r. 95% CI computed using Fisher's r-to-z transformation.

A further analysis revealed that the texts with higher lexical diversity and greater syntactic complexity were produced by the participants who made frequent modifications to AI-generated text (high engagement, $n$ = 26) in contrast to those who accepted the texts without any modifications (MATTR M = 0.76 versus 0.66, $p$ = .001) and (MCL M = 10.9 versus 7.5, $p < .001$), respectively. However, there were significant differences in the groups' cognitive load scores, indicating that active AI engagement did not have a significant effect on the cognitive burden when writing and modifying texts.

**Table 7**
Linguistic Outcomes by AI Editing Engagement (AI-Assisted Condition Only).

| Outcome | High Editing *M* (*SD*) | Low Editing *M* (*SD*) | $t(50)$ | $p$ | $d$ |
|---|---|---|---|---|---|
| MATTR | 0.76 (0.05) | 0.66 (0.06) | 6.88 | .001** | 1.85 |
| MCL | 10.9 (1.9) | 7.5 (1.2) | 8.14 | < .001*** | 2.13 |
| E/100W | 3.4 (1.1) | 2.8 (1.3) | 1.96 | .056 | 0.51 |
| NASA-TLX | 37.1 (8.7) | 39.6 (9.5) | -1.04 | .31 | -0.28 |

*Note.* High Editing = above-median AI text modification ($n$ = 26); Low Editing = below-median ($n$ = 26). d = Cohen's d. ** $p < .01$. *** $p < .001$.

**IV. DISCUSSION**

This study showed that writing support assisted by artificial intelligence substantially lowered cognitive load while increasing grammatical accuracy. However, these advantages also came with reduced syntactic complexity and lexical diversity among second-language undergraduate medical students. Overall, the results suggest that artificial intelligence assistance not only decreases the total cognitive demand but also reduces several aspects of mental workload, implying that artificial intelligence operates as a cognitive support tool during academic writing. Texts produced with artificial intelligence support achieved high grammatical accuracy, yet displayed lower syntactic complexity and lexical diversity. In the non-assisted condition, cognitive load and linguistic accuracy were inversely related, whereas no significant relationship emerged in the artificial intelligence-assisted condition. Students who revised the artificial intelligence-generated text submitted written work with high linguistic quality and without added cognitive burden. These findings indicate that actively engaging with artificial intelligence-generated content, rather than passively accepting it, may support more in-depth linguistic processing without increasing cognitive burden.

The very large effect size (d = 2.36) indicates that AI-assisted writing substantially lowered the perceived cognitive demands related to L2 medical writing, showing that AI acted as an effective cognitive support tool throughout complex writing activities. The decreases observed in all six NASA-TLX dimensions, especially mental demand and frustration, imply that AI-assisted writing enhanced task efficiency and reduced the emotional strain involved in producing academic texts in a second language. These findings align with the theory of Cognitive Load Theory (Sweller 1988) that proposes that when extraneous cognitive load can be reduced, then limited cognitive resources may be utilized more efficiently. They also agree with Manchón and Polio's (2022) suggestion that L2 writing requires far more cognitive effort than L1 writing, as writers must concurrently process language and generate content. However, the type of cognition relief needs to be thoughtfully considered. CLT distinguishes between extraneous load, which reduction is important and germane load, which reduction can negatively impact schema formation and long-term skill development (Orhani & Canhasi-Kasemi, 2026; Young et al., 2014). The substantial differences observed in syntactic complexity and lexical diversity of AI-generated texts suggest that AI assistance might potentially bypass productive generative processing (Peng et al., 2025). The proficiency level in which lower-proficiency students exhibited higher cognitive load reductions revealed that AI support can serve as an enabling tool for students facing cognitive challenges, while also potentially raising concerns about overreliance if its use is not gradually phased out (Alrefaie et al., 2026). This interpretation is also supported by recent studies, as evidenced by the cognitive and instructional advantages of introducing AI to writing for students with limited L2 proficiency (Feng, 2025; Zhang et al., 2026).

The present results offer empirical evidence that the use of AI-supported writing enhances the grammatical accuracy of texts and decreases the syntactic complexity and lexical diversity of texts, a relationship that has been studied to a lesser extent in the medical domain. Medical writing requires more than just proper grammar; it should also include sufficient syntax and lexical complexity to express clinical reasoning and facilitate evidence-based argumentation (Jashari, 2026; Park, 2013). The outcome of the students' "acceptance" of the AI-generated texts without revision (61.4%) suggests that learners had fewer opportunities to practice their linguistic autonomy in how they select words and construct sentences, thereby limiting the linguistic scope of their end products.

In contrast, students who actively revised AI-generated text produced compositions that showed significantly greater lexical diversity and syntactic complexity compared to those who copied and pasted the AI-produced text with only minimal revision, with similar cognitive load. The results indicate that the educational benefits of AI go beyond the mere availability of the technology and are highly dependent on the interaction between users and the technology. Teachers should, therefore, foster students' critical thinking and editing skills, not simply having them regurgitate AI-generated texts, to help students develop both their command of language and their ability to write independently. The findings that there was no significant correlation between the cognitive load and linguistic accuracy between the students in the AI-assisted condition and the other conditions agreed with CLT's hypothesis that linguistic accuracy costs would decrease when cognitive load is reduced in writing. In the non-AI condition, however, this correlation revealed a significant process dynamic; when students put a high amount of mental effort into the task, they generated less accurate texts, possibly because less resources were available to monitor language because of the distraction of content production. In the condition involving the use of AI tools for the external management of the linguistic processing that leads to errors, this dynamic process was disrupted not due to the improved cognitive efficiency of the students, but because of the use of AI tools for the external management of the linguistic process that produces errors (Borrego, 2025; Zhao, 2025).

## V. IMPLICATIONS OF THE FINDINGS

The conclusions of this study have important implications for pedagogical, curricular, assessment, and policy considerations when using AI-augmented writing in medical education. The overall evidence supports using a scaffolded approach to the integration of AI into writing development, providing support early on in the writing process and gradually fading it out as students develop greater independence and skills. Thus, the integration of AI tools may commence in the initial stages of students' academic writing trajectories, gradually decreasing and then stopping their reliance on AI as their writing proficiency increases. Lastly, the differences between passive acceptance and active revision of AI-generated text emphasize the importance of incorporating AI literacy into medical writing education. Students should be allowed to use AI writing tools and they should be taught how to interact with them critically when verifying the accuracy of AI-assisted writing, so as to enrich their word bank and help them become more sophisticated in their syntactic capacities. The results further carry significant implications for assessment practices in learning environments that are supported by AI. Therefore, rubrics used to evaluate the quality of writing in AI-enabled environments should not be solely based on grammatical correctness, as they may be jeopardized by the decrease in syntactic complexity and lexical diversity that often occurs. Instead, the scope of assessment should be expanded to include all of the linguistic skills necessary for any medical writing task.

## VI. LIMITATIONS OF THE STUDY

The study has some limitations that should be noted. The students self-reported their English proficiency; this means that their actual ability was not tested. Hence, the need for further research to employ the IELTS as a validated instrument. The editing behaviour log was used to examine the behavioural patterns instead of the underlying cognitive processes. Future studies could completement editing behaviour logs with other protocols (i.e., think-aloud) to capture the underlying cognitive process. Moreover, the sample was drawn from a single institution, thus limiting the generalisability of the findings to students in other medical education contexts in Saudi Arabia.

## VII. CONCLUSION

The findings of this study established that AI assistance reduced cognitive load, particularly for lower-proficiency students, while contributing to the writing of texts that had higher surface accuracy but lower syntactic complexity and lexical richness. The inverse correlation between cognitive load and accuracy amongst unaided students was absent in the AI-assisted condition. These findings suggested that the responsible integration of AI in medical education requires models that support the students' writing development while also preserving their productive cognitive engagement; AI literacy instructions should support the development of critical evaluation and revision, and assessment frameworks should be cantered on a wide range of professional writing competencies.

**DECLARATION OF THE USE OF AI**

Grammarly, for Microsoft Word, was used only to improve the language of the manuscript, including correcting grammatical errors and enhancing clarity. Limited technical assistance was also provided where necessary. AI was not used for developing the main ideas, conducting analysis, or contributing to the core content of this work.

**FUNDING**

The author extends her appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2026/01/41760).

**Biography**

Dr. Shrowg Alhomaidhi is an Assistant Professor of Applied Linguistics at the University of Prince Sattam bin Abdulaziz. Her academic expertise encompasses a range of aspects of Applied linguistics, as well as English language skills.

Email: s.alhomaidhi@psau.edu.sa & https://orcid.org/0000-0002-3010-6653

## REFERENCES


Abouelnour, M. et al. (2024). Interactive communicative discourse in the Holy Qur'an: A deliberative approach. *Edelweiss Applied Science and Technology*, 8(5), 1819–1830. https://doi.org/10.55214/25768484.v8i5.1915

Adel, M. A. et al. (2024a). The Muslim diaspora's nationalism and identity in Kamila Shamsie's Home Fire. *Hamdard Islamicus*, 47(2), 117–135. https://doi.org/10.57144/hi.v47i2.802

Adel, M. A., (2022). Towards an advanced interactive E-learning for the language. *Arab World English Journal (AWEJ)*, Special Issue on CALL, (8), 330-340. https://dx.doi.org/10.24093/awej/call8.22

Adel, M. A., (2019). The semiotics of paratexts in Haifa Betar's novel "Bleeding Dreams": A study in modern literary criticism. *An-Najah University Journal for Research - B (Humanities)*, 33(6), 963–972. https://doi.org/10.35552/0247-033-006-003

Adel, M. A., Mohamed, A. M., & Altwaiji, M. (2024b). The representation of women with disabilities in Shahira Al-Hasan's Cochlea's Secrets. *Journal of International Women's Studies*, 26(6), 8. https://vc.bridgew.edu/jiws/vol26/iss6/8

Adel, M. A. (2023). Aesthetics of religious coexistence and tolerance in Shafak's Turkish novel. *International Journal of Islamic Thought,* 24, 108–118. https://doi.org/10.24035/ijit.24.2023.273

Abou Adel, M. (2026). Fear-Driven Dialogue: A Linguistic Study of Death Knocks and Death. *Journal of Language Teaching and Research*. Vol. 17 No. 3. https://doi.org/10.17507/jltr.1703.07

Abou Adel, M., et al. (2025). Unsettling the Norm: A Posthumanist Reading of Sameness and Differences in Kathryn Erskine's Mockingbird. *Theory & Practice in Language Studies (TPLS)*. https://doi.org/10.17507/tpls.1511.03

Adel, M. et al. (2026). When algorithms write literature: A comparative study of Sally Rooney's Mr Salary and its AI simulation. *Research Journal in Advanced Humanities*, 7(3). https://doi.org/10.58256/ey89ca53

Adel, M. A., Ayrım, O., Alhourani, M., Asif, M., & Altwaiji, M. (2025). Melville in motion: Cultural transformations of Moby-Dick in Arabic-dubbed cartoons. *World Journal of English Language*, 15(4), 292–300. https://doi.org/10.5430/wjel.v15n4p292

Alenezi, M., Elhalafawy, A., Khalifa, S., Ismail, S., Alwethairi, O., Algobaei, F., & Alrefai, A. A. (2026a). Psychological adaptation of the disabled in Sharon Draper's Out of My Mind and Raquel Palacio's Wonder: A comparative post-traumatic study. *Research Journal in Advanced Humanities,* 7(2). https://doi.org/10.58256/p94xmk11

Alenzi, M. et al. (2026b). Beyond the feminine stereotypes: Tracing societal class barriers in Jojo Moyes' Me Before You. *International Journal of English Language and Literature Studies*. Vol. 15 NO. 1. https://doi.org/10.55493/5019.v15i1.5819

Alhourani et al. (2025). Decolonizing the Stage: Exploring Postcolonial Narratives and Identity in Rohina Malik's *Unveiled*. *World Journal of English Language*. https://doi.org/10.5430/wjel.v16n1p95

Alrefai, A. (2026). Intersectionality and marginalization in Viet Thanh Nguyen's The Sympathizer: A study of minority literature. *Research Journal in Advanced Humanities*, 7(3). https://doi.org/10.58256/vczd7298

Alrefaie, Z., Alhazimi, A., Almarabheh, A., Madkhali, T., & Elsamanoudy, A. (2026). AI assisted, mentor-guided narrative review writing task for medical students, a novel educational strategy to enhance research and academic writing. Medical Teacher, 1-10. https://doi.org/10.1080/0142159X.2025.2604240

Asif, M. et al. (2025). Honor killing and the cycle of social influence: A study of the Pakistani bride through Albert Bandura's social learning theory. *Research Journal in Advanced Humanities*, 6 (2). 1-14 https://doi.org/10.58256/pgj52m46

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, 100745. https://doi.org/10.1016/j.asw.2023.100745

Borrego, M. M. (2025). Error Analysis and Artificial Intelligence: Exploring Error Monitoring and Error Treatment Possibilities in EFL. *Revista Brasileira de Linguística Aplicada*, 25(2), e45036. https://doi.org/10.1590/1984-6398202545036

Cardon, P., Fleischmann, C., Aritz, J., Logemann, M., & Heidewald, J. (2023). The challenges and opportunities of AI-assisted writing: Developing AI literacy for the AI age. Business and Professional Communication Quarterly, 86(3), 257-295. https://doi.org/10.1177/23294906231176517

Chami, Y., Mahdy, E., Qutieshat, E., Adel, M. A., & Khater, M. (2026). Legal discourse analysis: Foundations of forensic linguistics and its applications in legislation and litigation. *Research Journal in Advanced Humanities*, 7(2). https://doi.org/10.58256/g09ere43

Devos, H., Gustafson, K., Ahmadnezhad, P., Liao, K., Mahnken, J. D., Brooks, W. M., & Burns, J. M. (2020). Psychometric properties of NASA-TLX and Index of cognitive activity as measures of cognitive workload in older adults. *Brain Sciences*, 10(12), 994.

Elhalafaway, A. et al. (2025). From Text to Performance: A Lehmannian Study of Postdramatic Strategies in Churchill's Escape Alone and Stephens' Light Falls. *Journal of Language Teaching and Research*. V6. pp. 1824-1832. https://doi.org/10.17507/jltr.1606.04

Ellahham, S. (2021). Communication in health care: Impact of language and accent on health care safety, quality, and patient experience. *American Journal of Medical Quality*, 36(5). https://doi.org/10.1097/01.JMQ.0000735476.37189.90

Feng, L. (2025). Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: A comparative study. *Journal of Educational Computing Research*, 62(8), 1741-1774.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P. A. Hancock & N. *Meshkati, Advances in Psychology* (Vol. 52, pp. 139-183).

Hashemian, M., & Heidari, A. (2013). The relationship between L2 learners' motivation/attitude and success in L2 writing. Procedia - *Social and Behavioral Sciences*, 70, 476-489. https://doi.org/10.1016/j.sbspro.2013.01.085

Hull, M. (2016). Medical language proficiency: A discussion of interprofessional language competencies and potential for patient risk. *International Journal of Nursing Studies*, 54, 158-172. https://doi.org/10.1016/j.ijnurstu.2015.02.015

Jashari, A. (2026). Enhancing coherence and persuasiveness in academic discourse through lexical cohesion and critical thinking. *Open Journal of Social Sciences*, 14(1).

Khalifa, S. et al. (2025). Self and Others in the English Translation of Taha Hussein's Al-Ayyām (The Days): A Posthumanist Perspective. *World Journal of English Language. URL:* https://doi.org/10.5430/wjel.v16n2p428

Khater, M. (2023). Refugee children's right to education: The education of Syrian refugees in Jordan—Reality and prospects. *Access to Justice in Eastern Europe*, 6(3), 108–119. https://doi.org/10.33327/AJEE-18-6.3-a000302

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence*, 5, 100156. https://doi.org/10.1016/j.caeai.2023.100156

Manchón, R., & Polio, C. (2022). The Routledge Handbook of Second Language Acquisition and Writing. Routledge New York.

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*, 49(5), 847-864. https://doi.org/10.1080/03075079.2024.2323593

Orhani, S., & Canhasi-Kasemi, E. (2026). Using Sweller's cognitive load theory to improve learning of derivative concepts. *European Journal of Science and Mathematics Education*, 14(2), 187-202

Park, S.-K. (2013). Lexical analysis of Korean university students' narrative and argumentative essays. English Teaching, 68(3).

Peng, J., Zhang, H., Tu, X., Zhang, X., Wu, Q., Wang, Y., & Xiao, D. (2025). Effectiveness of AI-assisted medical education for Chinese undergraduate medical students: a meta-analysis. *BMC Medical Education*, 25(1), 1207. https://doi.org/10.1186/s12909-025-07770-y

Pereira, C. A., Komarlu, T., & Mobeirek, W. (2023). The future of AI-assisted writing. arXiv preprint arXiv:2306.16641.

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing*

*Writing*, 57, 100752.
Sweller, J. (1988). Load during problem solving: 职 On/earn/ng. *Cognitive Science*, 12(2).
Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing*, 62. https://doi.org/10.1016/j.jslw.2023.101071
Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, 28(11), 13943-13967. https://doi.org/10.1007/s10639-023-11742-4
Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes,* 10(3), 130-139. https://doi.org/10.1016/j.jeap.2011.03.003
Young, J. Q., Van Merrienboer, J., Durning, S., & Ten Cate, O. (2014). Cognitive load theory: implications for medical education: AMEE Guide No. 86. Medical Teacher, 36(5), 371-384.
Zhang, Z., Wang, C., & Shang, Y. (2026). Are AI tools effective in reducing cognitive load or boosting writing self-efficacy for L2 postgraduates? The Asia-Pacific Education Researcher, 35(2), 343-354. https://doi.org/10.1007/s40299-025-01031-8
Zhao, D. (2025). The impact of AI-enhanced natural language processing tools on writing proficiency: An analysis of language precision, content summarization, and creative writing facilitation. *Education and Information Technologies,* 30(6), 8055-8086. https://doi.org/10.1007/s10639-024-13145-5