

# CAT Tools in Translator Education: An Experimental Investigation of Efficiency, Quality, and Mental Workload

Original Study

Keman Lai, Zheng Zheng\*, Xiuchun Chen  
School of Foreign Studies, Guangzhou Huali College  
\* Corresponding Author email: research.zheng@outlook.com

Submitted: 28.06.2025; Accepted: 05.02.2026.

**Abstract:** Computer-assisted translation (CAT) tools are increasingly incorporated into translator education, yet evidence on their effects on student translators' performance and perceived workload remains mixed. This study examines how computer-assisted translation (CAT) tool use relates to student translators' efficiency, translation quality, and perceived mental workload in English-to-Chinese translation. Thirty third-year undergraduate students completed two political translation tasks in a within-subjects crossover design, performing one task with a web-based CAT platform and the other without CAT support. Translation efficiency was measured by task completion time from screen recordings. Translation quality was assessed using a specifically designed rubric scored independently by two raters with good to excellent inter-rater reliability. Mental workload was measured immediately after each task using an adapted NASA-TLX. Paired analyses showed that CAT use reduced completion time and increased overall quality scores, while mental workload scores did not differ significantly between conditions. Robustness checks indicated that the time and quality advantages were not explained by period or text allocation, and that treatment effects were not driven by order-dependent carryover, although mental workload showed an order-related pattern. The findings support integrating CAT training into translation curricula through scaffolded practice that builds both tool-use proficiency and translation competence.

**Keywords:** CAT tools, translation efficiency, translation quality, mental workload, English-to-Chinese translation

## 1 INTRODUCTION

In the context of increasing globalisation, translation services have become essential, driven by the growing complexity and volume of multilingual communication (Mazlum et al. 2024). Computer-Assisted Translation (CAT) tools, integrating translation memory (TM), terminology bases (TB), machine translation (MT), and generative artificial intelligence (GenAI), have emerged as critical technologies aimed at enhancing translator productivity and efficiency (Bowker & Fisher 2010; Mones & Sahari 2024). Mastery of CAT tools has become important within translation education to align students' capabilities with industry requirements. However, the specific impacts of CAT tools on student translators'

performance and mental workload in English-to-Chinese translation contexts remain underexplored.

Previous research provides varied insights regarding CAT tools' efficacy. Alotaibi and Salamah (2023) demonstrated improvements in translation efficiency and terminological management with CAT tool assistance, while Skripak et al. (2022) noted drawbacks such as steep learning curves and reduced quality from over-reliance on automation. Some studies suggest CAT tools reduce cognitive load (e.g., Alotaibi & Salamah 2023), while others note increased psychological stress related to complex tool operations, such as Karpińska (2017). These contradictions point to the need for further investigation, particularly for student translators. This study addresses that gap by examining how CAT tools affect efficiency,

translation quality, and mental workload through a within-subjects experiment.

Despite the growing integration of CAT tools in translator training, empirical evidence on how such tools influence student translators' efficiency, translation quality, and mental workload remains limited, particularly in English-to-Chinese translation contexts. To address this gap, the present study examines the effects of CAT tool use on translation efficiency, translation quality, and perceived mental workload through a within-subjects experimental design involving undergraduate translation students.

## **2 ATTRIBUTES OF CAT TOOLS: A CRITICAL REVIEW**

### **2.1 CAT TOOLS ON TRANSLATION EFFICIENCY**

Translation efficiency is a multidimensional construct that has been discussed in the literature in relation to speed, productivity, and effort in the translation process. In empirical studies, however, efficiency is most commonly operationalised in terms of temporal measures, such as task completion time or words translated per unit of time, as these provide an objective and comparable indicator of productivity (Bowker & Fisher 2010; Wang & Gao 2016). In this study, translation efficiency is defined in this established sense and measured by the time taken to complete a translation task.

Many studies have examined how CAT tools impact translation efficiency. Proponents consistently report that these tools can significantly speed up the translation process through features such as translation memory reuse and terminology base management. Lagoudaki (2006), for example, found that translators attribute increased productivity to the ability to reuse previously translated segments, which reduces repeated work. Likewise, Wang (2015) and Zhang (2021) observed that student translators working with CAT tools completed translations more quickly, as functions such as translation memory and automatic formatting that reduce manual effort. Vieira (2014) pointed out that cognitive effort in machine-translated post-editing was significantly correlated with text features (e.g., Meteor scores, prepositional phrases in the source text, and sentence-level type-token ratios), while ST usage was negatively correlated with cognitive effort. This suggests that the cognitive effort of translators can be reduced by reducing the dependence on the source text (e.g., by reducing the use of ST) during the translation process. However, efficiency gains with CAT tools are not guaranteed across all scenarios. Some research suggests that the benefits depend on the translator's familiarity with the tool and the complexity of the text. For instance, O'Brien et al. (2017) found that in certain cases translators were actually less productive when using a CAT tool, possibly due to time spent managing the software or addressing its outputs. Similarly, Guerberof Arenas (2013) reported that post-editing machine translation within a CAT environment was slower than translating from scratch, indicating that in some

workflows the tool can introduce additional procedural burden. These findings suggest that although CAT tools often improve efficiency, their impact can vary and may be limited if users are inexperienced or if the translation task is not suited to tool support.

### **2.2 CAT TOOLS ON TRANSLATION QUALITY**

Translation quality refers to the accuracy, fluency, and adequacy of a translated text in meeting its intended purpose. It is a multi-dimensional concept evaluated by criteria such as correctness of content, clarity of expression, and consistency of terminology. As quality assessment can be subjective and context-dependent (Zheng 2025), the question of whether CAT tools improve or diminish translation quality remains debated in the literature (LeBlanc 2013).

On one hand, research indicates that CAT tools can support certain aspects of translation quality. Garcia (2010) and Lee and Liao (2011) both found that translations produced with TM assistance were judged to be of higher quality, especially in maintaining consistent terminology, compared to those produced without CAT support. In a similar vein, CAT tools are able to store and retrieve previously translated texts and terminology through the functions of Translation Memory (TM) and Terminology Management System (TMS), thus avoiding duplication of efforts. For example, when a translator encounters similar or identical text, the CAT tool can automatically provide suggestions for previous translations, which not only improves efficiency but also reduces the occurrence of repetitive errors (Usmanova et al. 2021; Qatrinnada & Nuraini 2024; Ye 2024). These benefits are most evident in technical texts, where consistency is critical and CAT features such as terminology bases directly contribute to quality. On the other hand, studies have identified several ways CAT tools might negatively affect quality. Wang (2015) argued that the sentence-by-sentence translation mode in many CAT tools can disrupt textual coherence, resulting in target texts that read less smoothly as a whole. Similarly, LeBlanc (2013) observed decline in quality when translators over-relied on TM suggestions without adapting them to context, indicating that blind trust in matches can yield inappropriate translations. Furthermore, Kassem (2021) found that student translators who heavily post-edited machine translations within a CAT tool produced lower-quality outputs than those translating manually, suggesting that over-reliance on the tool may impede the development of translation competence. These concerns highlight that if CAT tools are used uncritically, they may introduce errors or awkward phrasing that a translator might overlook. The mixed findings in research suggest that the effect of CAT tools on quality is context dependent. Karpińska (2017) notes that CAT tools perform better in informational texts characterised by conventionalised terminology, where terminological consistency is a key quality criterion. However, in literary texts, CAT tools may limit the translation process as they tend to provide machine suggestions that may lead to a loss of fluency

and naturalness in the text. It should be noted that the empirical investigation in the present study is restricted to political texts as a form of informational translation, and do not extend to literary or strongly argumentative genres, which may involve different cognitive and stylistic demands. Therefore, CAT tools can amplify efficiency and consistency in suitable contexts, yet translators may need to apply critical analysis for cases where the tool's suggestions might undermine the translation's accuracy.

### 2.3 CAT TOOLS ON MENTAL WORKLOAD

In addition to time and quality outcomes, an important consideration is how CAT tools affect the translator's mental workload, the overall cognitive effort and strain experienced during translation. This study uses the term mental workload in favour of cognitive load to capture not only memory and processing demands, but also the emotional and physical stressors involved in translation tasks. Cognitive load typically refers to the working memory resources needed for a task (Sweller 1988), while mental workload adopts a broader range of factors, including concentration, fatigue, and frustration (Young et al. 2015). By adopting the mental workload perspective, the present study considers the overall burden on the translator, from cognitive processing to emotional stress, when using a CAT tool.

Empirical studies on CAT tool use and translators' mental workload are still emerging, but they reveal both potential benefits and drawbacks. Evidence suggests that CAT tools can lighten the cognitive effort required. However, research on post-editing also indicates a risk of insufficient revision when student translators rely on fluent machine-generated output. For example, Yamada (2019) found a poorer error-correction rate in an NMT post-editing condition despite similar perceived cognitive effort, implying that some errors may be left unaddressed when the output is treated as an acceptable product. At the same time, physiological evidence suggests that post-editing can involve reduced cognitive effort. Lu and Sun (2018) noted that in post-editing, the translator's gaze duration and pupil diameter were smaller than those of human translators, indicating that post-editing reduces cognitive effort. Petrova (2019) noted that CAT tools can help translators manage translation projects and support different types of document processing. This integration of functionality reduces the mental workload on translators when dealing with complex tasks. In addition, Vela et al. (2019) observed neurophysiological signs of lower mental stress, such as more relaxed brain activity, in high-match post-translation edited segments. This suggests that CAT tools can indeed reduce the cognitive load of translation in some cases. Conversely, other studies indicate that CAT tools may sometimes increase mental workload. Karpińska (2017) showed that the complexity and difficulty of using CAT programs may increase the cognitive load of translators, suggesting that using CAT tools may require additional cognitive processing by translators, thus increasing the complexity of translation. Recently, Alves et al. (2016) experimentally pointed

out that participants' lack of familiarity with interactive tools or their reluctance to incorporate the automatically completed changes provided by the CASMACAT workbench produced negative results in terms of time and technical effort. These findings suggest that if a CAT tool is complex or if the user is not fully proficient with it, the tool's use can add to the mental effort rather than reduce it. In sum, while CAT tools have the potential to reduce translators' workload by handling routine tasks, their actual effect on mental workload can vary greatly depending on the usability of the tool and the translator's comfort with it.

### 2.4 RESEARCH GAPS AND QUESTIONS

Despite growing interest in CAT tool use, several research gaps remain, particularly in relation to student translators. Most empirical studies have focused on professional translators, leaving limited evidence on how CAT tools affect translation learners who are still developing linguistic competence, translation strategies, and tool literacy (Al-Batineh & Al Tenaijy 2024). As a result, it remains unclear whether findings from professional contexts can be directly extended to educational settings. In addition, existing studies often examine isolated outcomes, such as efficiency or quality, without jointly considering translation efficiency, translation quality, and mental workload. For student translators, these dimensions are closely interrelated, as improvements in speed or consistency may coincide with changes in cognitive effort or quality-related risks. There is, therefore, a need for empirical research that examines these outcome dimensions together. To address these gaps, the present study focuses on student translators and controls for text type by using political texts as a form of expository translation. The present study addresses the following research questions:

- (1) Does the CAT tool improve translation efficiency in translation practice?
- (2) Can the use of CAT tools in translation practice improve the quality of translation?
- (3) Is the CAT tool effective in reducing mental workload in translation practice?

## 3 METHODOLOGY

### 3.1 RESEARCH DESIGN AND PARTICIPANTS

This study adopted a within-subjects crossover experimental design to compare English-to-Chinese translation performance with and without a computer-assisted translation (CAT) tool. All participants completed translation tasks under both conditions, using a CAT tool vs. using a traditional (non-CAT method), controlling for individual differences in translation proficiency. To minimise order effects, participants were randomly divided into two groups with counterbalanced task orders. One group first translated using the CAT tool and then without it, while the other group followed the opposite order. With this crossover arrangement, each participant translated

one text with CAT assistance and the other without it, ensuring a balanced within-subject comparison. Moreover, a total of 30 participants, 15 per group (Group A and Group B), took part in the experiment. The sample size (N=30) is comparable to prior experimental studies in translator education and translation technology research (e.g., 24~33 student participants) and is consistent with sample sizes commonly used in controlled translation experiments (see Dong & Chen 2025; Koponen et al. 2019; Yang et al. 2023). A semester before this experiment, all participants had completed the foundational course *Translation Theory and Practice*, which assesses students' understanding of translation theories, English and Chinese academic writing skills, as well as basic translation ability. The course scores of the participants were closely clustered (N = 30, mean = 83.77, SD = 3.36, range = 78~90), with a low coefficient of variation (CV = 0.04), indicating relatively homogeneous proficiency and pre-acquired translation skills across the group prior to the experiment.

Each participant provided informed consent prior to participation. Before the experiment, participants had not received formal training in CAT tools as part of their programme or possessed systematic experience using CAT. To prepare them for the CAT condition, they received an approximately two-hour training session on YiCAT, a web-based computer-assisted translation platform (<https://www.yicat.vip/>), conducted in a computer laboratory. The training introduced core CAT functions, including the interface and text segmentation, importing, saving, and managing translation projects, consulting and updating the terminology base, interpreting and applying translation memory matches, editing, confirming, and revising segments, and exporting the final translation. Participants completed short guided practice exercises during the training, including a brief practice translation and targeted tasks such as applying TM matches and querying TB entries. The researchers monitored task completion and addressed questions to make sure that all participants could perform the required operations before the experiment. In addition, the experiment itself was conducted in a controlled translation laboratory environment to promote consistency and reliability. All participants used identical PCs configured with YiCAT for the CAT condition and WPS word processor together with spreadsheet form for the non-CAT condition. The sessions were monitored to achieve consistent experimental conditions and to minimise potential confounding variables.

### 3.2 MATERIALS AND INSTRUMENTS

Two English source texts (Text 1 and Text 2) of comparable difficulty and content were selected. Following the method by Liu et al. (2019) and Sun and Shreve (2014), text difficulty was assessed using Readability Studio (version 2024.0.1). More than twenty candidate texts of the same genre (political discourse) and with closely related topics were first analysed. The final two texts were selected as their results (see TABLE 1) showed the

most similar difficulty profiles across multiple key indices, such as Flesch-Kincaid Grade Level, Flesch Reading Ease (Flesch 1948, 2007), and Raygor Readability Estimate (Raygor 1977), which indicated that both texts corresponded to an advanced university reading level. The mean grade levels were 17.3 and 17.6, respectively, with estimated reader ages of approximately 22~23 years, and only minimal variation across indices. This multi-index analysis confirms that the two texts were highly comparable in linguistic complexity and overall translation difficulty, thus minimising text-related confounds in the within-subjects design.

**TABLE 1.** Content difficulty indices for the two source texts

Readability Index	Text 1	Text 2
Flesch–Kincaid Grade Level	16.4	17.1
Flesch Reading Ease	13	18
Fry	17+	17+
Gunning Fog	19+	19+
Raygor Estimate	17+	17+
SMOG	17.1	18
Average Grade Level (Mean)	17.3	17.6
Estimated Reader Age (Mean)	22.7	23.1

As both source texts belonged to the same domain, a domain-specific term base (TB) was specifically established for this study and applied consistently to both texts, containing standardised English-to-Chinese equivalents for key political terms. In addition, a domain-specific translation memory (TM) specifically compiled for this study from political texts was also made available to participants. The TB and TM was reviewed prior to use to remove duplicate or context-inappropriate entries.

The CAT condition was implemented using YiCAT, a web-based computer-assisted translation platform. As YiCAT is an online system, it does not operate with versioned software releases; the experiment was conducted using its standard institutional interface. Only TM and TB functions were enabled, and no automatic pre-translation was applied. The minimum match threshold followed the platform's default setting (70%). In the non-CAT condition, participants used WPS word processor for translation without CAT support, but with manual access to the same TM and TB content in spreadsheet format. In addition, to ensure experimental control and to isolate the potential effects of other emerging CAT functions, the automatic machine translation, pre-translation functions, external resource searches, and the generative AI assistant, were all disabled. This was done to prevent participants from relying on fully automated output as a finished product and to reduce automation bias, hence ensuring that translation decisions remained primarily human-driven and comparable across both conditions. All translations were completed on the lab computers with relevant tools accessible for both tasks. In addition, built-in screen recorder was used to record translation

activities. This provided a log of translation timing, which assisted in the measurement of task duration.

### 3.3 IDENTIFICATION OF VARIABLES

Translation efficiency was operationalised as task completion time, following established time-on-task practice in translation research (Bowker & Fisher 2010; Wang & Gao 2016). For each participant, the total completion time in the CAT condition was compared against their time in the non-CAT condition, allowing analysis of how the CAT tool impacted productivity. Shorter completion times indicated higher translation efficiency.

The quality of each translated text was assessed using a specifically designed rubric based on Williams' (2004) textual evaluation model. Developing a tailored evaluation framework was necessary as no single standardised translation quality assessment exists that fits all contexts (Zheng 2025). In line with Williams' model, the rubric treated accuracy as the primary assessment criterion. Accuracy was operationalised through the identification of semantic errors (core and general), terminology errors (specific and general), and critical content omissions. Language-related aspects, including wording, style, redundancy, and spelling, were incorporated as supporting dimensions reflecting fluency and acceptability in the target text. Within this framework, core semantic errors were operationalised as mistranslations that misconstrued the source text's meaning or communicative intent in a way that resulted in logically incorrect target-text content, even when surface syntax was acceptable. In contrast, general semantic errors referred to broader misapplications of meaning that affected accuracy at the local level but did not fundamentally change the overall message of the source text. The rubric yields a holistic quality score on a 0~100 scale, with higher scores indicating better overall quality. Scores were derived by applying deductions starting from 100 by error category. Two experienced translation instructors independently evaluated all translations using the rubric. Prior to formal rating, the raters were calibrated through joint discussion of the rubric criteria and sample translations. Final quality scores were calculated as the mean of the two raters' ratings. Inter-rater reliability was assessed using intra-class correlation coefficients (ICC). For translations produced with CAT tools, reliability was high (ICC(1,1) = 0.79; ICC(1,k) = 0.883). Similarly, strong agreement was observed for translations produced without CAT tools (ICC(1,1) = 0.854; ICC(1,k) = 0.921), indicating good to excellent reliability across conditions. When the two raters' scores differed noticeably, the raters revisited the rubric and discussed the case with input from the researchers to ensure consistent interpretation. These discussions were not logged, and no third rater was used. Substantial discrepancies between the two raters were rare in the present study.

Perceived cognitive workload during the translation tasks was measured using the NASA Task Load Index (NASA-TLX) questionnaire (Hart & Staveland 1988). Following Liu et al. (2019) and Sun and Shreve (2014), four

subscales were used: mental demand, effort, frustration, and perceived performance. Each subscale was rated on a 10-point scale (1–10), and the ratings were summed to produce a raw overall workload score for each task (total range = 4~40). The NASA-TLX is a widely used, validated instrument that captures self-reported workload on multiple aspects. However, several limitations have been noted in prior research. For example, Noyes and Bruneau (2007) reported higher scores for a computer-based TLX than for a paper version, indicating that workload ratings can be sensitive to the mode of administration. Furthermore, concerns have been raised about the reliance on self-report, which cannot on its own explain how the task was performed, and about methodological variation in scoring (weighted or raw scores), both of which can affect interpretation across studies (de Waard & Lewis Evans 2014). Despite these limitations, NASA-TLX still offers a practical, low-burden way to obtain comparable mental workload ratings immediately after each translation task, supporting within-participant comparisons between CAT and non-CAT conditions. We acknowledge that reliance on self-reported NASA-TLX scores alone, without concurrent physiological or behavioural measures, is a constraint of the present study. In the experiment, immediately after completing each translation, participants filled out the NASA-TLX, which prompted them to rate their experience on several subscales related to mental effort. These ratings were then combined to produce an overall mental workload score for the task. By administering the questionnaire after each task, the study obtained a pair of workload scores for each participant, one for the CAT-assisted translation and one for the non-assisted translation.

### 3.4 PROCEDURE

The experiment was carried out in two main sessions. In the first session of the experiment, the two groups performed their initial translation tasks under opposite order. Both groups worked simultaneously in the lab under supervision. Participants were instructed to translate as accurately as possible while working at a comfortable pace. They were allowed to pause or edit their work freely; all such actions were captured by the screen-recording software. In the second session, the groups switched tasks and tools to complete the crossover design. This means each participant faced the alternate condition in session 2: those who first used the CAT tool now translated without it, and vice versa, each on the other text. The procedures during this round were identical to session 1, such as same time allocations, environment, and instructions. By the end of session 2, each participant had completed one CAT-assisted and one non-assisted translation, providing paired observations for analysis.

Throughout both sessions, screen-recording software tracked the time each participant took to finish the translation. The recording began when participants started reading the source text and ended when they finalised their translation. The participants were unaware of their exact timing, which was hidden, during the task,

preventing any timing feedback from influencing their work. In addition, immediately after completing each translation task, participants filled out the NASA-TLX workload questionnaire on the computer. The researchers provided the NASA-TLX form as soon as a participant indicated they had finished their translation. Participants rated their mental workload while the experience was fresh in memory. This procedure was repeated after both the session 1 and session 2 tasks, resulting in two separate NASA-TLX submissions per participant. The immediate post-task administration minimised memory decay and ensured that the workload data corresponded closely to the just-completed task. Furthermore, after both rounds were finished, participants submitted their translated texts, which were saved from the lab computers. These translations were then evaluated for quality. The raters used the pre-defined scoring rubric to assess each translation. They worked independently, assigning scores for each translation's quality without being informed whether a given text was produced with or without the CAT tool, thus reducing potential bias in judgment. Their scores for each translation were later averaged to yield a single quality score per translation. Moreover, participants had no access to external resources, except for the content provided through YiCAT (for the CAT condition) and the offline resources (for the non-CAT condition), to ensure that the use of external aids did not confound the results.

#### 4 DATA ANALYSIS

This study employed a quantitative analytical approach to compare outcomes across two translation conditions. For each participant, three paired outcome measures were obtained: translation time with versus without the CAT tool, translation quality score with versus without the CAT tool, and perceived workload measured by the adapted NASA-TLX with versus without the CAT tool. These paired observations were analysed using paired-samples t-tests. The paired t-test is appropriate here because the two sets of scores in each comparison are dependent (each participant's scores in the two conditions form a pair) and the design involves related samples rather than independent groups (Hsu & Lachenbruch 2014). In addition, given the crossover design, in which participants completed two sessions under different conditions with counterbalanced order, analyses were conducted using analysis of variance to account not only for the within-participant condition effect (CAT vs non-CAT), but also for design-related factors such as period and sequence that may influence performance (Grizzle 1965; Langenberg et al. 2023). Specifically, multivariate

ANOVA (MANOVA) was performed for translation time, translation quality, and workload, with Condition (CAT vs non-CAT) as the primary factor of interest. To evaluate potential learning or transfer between sessions, the models additionally included Period (Session 1 vs Session 2) and Sequence (CAT to non-CAT vs non-CAT to CAT), as well as Text allocation, allowing the analysis to test whether the condition effect remained stable after controlling for these factors. A carryover diagnostic was also conducted by computing participant-level difference scores between the two conditions for each dependent variable and comparing these difference scores across the two sequence groups using ANOVA; a non-significant sequence-group difference indicates no evidence that the treatment effect depends on order. All analyses were conducted using SPSSPRO (<https://www.spsspro.com/>), with an alpha level of 0.05. Before running those tests, the distributions of relevant variables were all examined to verify key assumptions, particularly approximate normality. Shapiro-Wilk tests indicated no significant departures from normality ( $p > 0.05$ ), supporting the use of the statistical methods.

### 5 DATA RESULTS

#### 5.1 PAIRED T-TEST RESULTS OF TRANSLATION EFFICIENCY

Participants translated significantly faster with the CAT tool than with the non-CAT methods. As shown in TABLE 2, the average completion time with CAT assistance (T1) was 805.37 seconds (SD = 343.19), compared to 1023.23 seconds (SD = 358.38) without CAT (T2). This 218-second reduction in mean translation time represents about a 21% speed increase. A paired samples t-test confirmed that this difference was statistically significant,  $t = -3.18$ ,  $p = 0.003$ , with a moderate-to-large effect size (Cohen's  $d = 0.58$ ). It should be noted that this effect size reflects a meaningful within-participant improvement in this sample and task setting, but it should not be interpreted as guaranteeing the same magnitude in other populations or text types.

#### 5.2 PAIRED T-TEST RESULTS OF TRANSLATION QUALITY

Using the CAT tool also led to a clear improvement in translation quality as measured by holistic scores from professional raters. As shown in TABLE 3, the average quality score with CAT (on a 0-100 scale) was 86.08 (SD = 3.69), compared to 79.30 (SD = 5.53) without CAT. This was a mean increase of approximately 6.78 points when translations were assisted by CAT. A paired t-test showed that the difference is statistically significant,  $t = 6.28$ ,  $p <$

**TABLE 2.** Paired t-test results of translation efficiency with and without CAT tool

Condition	Mean (SD)	Mean difference	95% CI	t	df	p	Cohen's d
With CAT (T1)	805.37 (343.19)						
Without CAT (T2)	1023.23 (358.38)	217.87	[77.88, 357.85]	-3.18	29	0.003	0.58

**TABLE 3.** Paired t-test results of translation quality scores with and without CAT tool

Condition	Mean (SD)	Mean difference	95% CI	t	df	p	Cohen's d
With CAT (T1)	86.08 (3.69)						
Without CAT (T2)	79.30 (5.53)	6.78	[4.57, 8.99]	6.28	29	<0.001	1.146

Note. For very small probabilities, exact p values are smaller than the reporting precision of the software; these are reported as  $p < .001$

0.001, indicating a large effect ( $d = 1.146$ ). In other words, participants produced substantially higher-quality translations on average when using the CAT tool.

**5.3 PAIRED T-TEST RESULTS OF MENTAL WORKLOAD**

Regarding mental workload, the data showed no significant difference in perceived cognitive load between translating with the CAT tool and without it. As shown in TABLE 4, the NASA-TLX overall workload scores were nearly the same: with CAT, the mean score was 24.67 (SD = 5.23), and without CAT it was 25.30 (SD = 4.69). Higher scores indicate greater perceived workload, so if anything, the CAT condition was slightly lower, but the difference was minimal. The paired t-test confirmed that this difference was not statistically significant,  $t = -0.70$ ,  $p = 0.487$ . The effect size was very small (Cohen's  $d = 0.13$ ), indicating basically no difference in self-reported mental workload between the two conditions.

**5.4 ROBUSTNESS CHECKS: MANOVA AND CARRYOVER DIAGNOSTICS**

To address concerns related to the crossover design and to examine the combined pattern across outcome measures, MANOVA was conducted with Time, TQavg, and NASA as dependent variables and Condition (CAT vs non-CAT), Sequence (CT vs TC), Period (Session 1 vs Session 2), and Text (Text 1 vs Text 2) as factors. As displayed in TABLE 5 and TABLE 6, the multivariate tests showed a significant overall effect of Condition (Wilks'  $\Lambda = 0.645$ ,  $p < 0.001$ ) and Sequence (Wilks'  $\Lambda = 0.830$ ,  $p = 0.019$ ), while Period ( $p = 0.635$ ) and Text ( $p = 0.850$ ) were not significant. Follow-up univariate tests indicated that Condition significantly affected Time ( $F(1,55) = 5.643$ ,  $p = 0.021$ ) and TQavg ( $F(1,55) = 30.288$ ,  $p < 0.001$ ), but not NASA ( $F(1,55) = 0.278$ ,  $p = 0.600$ ). In contrast, Sequence significantly affected NASA ( $F(1,55) = 10.54$ ,  $p = 0.002$ ), but not Time or TQavg.

As an additional carryover diagnostic, participant-level difference scores between CAT and non-CAT

**TABLE 4.** Paired t-test results of NASA-TLX scores with and without CAT tool

Condition	Mean (SD)	Mean difference	95% CI	t	df	p	Cohen's d
With CAT (T1)	24.67 (5.23)						
Without CAT (T2)	25.30 (4.69)	0.63	[-1.21, 2.47]	-0.70	29	0.487	0.13

**TABLE 5.** Multivariate tests: Wilks' Lambda

Effect	Wilks' $\Lambda$	F	p
Condition (CAT vs non-CAT)	0.645	53	<0.001
Sequence (CT vs TC)	0.830	53	0.019
Period (Session 1 vs Session 2)	0.969	53	0.635
Text (Text 1 vs Text 2)	0.985	53	0.850

Note. Dependent variables: Time, TQavg (translation quality average score), and NASA (adapted NASA-TLX total score). Condition = CAT-assisted vs non-CAT (traditional) condition; Sequence = CT (CAT to non-CAT) vs TC (non-CAT to CAT); Period = Session 1 vs Session 2; Text = Text 1 vs Text 2 allocation.

**TABLE 6.** Test results of between-subjects effects

Dependent variable	Condition		Sequence		Period		Text	
	F(1,55)	p	F(1,55)	p	F(1,55)	p	F(1,55)	p
Time	5.643	0.021	0.011	0.915	0.977	0.327	0.501	0.482
TQavg	30.288	<0.001	0.000	0.989	0.437	0.511	0.649	0.424
NASA	0.278	0.600	10.54	0.002	0.55	0.461	0.008	0.929

Note. Time = task completion time; TQavg = mean of two raters' quality scores; NASA = adapted NASA-TLX total score (sum of the four subscales used in this study). Condition = CAT-assisted vs non-CAT condition; Sequence = CT vs TC; Period = Session 1 vs Session 2; Text = Text 1 vs Text 2 allocation.

**TABLE 7.** Difference scores (D) by sequence

Outcome (difference score)	TC (n=15) Mean	TC SD	CT (n=15) Mean	CT SD	Mean difference (TC – CT)	95% CI
Dtime	313.067	427.785	122.667	297.806	190.400	[-86.78, 467.58]
Dtq	7.667	5.457	5.900	6.409	1.770	[-2.69, 6.22]
Dnasa	1.533	5.592	-0.267	4.166	1.800	[-1.90, 5.50]

Note. Sequence groups: TC = non-CAT to CAT; CT = CAT to non-CAT. Participant-level difference scores were computed as: Dtime = Time(TRAD) – Time(CAT), Dtq = TQavg(CAT) – TQavg(TRAD), and Dnasa = NASA(TRAD) – NASA(CAT). Positive D values indicate faster time, higher quality, and lower workload in the CAT condition, respectively. Mean difference is computed as TC – CT; CI = 95% confidence interval for the between-group difference in D.

**TABLE 8.** Comparing difference scores across sequence groups

Difference score	Levene's test p	F	p	Partial $\eta^2$
Dtime	0.130	2.001	0.168	0.067
Dtq	0.376	0.661	0.423	0.023
Dnasa	0.236	1.000	0.326	0.034

Note. Levene's test assesses homogeneity of variance across sequence groups. Partial  $\eta^2$  is reported as an effect-size index for the sequence-group contrast on each difference score.

conditions were computed for each outcome (Dtime, Dtq, and Dnasa) and compared across sequence groups using one-way ANOVA. None of these difference scores differed significantly by sequence (Dtime:  $F = 2.001$ ,  $p = .168$ ; Dtq:  $F = 0.661$ ,  $p = .423$ ; Dnasa:  $F = 1.000$ ,  $p = .326$ ), indicating that the magnitude of the CAT–non-CAT contrast did not depend on whether participants completed the CAT condition first or second. These supplementary analyses support the robustness of the main findings by showing that the observed condition effects on time and quality persist after accounting for sequence, period, and text allocation, with no evidence that treatment effects were driven by order-dependent carryover. (TABLE 7–8.)

## 6 DISCUSSION

### 6.1 TRANSLATION EFFICIENCY

The results of translation efficiency showed that the CAT tool substantially improved translation efficiency. This finding echoes with prior research indicating that CAT tools can streamline the translation workflow by automating repetitive tasks and providing quick access to resources. For example, Bowker (2002) argued that features like translation memory and terminology management allow translators to work faster, and Hutchins (1998) similarly observed that purely manual translation tends to take longer time than technology-assisted translation. The result in the present study is consistent with these observations: translators saved time by adopting the CAT tool's translation memory and term base suggestions, instead of manually searching for resources. In fact, the integration of the YiCAT platform in the translation process appears to have streamlined subtasks, allowing participants to reuse previously translated segments and focus their effort more efficiently.

It is worth noting that our efficiency gain (about 217 seconds on average) is in line with productivity improvements reported in other studies. For instance, Ahsan et al. (2021) in the English-Hindi direction, post-editing reduced the translation time by 63% and the number of keystrokes by 59%, and Garcia (2010) reported up to a 50% increase. The magnitude of our observed effect (Cohen's  $d = 0.58$ ) reinforces the notion that CAT tools can significantly accelerate translation speed for student translators. At the same time, not all researchers have found such favourable outcomes. O'Brien et al. (2017) observed cases in which using translation memory decreased productivity and even harm quality, and Kassem (2021) found that students working with CAT tools were less productive than those using traditional methods. A plausible interpretation is that efficiency benefits depend on tool familiarity and workflow competence, rather than tool availability alone. In the present study, participants received structured training and worked in a controlled lab setting with a relatively accessible interface, which may have reduced tool-handling overhead. From a translator-education perspective, this finding has direct implications for training efficacy. Efficiency gains are more likely when students are trained not only in basic operations (e.g., segmentation, TM matches, and TB consultation), but also in efficient workflow habits, such as deciding whether to accept TM matches, and maintaining term consistency while revising at the discourse level. Therefore, CAT instruction may include guided practice and feedback on workflow decisions, rather than treating CAT use as a purely technical add-on.

### 6.2 TRANSLATION QUALITY

The observed improvement in translation quality supports the view that CAT-assisted translation can reinforce quality through improved terminological consistency

and reduced omissions. This finding aligns with Petrova's (2019) observation that translation memory enhances terminological consistency and accuracy, and with broader accounts suggesting that CAT tools can reduce certain errors by providing structured access to translation resources and consistency support. The results in the present study indicate that the quality gain should not be interpreted as uniform across all aspects of performance. During rubric-based rating, the two raters made notes on meaning-related problems, which were discussed in terms of two broad subtypes: core semantic errors, where the translation misconstrues the source meaning or communicative intent in a way that results in logically incorrect content, and general semantic errors, which involve local meaning imprecision or minor lapses that do not fundamentally alter the overall message. These observations were not based on a separate quantitative count and qualitative coding of error subtypes. Nevertheless, the raters' notes suggested that the CAT-assisted condition occasionally produced high-impact meaning distortions, while the non-CAT condition more often involved minor meaning imprecision. Two anonymised examples are provided in the Appendix to illustrate this contrast.

One plausible explanation is that segment-by-segment processing in CAT environments can encourage sentence-level decision-making, which may increase the likelihood of overly literal renditions that do not fully capture the intended meaning in context. This concern has also been raised by Wang (2015), who argues that sentence-level segmentation can disrupt coherence and affect fluency. In addition, reliance on TM matches or tool-provided prompts may lead some students to accept suggestions too readily without sufficiently checking meaning at the discourse level, a risk also noted in discussions of tool reliance in translator training (e.g., Karapetyan 2023; Yang 2024). In contrast, the CAT environment may still help students avoid certain minor meaning problems by providing structured reference support and terminological guidance. In line with this, the CAT condition showed a slight advantage in terminology-related performance, with fewer terminology errors overall than in the non-CAT condition, suggesting improved terminological consistency when a term base is available.

The findings suggest that CAT tools can improve student translators' overall output quality, but may also be associated with a shift in the types of meaning-related problems that occur. Pedagogically, this supports the inclusion of explicit training that goes beyond tool operation to include revision and critical evaluation strategies, so that students learn to monitor global meaning and discourse coherence rather than relying narrowly on segment-level suggestions.

### 6.3 MENTAL WORKLOAD

Statistically, the results indicate that using the CAT tool did not significantly change student translators' perceived mental workload. Participants reported very similar levels

of mental effort in both the CAT and non-CAT conditions, suggesting that the tool's potential to reduce mental strain was not realised in the present sample, but that it also did not add substantial workload. This pattern can be interpreted in light of automation theory and prior work on translator cognition. Parasuraman and Riley's (1997) framework suggests that introducing automation does not necessarily reduce workload, because any reduction in task demands may be offset by new monitoring and decision-making demands. In the present study, while the CAT tool supported certain operations, it also required participants to interact with the interface and to evaluate and manage tool-provided prompts, which may have balanced out any workload relief. The results are also consistent with Gray et al.'s (2006) observation that perceived workload can remain relatively stable in familiar cognitive tasks as individuals adjust to task demands. Given that participants translated texts of comparable difficulty in both conditions, they may have adapted to the requirements of each workflow.

The literature similarly presents mixed evidence on whether CAT tools lighten or increase translators' cognitive burden. On one hand, Petrova (2019) reported that translators experienced reduced mental workload when using CAT tools, presumably because the integration of functionality helped manage translation projects and support various types of document processing. On the other hand, O'Brien et al. (2017) found that certain translation memory systems increased mental workload, as translators had to invest mental effort in evaluating and editing imperfect suggestions. Our results fall in between these extremes, showing no clear advantage either way. One possible reason is the relative inexperience of our participants with CAT tools. Alotaibi and Salamah (2023) note that students initially face a steep learning curve with new translation technology, which can add stress and effort as they learn the tool. Although participants received introductory training and short guided practice before the experiment, we did not administer a separate questionnaire or interview to measure training outcomes or perceived proficiency in a systematic way. The role of training therefore cannot be tested directly in the present study and should be treated as an interpretive account rather than a demonstrated moderator. Future research could incorporate explicit measures of CAT experience and perceived competence, and combine self-reported workload with process indicators, to test whether increased familiarity reduces workload over time. Pedagogically, these findings suggest that educators should not assume CAT tools alone will lower students' mental workload; rather, a scaffolded introduction of CAT use across courses may help learners build workflow competence and eventually realise any workload-related benefits once the initial learning demands diminish.

### 6.4 ROBUSTNESS AND CROSSOVER EFFECTS

The robustness checks strengthen the main interpretation by showing that the observed CAT advantages in efficiency and quality are not plausibly attributable

to session-related practice or fatigue or to residual text differences. In the MANOVA, Condition remained a significant predictor of the combined outcome pattern, and follow-up tests confirmed that Condition significantly affected Time ( $F(1,55) = 5.643, p = 0.021$ ) and TQavg ( $F(1,55) = 30.288, p < 0.001$ ), while Period and Text were non-significant. This pattern is consistent with established guidance on two-period crossover analysis, where modelling period or sequence helps rule out alternative explanations for treatment effects (Grizzle 1965; Shi & Ye 2024). In other words, the findings indicate that the time and quality improvements associated with CAT use in this experiment appear to be robust within the constraints of the crossover design, rather than ‘artefacts’ of second-session familiarity or uneven text difficulty.

Meanwhile, the robustness analysis highlights a feature that is relevant for interpretation and pedagogy. Sequence showed a significant effect on NASA ( $F(1,55) = 10.54, p = 0.002$ ) even though the CAT vs non-CAT contrast in NASA was non-significant. This indicates that students’ perceived mental workload may depend on whether they used CAT in the first or second session, even though their time and quality outcomes did not show the same order dependence. Such an order-related pattern is not always observed in NASA-TLX studies. For example, prior workload research has reported no order effects in some task contexts, implying that sequence effects may be task- and population-dependent (e.g., Dickinson et al. 1993). One plausible account in the present setting is that student users’ perceived effort may be shaped by early tool-handling demands and adaptation, an issue also consistent with work highlighting sensitivity and individual differences in NASA-TLX responding (e.g., Barajas-Bustillos et al. 2023). Importantly, the carryover diagnostic based on participant-level condition difference scores found no significant sequence-group differences for Dtime, Dtq, or Dnasa, indicating that the magnitude of the CAT–non-CAT contrast did not depend on order and supporting the conclusion that the main treatment effects were not driven by order-dependent carryover. In translator education, the sequence effect on perceived mental workload suggests that the use of CAT tool functions more like a developing skill than an immediate cognitive relief. When CAT is introduced earlier, students may devote extra effort to tool operation, which can offset any workload reduction even if speed and quality improve. This supports a scaffolded approach: begin with short, guided exercises that build procedural fluency with CAT tool functions, then progressively shift to tasks that require stronger judgement and revision at the discourse level, so that tool-handling becomes less effortful and attention can be redirected to meaning and coherence.

## 7 CONCLUSION

This study offers evidence relevant to translator education by examining how CAT tool use relates to student translators’ speed, translation quality, and perceived

mental workload in an English-to-Chinese task. By bringing performance outcomes together with a cognitive indicator, the study helps clarify what students may gain from CAT use in classroom settings and what kinds of demands may remain when they work with tool-supported workflows.

Our empirical evidence indicates that CAT tools provide clear benefits for student translators in terms of efficiency and translation quality, without introducing additional mental workload, even if they do not significantly reduce that burden. Across the three core dimensions examined, translation efficiency, translation quality, and perceived mental workload, the findings reinforce that CAT tools can serve as a valuable asset in translator education by boosting productivity and supporting higher-quality outputs. These advantages align with the objectives of contemporary translator training, where students are expected to manage substantial text volumes while maintaining acceptable quality standards with the aids of technologies (Massey & Ehrensberger-Dow 2011).

Meanwhile, the results point to important pedagogical considerations. Although CAT use was associated with higher overall quality scores, the findings suggest that these gains may coincide with occasional high-impact meaning errors, highlighting the need for students to engage critically with CAT support rather than relying unreflectively on tool-provided suggestions. The anonymised examples provided in the Appendix illustrate these meaning-related problems and reinforce the need to develop students’ metalinguistic monitoring of meaning and intent beyond segment-level matches. Moreover, the absence of a clear reduction in mental workload, together with the observed order-related pattern in workload, suggests that the cognitive demands of CAT use may depend on when and how the tool is introduced within a learning sequence. This implies that simply introducing technology is not a substitute for skill development. On the other hand, students must develop both strong translation competence and familiarity with CAT-mediated workflows and operations to benefit fully.

The findings in the present study also contribute to the understanding of how CAT tools can be integrated into translator workflows and curricula. By leveraging the efficiency and consistency advantages of CAT tools while addressing their limitations through structured training, educators can help student translators strike a balance between speed, quality, and mental effort. Accordingly, translation programmes may benefit from embedding CAT training across multiple stages of instruction, using guided and project-based tasks that encourage both efficient tool use and careful revision for meaning. Assessment practices may likewise be expanded to evaluate not only translation outputs, but also students’ workflow decisions and their ability to apply critical judgement when working with CAT-supported translation environments.

The present study is subject to several limitations. First, the implications are most directly relevant to similar undergraduate translator training contexts and political text tasks, and should be applied cautiously to other

learner populations, genres, or CAT configurations. Second, the small sample size of 30 third-year undergraduate students, each having received about two hours of CAT tool training, may not fully represent the broader population of translation students, and a more diverse cohort might yield different results. No a priori power analysis was conducted, so the sample size was determined pragmatically and may limit the strength of inferences, particularly for small effects. Although participant homogeneity was supported by similar performance in a foundational translation course, no independent standardised language proficiency test was administered, which may limit the generalisability of the findings. Third, no separate pre-test or standardised self-report scale of prior CAT experience was administered; however, prior CAT exposure was minimal in this cohort. Future studies could include a short CAT experience questionnaire and incorporate it as a covariate. In addition, the focus on a single CAT tool, YiCAT, limits the generalisability of the findings, as other tools with different functionalities could influence efficiency, quality, and mental workload in distinct ways. Moreover, while the translation quality scoring strived to ensure comparability across conditions, finer-grained error categorisation was not subjected to separate reliability testing, which should be addressed in future research. Semantic error subtypes were discussed on the basis of raters' qualitative notes during scoring rather than a separate, systematically quantified error-coding procedure with dedicated reliability checks. The present study did not conduct a systematic discourse-level analysis, such as cohesion and coherence measures, which could be explored in future research to clarify text-level impacts of segment-based workflows. Finally, the reliance solely on self-reported measures of mental workload using the NASA-TLX, without supplementary objective data such as eye-tracking or keystroke logging, may affect the accuracy of the workload estimates. It should also be noted that only a subset of NASA-TLX subscales was employed and that workload scores were based on raw, unweighted ratings. In short, these limitations indicate that the findings should be interpreted cautiously and warrant further investigation with larger, more diverse samples, additional tools and text types, and more rigorous process and discourse-level measures.

### ACKNOWLEDGEMENTS

The authors would like to thank the student participants for their time and engagement in the experiment. We are also grateful to the two raters, who served for the quality assessment and provided valuable professional insights during the process. We appreciate the constructive comments from the anonymous reviewers, which have helped improve the clarity and rigour of this study.

### FUNDING

This research was funded by the Quality Engineering Project of the Guangdong Provincial Department of Education for 2021 – “Translation Course Teaching and Research Office of Guangzhou Huali College” (Yue Jiao

Gao Han [2021] No. 29) [2021 Nian Guangdong Sheng Jiaoyuting Zhiliang Gongcheng Xiangmu “Guangzhou Huali Xueyuan Fanyi Kecheng Jiaoyanshi” (Yue Jiao Gao Han 2021, 29Hao)]; and, the 2025 Smart Course Construction Project of Guangzhou Huali College for the English-Chinese/Chinese-English Translation course (Guangzhou Huali Yuan Jiao Zi 2025 24Hao Xuhao1) [Guangzhou Huali Xueyuan 2025 Nian Zhihui Kecheng Jianshe Xiangmu, Yinghan/Hanying Biyi, (Guangzhou Huali Yuan Jiao Zi 2025 24Hao Xuhao1)].

### ETHICS APPROVAL

The ethics approval was obtained from the research department of Guangzhou Huali College on 1st November 2024.

### DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author.

### DATA AVAILABILITY STATEMENT

To protect participant confidentiality, only the datasets used for the analyses reported in this study are publicly accessible. The datasets are available on the Open Science Framework (OSF) at: [https://osf.io/szajn/overview?view\\_only=9c32a82953ef4bc1b5413aafba323851](https://osf.io/szajn/overview?view_only=9c32a82953ef4bc1b5413aafba323851)

### REFERENCES

- Ahsan, A., Mujadia, V., Sharma, D. M., 2021. Assessing post-editing effort in the English Hindi direction. *arXiv preprint arXiv:2112.09841*, available at: < <https://doi.org/10.48550/arXiv.2112.09841> >.
- Al-Batineh, M., Al Tenaijy, M., 2024. Adapting to technological change: An investigation of translator training and the translation market in the Arab world. *Heliyon*, 10(7), available at: < <https://doi.org/10.1016/j.heliyon.2024.e28535> >.
- Alotaibi, H., Salamah, D., 2023. The impact of translation apps on translation students' performance. *Education and Information Technologies*, 28(8), 10709-10729, available at: < <https://doi.org/10.1007/s10639-023-11578-y> >.
- Alves, F., Koglin, A., Mesa-Lao, B. et al., 2016. Analysing the impact of interactive machine translation on post-editing effort. In *New directions in empirical translation process research: Exploring the CRITT TPR-DB* (pp. 77-94), available at: < [https://link.springer.com/chapter/10.1007/978-3-319-20358-4\\_4](https://link.springer.com/chapter/10.1007/978-3-319-20358-4_4) >.
- Barajas-Bustillos, M. A., Maldonado-Macías, A. A., Serrano-Rosa, M. A., et al., 2023. Impact of experience on the sensitivity, acceptability, and intrusiveness of two subjective mental workload techniques: The NASA-TLX and workload profile. *WORK: A Journal of Prevention, Assessment & Rehabilitation*, 75(4), 1265–1275, available at: < <https://doi.org/10.3233/WOR-211324> >.
- Bowker, L., 2002. *Computer-aided translation technology: A practical introduction*. University of Ottawa Press.
- Bowker, L., Fisher, D., 2010. *Computer-aided translation*.

- In *Handbook of translation studies* (Vol. 1, pp. 60-65). John Benjamins Publishing, available at: < <https://benjamins.com/online/hts/articles/comp2?srsId=AfmBOoo6mBoJUj7d-C77lqigkj2gsYMMEMFWGp3NH7NWAWKiSPKGI59e> >.
- de Waard, D., Lewis Evans, B., 2014. Self-report scales alone cannot capture mental workload: A reply to De Winter, Controversy in human factors constructs and the explosive use of the NASA-TLX: A measurement perspective. *Cognition, Technology & Work*, 16(3), 303–305, available at: < <https://doi.org/10.1007/s10111-014-0277-z> >.
- Dickinson, J., Byblow, W. D., Ryan, L. A., 1993. Order effects and the weighting process in workload assessment. *Applied Ergonomics*, 24(5), 357–361, available at: < [https://doi.org/10.1016/0003-6870\(93\)90075-K](https://doi.org/10.1016/0003-6870(93)90075-K) >.
- Dong, D., Chen, M. L., 2025. Metacognitive strategies in translation: a comparative study of student and professional translators. *Humanities and Social Sciences Communications*, 12(1), 1–9, available at: < <https://doi.org/10.1057/s41599-025-05289-7> >.
- Flesch, R., 1948. A new readability yardstick. *Journal of Applied Psychology*, 32(3), 221, available at: < <https://doi.org/10.1037/h0057532> >.
- Flesch, R., 2007. Flesch-Kincaid readability test. Retrieved October, 26(3), 2007.
- Guerberof Arenas, A., 2013. What do professional translators think about post-editing? *The Journal of Specialised Translation*, (19), 75-95, available at: < <https://doi.org/10.26034/cm.jostrans.2013.424> >.
- Garcia, I., 2010. Is machine translation ready yet? *Target: International Journal of Translation Studies*, 22(1), 7-21, available at: < <https://doi.org/10.1075/target.22.1.02gar> >.
- Gray, W. D., Neth, H., Schoelles, M. J., 2006. The functional task environment. In A. F. Kramer, D. A. Wiegmann, & A. Kirlik (Eds.), *Attention: From theory to practice* (pp. 100-118). Oxford University Press, available at: < <https://nbn-resolving.de/urn:nbn:de:bsz:352-284116> >.
- Grizzle, J. E., 1965. The two-period change-over design and its use in clinical trials. *Biometrics*, 21(2), 467–480, available at: < <https://www.jstor.org/stable/2528104> >.
- Hsu, H., Lachenbruch, P. A., 2014. Paired t test. *Wiley StatsRef: statistics reference online*, available at: < <https://doi.org/10.1002/9781118445112.stat05929> >.
- Hart, S. G., Staveland, L. E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139-183, available at: < [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9) >.
- Hutchins, J., 1998. The origins of the translator's workstation. *Machine Translation*, 13(4), 287-307, available at: < <https://doi.org/10.1023/A:1008123410206> >.
- Kassem, M. A. M., 2021. The effect of utilizing CAT technology on English majors' translation and motivation. *Asian EFL Journal*, 28, 135-155, available at: < <https://psau.sa.elsevierpure.com/en/publications/the-effect-of-utilizing-cat-technology-on-english-majors-translat/> >.
- Karapetyan, M., 2023. Teaching languages in the digital age: incorporating machine translation. *Translation Studies: Theory and Practice*, 3(2), 58-69, available at: < <http://doi.org/10.46991/TSTP/2023.3.2.058> >.
- Karpińska, P., 2017. Computer Aided Translation-possibilities, limitations and changes in the field of professional translation. *The Journal of Education, Culture, and Society*, 8(2), 133-142, available at: < <http://doi.org/10.15503/jecs20172.133.142> >.
- Koponen, M., Salmi, L., Nikulin, M., 2019. A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation*, 33(1), 61–90, available at: < <https://doi.org/10.1007/s10590-019-09228-7> >.
- Lagoudaki, E., 2006. Translation Memories Survey 2006: Users' Perceptions Around TM Use. In *Translating and the Computer 28* (pp. 1-29). London, available at: < <https://aclanthology.org/2006.tc-1.2.pdf> >.
- Langenberg, B., Janczyk, M., Koob, V., et al., 2023. A tutorial on using the paired t test for power calculations in repeated measures ANOVA with interactions. *Behavior Research Methods*, 55(5), 2467–2484, available at: < <https://doi.org/10.3758/s13428-022-01902-8> >.
- LeBlanc, M., 2013. Translators on translation memory (TM). Results of an ethnographic study in three translation services and agencies. *Translation & Interpreting*, 5(2), 1-13, available at: < <https://www.trans-int.org/index.php/transint/article/view/228> >.
- Lee, J., Liao, P., 2011. A Comparative Study of Human Translation and Machine Translation with Post-editing. *Compilation and Translation Review*, 4(2), 105-149, available at: < [https://openurl.ebsco.com/EPDB%3Agcd%3A2%3A1576482/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A67659290&crl=c&link\\_origin=scholar.google.com](https://openurl.ebsco.com/EPDB%3Agcd%3A2%3A1576482/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A67659290&crl=c&link_origin=scholar.google.com) >.
- Liu, Y., Zheng, B., Zhou, H., 2019. Measuring the difficulty of text translation: The combination of text-focused and translator-oriented approaches. *Target: International Journal of Translation Studies*, 31(1), 125–149, available at: < <https://doi.org/10.1075/target.18036.zhe> >.
- Lu, Z., Sun, J., 2018. ren2 gong1 fan1 yi4 he2 yi4 hou4 bian1 ji2 zhong1 ren4 zhi1 jia1 gong1 de yan3 dong4 shi2 yan4 yan2 jiu1 [An eye-tracking experimental study of cognitive processing in human translation and post-editing]. *wai4 yu3 jiao4 xue2 yu3 yan2 jiu1 [Foreign Language Teaching and Research]*, 50(5), 760-769, 801, available at: < [https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf\\_q\\_gR\\_iJ3ioNm1kTlx-LqO5zpn6XrUBGlrGS16\\_ykg2tWQbdBILQoZQ5TX-P2cgH1PboLoX-5Ao\\_2vF-tyU8RamuTc7k4G57rLHER\\_LnXmZibW2jmVY-OqRysOVfF0Jw-4WK3ka-39ZMhX\\_vFKxdjXp9XylHnIQ\\_AKWgeVI5xA8Sg5Joukg=&uniplatform=NZKPT&language](https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf_q_gR_iJ3ioNm1kTlx-LqO5zpn6XrUBGlrGS16_ykg2tWQbdBILQoZQ5TX-P2cgH1PboLoX-5Ao_2vF-tyU8RamuTc7k4G57rLHER_LnXmZibW2jmVY-OqRysOVfF0Jw-4WK3ka-39ZMhX_vFKxdjXp9XylHnIQ_AKWgeVI5xA8Sg5Joukg=&uniplatform=NZKPT&language) >.

- Massey, G., Ehrensberger-Dow, M., 2011. Technical and instrumental competence in the translator's workplace: Using process research to identify educational and ergonomic needs. *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, 14, available at: < <https://doi.org/10.4000/ilcea.1060>>.
- Mazlum, F., Shadman, N., Khademnabi, M. M., 2024. Need for Translation Services in the Globalized World: A Periphery-Informed Explanation. *Sustainable Multilingualism/Darnioji daugiakalbyst*, 24, 236-256, available at: < <https://doi.org/10.2478/sm-2024-0009>>.
- Moneus, A. M., Sahari, Y., 2024. Artificial intelligence and human translation: A contrastive study based on legal texts. *Heliyon*, 10(6), available at: < <https://doi.org/10.1016/j.heliyon.2024.e28106>>.
- Noyes, J. M., Bruneau, D. P. J., 2007. A self-analysis of the NASA-TLX workload measure. *Ergonomics*, 50(4), 514-519, available at: < <https://doi.org/10.1080/00140130701235232>>.
- O'Brien, S., Ehrensberger-Dow, M., Hasler, M., Connolly, M., 2017. Irritating CAT tool features that matter to translators. *Hermes: Journal of Language and Communication in Business*, 56, 145-162, available at: < <https://doras.dcu.ie/23262/>>.
- Parasuraman, R., Riley, V., 1997. Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253, available at: < <https://doi.org/10.1518/001872097778543886>>.
- Petrova, V., 2019. Translation quality assessment tools and processes in relation to CAT tools. In *Proceedings of the Human-Informed Translation and Interpreting Technology Workshop (HiT-IT 2019)* (pp. 89-97), available at: < <https://aclanthology.org/W19-8711.pdf>>.
- Qatrinnada, S., Nuraini, K., 2024. Voice of Automatic Translation Work Versus Transcreation Work. *Journal La Sociale*, 5(3), 869-881, available at: < <https://doi.org/10.37899/journal-la-sociale.v5i3.1199>>.
- Raygor, A. L., 1977. The Raygor readability estimate: A quick and easy way to determine difficulty. In *Reading: Theory, Research, and Practice*, 259-263.
- Shi, D., Ye, T., 2024. Behavioral carry-over effect and power consideration in crossover trials. *Biometrics*, 80(2), available at: < <https://doi.org/10.1093/biomtc/ujae023>>.
- Skripak, I. A., Shatskaya, A. V., Ukhanova, E. V. et al., 2022. Information technologies and language: The impact of CAT systems on improving the efficiency of translators' training. *Theory and Practice in Language Studies*, 12(11), 2358-2364, available at: < <https://doi.org/10.17507/tpls.1211.16>>.
- SPSSPRO (n.d.). SPSSPRO: Online statistical analysis platform [Online software], available at: < <https://www.spsspro.com/>>.
- Sun, S., Shreve, G. M., 2014. Measuring translation difficulty: An empirical study. *Target: International Journal of Translation Studies*, 26(1), 98-127, available at: < <https://doi.org/10.1075/target.26.1.04sun>>.
- Sweller, J., 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285, available at: < [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)>.
- Usmanova, Z. A., Zudilova, E. N., Arkatov, P. A., Vitkovskaya, N. G., Kravets, E. V., 2021. Impact of computer-assisted translation tools by novice translators on the quality of written translations. *Laplage em Revista*, 7(Extra C), 714-721, available at: < [https://www.researchgate.net/publication/353469289\\_Impact\\_of\\_computer-assisted\\_translation\\_tools\\_by\\_novice\\_translators\\_on\\_the\\_quality\\_of\\_written\\_translations](https://www.researchgate.net/publication/353469289_Impact_of_computer-assisted_translation_tools_by_novice_translators_on_the_quality_of_written_translations)>.
- Vela, M., Pal, S., Zampieri, M., Naskar, S. K., van Genabith, J., 2019. Improving CAT tools in the translation workflow: New approaches and evaluation. *arXiv preprint arXiv:1908.06140*, available at: < <https://doi.org/10.48550/arXiv.1908.06140>>.
- Vieira, L. N., 2014. Indices of cognitive effort in machine translation post-editing. *Machine Translation*, 28(3), 187-216, available at: < <https://doi.org/10.1007/s10590-014-9156-x>>.
- Wang, S., 2015. Qian3 tan4 ji4 suan4 ji1 fu3 zhu4 fan1 yi4 gong1 ju4 de you1 shi4 yu3 lie4 shi4 [A brief discussion on the advantages and disadvantages of computer-aided translation tools]. *Hai4 wai4 Ying1 yu3 [Overseas English]*, (19), 100-101, available at: < [https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf\\_oQkAjoZeyo4HQEbFLYDpALkPhWh6gLqE1BMLQkN8DWdqD4ucuLiV6MR5M5ANQB2v10E miuWmsnnKl6K816wJipKp9MgLLc082b\\_3jSnV9DxH98tgMn4392dwZPB-d1WC1urg-a4TnOmFRqYj-SgLikk0MdIMP04I-HMFBpfb13yw==&uniplatform=NZKPT&language](https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf_oQkAjoZeyo4HQEbFLYDpALkPhWh6gLqE1BMLQkN8DWdqD4ucuLiV6MR5M5ANQB2v10E miuWmsnnKl6K816wJipKp9MgLLc082b_3jSnV9DxH98tgMn4392dwZPB-d1WC1urg-a4TnOmFRqYj-SgLikk0MdIMP04I-HMFBpfb13yw==&uniplatform=NZKPT&language)>.
- Wang, X., Gao, T., 2016. Qian3 tan4 ji4 suan4 ji1 fu3 zhu4 fan1 yi4 gong1 ju4 dui4 yi4 wen2 zhi4 liang4 de you1 hua4 zuo4 yong4 [A brief discussion on how computer-aided translation tools enhance translation quality]. *Hai4 wai4 Ying1 yu3 [Overseas English]*, (9), 125-126, available at: < [https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf\\_oK\\_WYE10qVFfnXhaBS6EsWPRVzb4E8FaAHfOVKTXzNBj1LO-VaYyjn7s1fBG1sJZzEk5AMzAMNi-KxDwWzTpnY9wpqWdP7zxmsloc\\_Cq1\\_tWiP1hHEjZWnyGiN8qfEbbXWBtjeVNFxjXudwUBKegD7fa-Ka9TYeJrmsftlGj8NDOg==&uniplatform=NZKPT&language](https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf_oK_WYE10qVFfnXhaBS6EsWPRVzb4E8FaAHfOVKTXzNBj1LO-VaYyjn7s1fBG1sJZzEk5AMzAMNi-KxDwWzTpnY9wpqWdP7zxmsloc_Cq1_tWiP1hHEjZWnyGiN8qfEbbXWBtjeVNFxjXudwUBKegD7fa-Ka9TYeJrmsftlGj8NDOg==&uniplatform=NZKPT&language)>.
- Williams, M., 2004. *Translation quality assessment: An argumentation-centered approach*. University of Ottawa Press, available at: < <https://www.jstor.org/stable/j.ctt1ckphnh>>.
- Yamada, M., 2019. The impact of Google Neural Machine Translation on Post-editing by student translators. *The Journal of Specialised Translation*, (31), 87-106, available at: < <https://doi.org/10.26034/cm.jostrans.2019.178>>.
- Yang, A., 2024. On the influence of artificial intelligence on foreign language learning and suggested learning strategies. *International Journal of Education and*

- Humanities*, 4(2), 107-120, available at: < [https://doi.org/10.58557/\(ijeh\).v4i2.214](https://doi.org/10.58557/(ijeh).v4i2.214)>.
- Yang, Y., Liu, R., Qian, X., Ni, J., 2023. Performance and perception: machine translation post-editing in Chinese-English news translation by novice translators. *Humanities and Social Sciences Communications*, 10(1), 1–8, available at: < <https://doi.org/10.1057/s41599-023-02285-7> >.
- Ye, F., 2024. Research on the application of information technology in English translation teaching in colleges and universities. *The Educational Review, USA*, 8(3), available at: < <http://dx.doi.org/10.26855/er.2024.03.021>>.
- YiCAT (n.d.). YiCAT: Web-based computer-assisted translation platform [Online software], available at: < <https://www.yicat.vip/>>.
- Young, M. S., Brookhuis, K. A., Wickens, C. D., Hancock, P. A., 2015. State of science: Mental workload in ergonomics. *Ergonomics*, 58(1), 1-17, available at: < <https://doi.org/10.1080/00140139.2014.956151>>.
- Zhang, L., 2021. Ji4 suan4 ji1 fu3 zhu4 fan1 yi4 yu3 da4 xue2 sheng1 fan1 yi4 neng2 li4 de guan1 xi4 [The relationship between computer-aided translation and university students' translation competence]. *Hai4 wai4 Ying1 yu3 [Overseas English]*, (7), 176-177, available at: < [https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf\\_pjKfK5LBNewxoE3bmdKw9KZh-Tb\\_zyFD111\\_qvhv9c9R0gCOnbQZ3maxg-ds2XObrv5-xQr0hnft1MCnwezTyQAoFEI5EH5jMxwmR5FB3lqmYNBt5btuQoj6ERYdABhLeBRRBJckQAagF5B3nJf1n72V4NnshYzVjgsHPolCPZg==&uniplatform=NZKPT&langug](https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf_pjKfK5LBNewxoE3bmdKw9KZh-Tb_zyFD111_qvhv9c9R0gCOnbQZ3maxg-ds2XObrv5-xQr0hnft1MCnwezTyQAoFEI5EH5jMxwmR5FB3lqmYNBt5btuQoj6ERYdABhLeBRRBJckQAagF5B3nJf1n72V4NnshYzVjgsHPolCPZg==&uniplatform=NZKPT&langug)>.
- Zheng, Z., 2025. From bilingualism to professionalism: The effect of bilingual dominance and professional experience on translation quality in Chinese-to-English translation. *Cogent Arts & Humanities*, 12(1), available at: < <https://doi.org/10.1080/23311983.2025.2465936>>.