

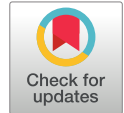
İstanbul Üniversitesi Çeviribilim Dergisi Istanbul University Journal of Translation Studies

Research Article | Araştırma Makalesi

Open Access | Açık Erişim

Measuring the Usability of DeepL: Impact of Translation Technology Courses in Translator Training

DeepL'in Kullanılabilirliğinin Ölçülmesi: Çevirmen Eğitiminde Çeviri Teknolojisi Derslerinin Etkisi



Evren Barut¹ 

¹ Afyon Kocatepe University, Faculty of Science and Literature and Department of Translation Studies, Afyonkarahisar, Türkiye,

Abstract

Neural machine translation (NMT) tools, such as DeepL, are becoming increasingly visible in Translation education. Students are experiencing these tools in the translation process. However, the impact of translation technology courses on perceived usability has been examined in a limited number of quantitative studies. This study aims to measure undergraduate students' perceived usability of DeepL using the System Usability Scale (SUS) and to evaluate the impact of taking a translation technology course on these perceptions. In this quantitative study, the data were analyzed using descriptive statistics, exploratory factor analysis, and an independent-samples t-test. The results showed that DeepL's overall usability was at an "acceptable/good" level, and the scale's internal consistency was adequate. The factor analysis revealed a dominant, overarching usability factor. Students who had completed a course on translation technologies perceived DeepL as significantly more usable than their peers without such training. Statistical tests confirmed this disparity, which corresponded to a moderate effect size. These outcomes imply that targeted education in translation technology not only promotes more frequent use but also fosters more favorable subjective assessments of such tools. The study emphasizes the need to incorporate such tools into academic programs through a structured, critically engaged pedagogical approach.

Öz

DeepL gibi nöral makine çevirisi (NMÇ) araçları, çeviri eğitiminde giderek daha görünür hale gelmektedir. Artık öğrenciler bu araçları çeviri sürecinde giderek daha yaygın biçimde deneyimlemektedir. Bununla birlikte, çeviri teknolojisi derslerinin kullanılabilirlik algısı üzerindeki etkisi, sınırlı sayıda nicel çalışmada incelenmiştir. Bu çalışma, lisans öğrencilerinin DeepL'nin algılanan kullanılabilirliğini Sistem Kullanılabilirlik Ölçeği (SUS) kullanarak ölçmeyi ve çeviri öğrencileri arasında çeviri teknolojisi dersi alan ve almayanlar arasındaki kullanılabilirlik algısını değerlendirmeyi amaçlamaktadır. Bu nicel çalışmada veriler; betimleyici istatistikler, açıklayıcı faktör analizi ve bağımsız örneklem t-testi kullanılarak analiz edilmiştir. Sonuçlar, DeepL'nin genel kullanılabilirliğinin "kabul edilebilir/iyi" düzeyde olduğunu ve ölçeğin iç tutarlılığının yeterli olduğunu göstermiştir. Faktör analizi, baskın, genel bir kullanılabilirlik unsurunu ortaya koymuştur. Çeviri teknolojileri üzerine bir dersi tamamlamış öğrenciler, DeepL'yi bu tür bir eğitim almamış akranlarına göre önemli ölçüde daha kullanışlı olarak algılamışlardır. İstatistiksel testler, orta düzeyde bir etki büyüklüğüne karşılık gelen bu farklılığı doğrulamıştır. Bu sonuçlar, çeviri teknolojisinde hedeflenen eğitimin yalnızca daha sık kullanımı teşvik etmekle kalmayıp, aynı zamanda bu tür araçların daha olumlu öznal değerlendirmelerini de desteklediğini göstermektedir. Çalışma, bu tür araçların yapılandırılmış, eleştirel bir pedagojik yaklaşımla akademik programlara entegre edilmesinin gerekliliğini vurgulamaktadır.

Keywords

translation studies · neural machine translation · translator training · DeepL · system usability scale (SUS)



Citation | Atıf: Barut, E. (2026). Measuring the usability of DeepL: Impact of translation technology courses in translator training.. *Istanbul Üniversitesi Çeviribilim Dergisi-Istanbul University Journal of Translation Studies*, (24), 209-224. <https://doi.org/10.26650/iujts.2026.1882848>

This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License. 

© 2026. Barut, E.

Corresponding author | Sorumlu Yazar: Evren Barut ebarut@aku.edu.tr



INTRODUCTION

In fulfilling their role to prepare students for a global competition, higher education institutions shall also remain responsive to profound technological shifts. Within translation studies, this evolution traces a clear path: from computer-assisted tools to machine translation, and now to neural machine translation (NMT) systems. These advancing technologies are fundamentally redefining both the profession itself and the methods used to educate future translators (Li et al., 2024, p. 21). The move toward integrated digital workflows is displacing conventional translation methods. Consequently, future translators must now master a dual skillset: deep language proficiency alongside sophisticated technical expertise.

The advent of neural machine translation tools presents significant advantages, including accelerated workflows, enhanced consistency, and cost efficiency. However, their influence extends beyond mere productivity, profoundly affecting both the quality of translations and the very structure of production pipelines. Leveraging sophisticated neural networks, contemporary systems such as DeepL have surpassed earlier technologies in contextual understanding and stylistic fluency, generating results that, for many language combinations, rival human-produced translations. This evolution compels a parallel transformation in translator education. Students must cultivate the ability to interact with these tools not as simple aids, but as collaborators within a critical workflow. This shift requires mastering advanced competencies, including the rigorous evaluation of machine-generated text, the application of strategic post-editing, and the ability to execute comprehensive rewrites when necessary. (Wang et al., 2024, p. 17099).

In light of these developments, there is a pressing expectation for translation education programs to modernize their curricula by formally integrating AI-supported tools. The goal is to provide students with a balanced foundation of robust practical competencies and essential theoretical knowledge. Failure to do so risks leaving graduates ill-prepared for a translation industry that is undergoing rapid digital transformation, potentially widening the skills gap between academia and the professional world. Thus, understanding the impact of AI translation tools on higher education has become a critical research priority. This necessitates a specific focus on two interrelated questions: how users perceive the usability of these tools, and how formal training alters that perception.

Within translator education, the concept of “usability” encompasses the effectiveness, efficiency, and satisfaction that students derive from both learning and applying translation technologies. Higher perceived usability directly correlates with a greater likelihood of student adoption, facilitating the tool’s integration into both coursework and independent practice, and ultimately promoting its transfer to professional settings. Thus, usability must be viewed not merely as a technical attribute of software, but as a fundamental pedagogical element that shapes the learning experience itself.

DeepL has emerged as a predominant NMT-based tool among translation students, largely due to its capacity to generate outputs of notable quality in terms of grammatical coherence and semantic nuance. Conceptually, such task-specific NMT systems should be distinguished from prompt-based generative AI applications (e.g., ChatGPT): although both fall under the broad umbrella of artificial intelligence, DeepL functions as a dedicated, domain-narrow translation engine rather than a general-purpose generative model (Chen, 2024). This distinction is also borne out empirically, as comparative evaluations have shown that dedicated NMT systems tend to match or outperform general-purpose generative models on standard translation-quality metrics, with the latter being designed primarily as conversational rather than translation systems (Son & Kim, 2023). However, in an educational framework, output quality alone is an insufficient

proxy for usability. True usability is a multidimensional construct, involving factors such as interface intuitiveness, the clarity of commands, the learning curve's steepness, task completion time, and the system's tolerance for user error. For translator training, this translates to how readily students master the tool, the speed and consistency with which they accomplish assignments, and crucially, the extent to which they learn to employ it in a critical, controlled, and ethical manner.

This study employs the System Usability Scale (SUS), a concise, reliable, and widely adaptable instrument for assessing general usability. By quantifying DeepL's perceived usability through the SUS, the research serves a dual purpose: it renders students' technological experiences empirically visible, and it provides a means to test the potential influence of formal translation technology instruction on these perceptions.

RESEARCH BASIS

The widespread adoption of neural machine translation tools in translator education has not been matched by robust empirical research. A solid foundation for understanding how users perceive their usability in higher education, and how formal training alters this perception, is still lacking. While tools like DeepL demonstrate clear proficiency in linguistic output, this technical capability alone leaves critical pedagogical questions unanswered (Salinas & Burbat, 2023, pp. 245–247). Key among these are the dynamics of how students adopt such tools in learning environments, their subjective evaluations of ease of use, and the consequential impact of these perceptions on tangible learning outcomes (Li, 2024; Munawwarah & Martriwati, 2024). Prevailing research in the discipline concentrates on broad technology acceptance models and general attitudinal surveys. This focus, however, has created a significant gap in the literature, as there is a marked shortage of granular quantitative data addressing the core facets of translator training. These facets include detailed pedagogical applications, the cultivation of advanced post-editing skills, and frameworks for principled ethical practice. Without such evidence, our capacity to understand how these tools function within authentic learning environments remains fundamentally limited.

The integration of technology into translator education also contends with systemic barriers, including insufficient hands-on training, a shortage of instructors with dual pedagogical and technical expertise, and the belated, fragmented incorporation of relevant courses into curricula (Alzghoul et al., 2025; Zaghloul & Khasawneh, 2023). The accelerating pace of AI innovation demands a fundamental rethinking of translation curricula, assessment methods, and expected competencies. Consequently, a critical need exists to investigate how formal training in translation technology directly influences student evaluations of AI-based tools (Long, 2025, p. 1510). This study addresses that need by exploring the perceived usability of DeepL among undergraduate translation students and assessing how a dedicated technology course shapes their perceptions. It is guided by the following research questions. What are the perceptions of prospective translators studying in undergraduate Translation and Interpreting programs in higher education regarding the general usability of neural machine translation tools such as DeepL (RQ1)? Do perceived usability scores for DeepL differ significantly between translation undergraduates who have taken a translation technology course and those who have not (RQ2)?

This study employs the System Usability Scale to quantitatively assess translation students' experiences with DeepL, examining how formal technology education may influence their digital competency, including user acceptance, perceived usability, and AI literacy.

LITERATURE REVIEW

This section reviews the conceptual and empirical foundations of the study in three steps. It first introduces usability and the System Usability Scale (SUS) as the measurement framework adopted here; it then

examines the role of neural machine translation and related technologies in translator training; and it finally situates the study within the literature on technology acceptance and AI literacy.

System Usability Scale and Usability

Usability, a central concept in human-computer interaction and educational technology, refers to how easily, effectively, and satisfactorily users can employ a system. According to the ISO 9241-11 standard, it is defined as “the degree to which specific users can achieve specific goals in a specific usage context with effectiveness, efficiency, and satisfaction” (Derisma, 2020, p. 184). This definition shifts the focus from mere functionality to the quality of user experience, positioning effectiveness, efficiency, and satisfaction as its core components (Lu Jie et al., 2025, p. 2). Within educational settings, high usability is vital as it supports the learning process, minimizes extraneous cognitive load, and promotes the sustained adoption of a tool (Fini, 2009, p. 4). Success in digital learning environments further depends on technical elements such as accessibility, learnability, and operational efficiency.

To measure this construct, the System Usability Scale (SUS) offers a concise and practical solution. Developed by John Brooke in 1986 to evaluate electronic office systems, this 10-item scale captures subjective user perceptions (Brooke, 1996, p. 3). Despite its simplicity, the SUS has consistently shown high reliability and strong content validity, with Cronbach’s alpha values often exceeding “.91” (Bangor et al., 2008, p. 581). These robust psychometric properties are maintained across various language adaptations, contributing to its widespread use in both industry and academic researches (Clark et al., 2021; Gao et al., 2020). Its quick administration and reliability even with small sample sizes make it particularly useful for assessing user experience in educational technology contexts (Derisma, 2020, p. 183).

Nevertheless, the underlying factor structure of the SUS continues to be a subject of scholarly discussion. Although the scale was initially designed as a single-dimensional general usability indicator, later studies, particularly the analyses of Lewis and Sauro, have revealed that the SUS may have a two-factor structure consisting of 8 items of “Usability” and 2 items of “Learnability” (Borsci et al., 2009; Lewis & Sauro, 2009). While these two factors are highly correlated with the total SUS score, the learnability dimension in particular can offer additional information regarding how quickly and comfortably a user learns a new system. One of the widely accepted frameworks for interpreting SUS scores is presented by Bangor and colleagues (2008): scores of 85 and above indicate “very high usability”, 70–85 “good–excellent”, 50–70 “acceptable but open to improvement”, and below 50 “unacceptable usability” (Orfanou et al., 2015, p. 229). Such interpretation guides allow for a standardized assessment of the usability level of different systems, such as educational technologies and translation tools (Vlachogianni & Tselios, 2023, p. 7).

AI-based and Machine Translation Tools in Translator Training

In recent years, AI-based translation tools, along with neural and traditional machine translation systems, have become integral components of translator education (Birdsell, 2022; Rapa et al., 2024). Driven by advances in big data, cloud computing, and deep learning, these technologies have progressed rapidly, making proficiency in their use an essential part of the modern translator’s curriculum (Su & Li, 2023, p. 2). Translation programs are systematically reevaluating and redesigning their content to incorporate AI and machine translation as core pedagogical elements (Chen, 2024; Kenny, 2019; Zhang & Doherty, 2025).

Students use AI-based tools for a wide range of tasks such as; searching for terms, verifying meaning and context, preparing for translation tasks, comparing their own output with machine-generated text, aiding academic writing (Polakova & Klimova, 2023, p. 2). Platforms like DeepL enable learners to critically revise their drafts by comparing them against machine output, thereby honing both their writing and analytical skills (Liang, 2025, pp. 82–84; Munawwarah & Martriwati, 2024). They are also widely used for comprehending

source texts, improving writing, and editing, offering immediate feedback on grammar and semantic fluency. Despite some institutional restrictions, student use remains prevalent and often routine, indicating that machine translation has evolved from an alternative aid into a standard feature of computer-assisted translation environments (Salinas & Burbat, 2023, p. 256).

The primary advantages of these tools include accelerated workflows, terminology support, and overall efficiency gains. By granting instant access to extensive bilingual corpora, they help students grasp linguistic subtleties and cultural context, while simultaneously building a technological skill set that enhances their professional readiness (Alzghoul et al., 2025, pp. 398–399; Munawwarah & Martriwati, 2024). Additionally, they offer cognitive and affective benefits, such as fostering autonomous learning, reinforcing grammar, and reducing language-related anxiety. Among these tools, DeepL is frequently highlighted for its superior accuracy, fluency, and naturalness (Jiang, 2022; Li, 2024; Liang, 2025).

However, significant risks accompany their use, particularly with complex, culturally nuanced, or highly specialized texts. In such contexts, errors can cause serious misunderstandings and compromise translation quality, especially regarding idioms, cultural references, or technical terminology (Alzghoul et al., 2025, p. 399). Overreliance may also erode students' critical reading, problem-solving abilities, and active vocabulary recall (Munawwarah & Martriwati, 2024; Özmat & Akkoyunlu, 2024). Some students report anxiety about the tools' reliability, ethical boundaries, and potential long-term effects on their own translation competence (Li, 2023, pp. 3039–3044).

Empirical research on DeepL reflects these multifaceted effects. Studies note its positive impact on writing skills and perceived language proficiency, while examinations of student perceptions reveal a mix of recognized benefits and concerns (Kusuma Nata Laksana & Cahya Komara, 2024; Polakova & Klimova, 2023). In practice, its use has been shown to improve syntactic complexity, lexical range, fluency, and accuracy in student writing (Liang, 2025, p. 60). Although generally praised for outperforming other systems in accuracy and fluency, DeepL can still struggle with contextual appropriateness and finer semantic distinctions (Li, 2024, p. 714).

While the literature confirms the significant and multifaceted role of tools like DeepL in translation education, there is a distinct lack of research that quantifies their perceived usability through standardized measures such as the System Usability Scale (SUS). This study explicitly targets this empirical gap, employing the SUS to deliver a quantitative evaluation of DeepL's usability, as formulated in RQ1.

Technology Acceptance, and AI Literacy within Translator Training

In translator training, a student's judgment of how helpful a tool is will largely determine whether they embrace it and use it consistently. Research on translation tools shows the likelihood of students applying such technologies in both formal and informal learning (Alzghoul et al., 2025, p. 399). This process is further underpinned by the concept of self-efficacy, which refers to one's belief in their own ability to succeed. Specifically, computer self-efficacy and translation technology self-efficacy (TTSE) are vital, as they directly influence a student's perception of a tool's ease of use and their broader propensity to embrace new technologies (Li et al., 2024; Preciado & Vallejo, 2025; Wang et al., 2024).

Dedicated courses in translation technology can significantly reshape this landscape. They have the potential to improve students' technical skills, reduce anxiety associated with new tools, and foster more strategic and confident usage (Pietrzak, 2020). By teaching the functions of modern tools, such training better prepares students for professional demands and helps them adapt tools to various tasks (Alzghoul et al., 2025; Mohammed et al., 2020). However, these courses do not fully resolve all concerns. Persistent questions remain regarding the long-term impact of tools like neural machine translation on core translation

competence, alongside ongoing ethical dilemmas and uncertainties about the evolving role of the translator (Li, 2023; Tian et al., 2023).

AI literacy and data literacy are becoming fundamental skills for modern translator. AI tools use to encompass critical evaluation, effective collaboration with AI systems, and ethical decision-making (Chen, 2024; Krüger, 2023; Prado, 2024; Zhang & Doherty, 2025). This shift demands an evolution in curricula. Beyond linguistic mastery, translator education must now systematically incorporate skills for managing AI outputs, developing post-editing strategies, understanding AI-inclusive workflows, and applying ethical frameworks for AI use (Krüger, 2022, 2023, 2024).

AIM AND METHODOLOGY

The aim of this study is to measure undergraduate translation students' perceived usability of DeepL using the System Usability Scale (SUS) and to determine whether completing a translation technology course is associated with differences in these usability perceptions. In line with this aim, the following subsections present the research design, the participants and sampling method, data collection and ethical considerations, and the data analysis procedures.

Research Design

This study employs a quantitative, cross-sectional design to investigate how translation students perceive the usability of DeepL. Using a survey-based approach, the research descriptively and comparatively analyzes participant perceptions at a fixed point in time, a methodology common in studies of technology acceptance and usability within the field (Alzghoul et al., 2025; Kasperè et al., 2021; Yang & Wang, 2019). This quantitative framework facilitates the statistical examination of data gathered via standardized scales, allowing for the systematic and comparable presentation of findings on usability.

The study was conducted across multiple state and private universities in Turkey offering undergraduate programs in Translation and Interpreting. Participation was solicited through an academic network of department faculty, who encouraged their students to take part. To maintain focus on the core subject of user experience, the survey did not collect detailed personal demographic data. Instead, it concentrated solely on the experiences of students within the relevant disciplinary context

Participants and Sampling Method

This study involved a non-probability convenience sample of undergraduate students enrolled in Translation and Interpreting programs at Turkish universities. The survey link was distributed through a professional academic network, inviting voluntary participation. This method, common in translation education research, was chosen for its practical access to the target population within existing logistical constraints (Özmat & Akkoyunlu, 2024, pp. 153–154).

Data collection began with 234 initial responses. The dataset was then refined using an automated screening protocol. One participant who declined the informed consent agreement was excluded. Furthermore, to ensure sample integrity, the survey was designed to automatically conclude for any respondent who indicated they were not a Translation and Interpreting student. This process yielded a final, valid dataset comprising 233 participants from the target demographic. The raw data underlying these analyses are openly available (Barut, 2026).

As a result of these procedures, the analytical sample size for statistical analyses was determined as $N = 137$. Inclusion criteria were defined as follows:

- Being a student in the Translation and Interpreting undergraduate program,

- Answering “Yes” to the question “Did you use DeepL?”
- Having answered all 10 items of the System Usability Scale (SUS).

Participants who did not meet these criteria were excluded from the analytical sample. Detailed distributions for these subgroups are outside the primary focus of this study and are therefore not reported separately.

The main group comparison within the analytical sample is based on the variable of “having taken a course in translation technology and/or translation tools”. In this context:

- Main Group: took a course in translation technology/translation tools (from now on, as “course takers”): $n = 76$,
- Control Group: did not take this course (from now on, as “non-takers”): $n = 61$

Two subgroups were created. This grouping aims to test the possible effect of translation technology training on the perceived usability of DeepL. Summary information on the basic characteristics of the participants and group sizes is presented collectively in [Table 1](#) for reference in later sections of the study.

Table 1

Participant Flow and Formation of the Analytical Sample

Sample Overview		<i>n</i>
1	Initial Sample; all participants who completed the survey	234
2	Participants without informed consent	1
3	Participants with informed consent	233
4	Analytical Sample: Translation students using DeepL and completed the SUS fully	137
5	Translation Technology / Tools Course Status	
5a.	Course-takers as main group (translation students who took the course)	76
5b.	Non-takers as control group (translation students who did not take the course)	61

A total of 234 participants completed the survey. One of them did not give voluntary consent. Among the participants who gave consent, those who graduated from different undergraduate programs and did not complete the SUS scale were excluded from the analytical sample. Since the focus of the research was Translation and Interpreting students, the main analyses were conducted only with participants studying in these programs and who stated that they used DeepL ($n = 137$).

Data Collection and Ethical Considerations

The survey form used as a data collection tool consists of two parts: a short questionnaire containing demographic questions and the System Usability Scale (SUS). Participants’ basic characteristics were questioned to define the analytical sample and comparison groups with demographic questions. Multiple-choice questions were asked to determine their status as a Translation and Interpreting student, their experience using DeepL, and whether they had taken a translation technology course. The study focused on comparing course-takers as the main group with non-takers as the control group.

The main measurement tool of the study is the 10-item System Usability Scale (SUS), originally introduced by Brooke in 1986 (Brooke, 1996). The scale has a 5-point Likert-type response format. In accordance with the standard SUS scoring procedure, and taking reverse questions into account, a total score on the scale of 0–100 was obtained for each user based on the answers given to the 10 items (Brooke, 1996, p. 4). The study utilizes SUS scores as the primary quantitative measure of the DeepL user experience, enabling concrete analysis of its perceived usability.

The survey was administered using a validated Turkish adaptation of the SUS, which maintains the scale's reliability and validity for this context (Demirkol & Şeneler, 2018). This instrument effectively captures subjective evaluations across key usability dimensions—including ease of use, consistency, learnability, and overall satisfaction—aligning with its established application in educational technology research (Gao et al., 2020; Kasperé et al., 2021). In this analysis, the SUS data serves two main purposes: first, to describe the sample's overall usability perception of DeepL (RQ1), and second, to statistically compare these perceptions between students who have and have not completed a relevant technology course (RQ2). The scale showed acceptable internal consistency and reliable statistics according to the given data in results.

The data were collected online via Google Forms. The survey link was shared through an academic network comprising faculty members from Translation and Interpreting departments at various universities in Turkey; along with the network, faculty members invited undergraduate students in their programs to participate in the study. Participants voluntarily completed the survey, approximately in 5 minutes at their convenience. The data collection period was between June–September 2025.

The survey commenced with an ethical statement and an informed consent form. Participants who declined consent or did not belong to the target demographic of Translation and Interpreting students were automatically redirected to an exit page. Those who provided consent and met the criteria proceeded to the demographic section and the System Usability Scale (SUS), where the primary data were collected.

The study adhered to relevant ethical guidelines, including the principles of the Helsinki Declaration. Participants were informed that they could terminate the survey at any time, that their responses would be recorded anonymously, and that the data would only be used for research purposes. The research was ethically approved by the [institution name withheld for blind review] Social and Human Sciences Scientific Research and Publication Ethics Board with decision number 2025/269, taken at meeting number 06 on June 18, 2025. All data will be stored in environments accessible only to researchers and will not be shared with groups not involved in the academic purposes.

Data Analysis

Data analysis was conducted using appropriate statistical software. Descriptive statistics were calculated for SUS scores (0–100), hereinafter referred as SUS_100, and demographic variables to describe the sample. The internal consistency of SUS was assessed using Cronbach's α coefficient; construct validity was examined using exploratory factor analysis (EFA) (with goodness-of-fit checked using KMO and Bartlett tests). SUS scores were compared according to whether or not they had taken a Translation Technology/Tools course using an independent samples t-test; an additional check was performed using the Mann–Whitney U test to ensure sensitivity to any unmet assumptions. The significance level was set at $\alpha = .05$ for all tests (Alzghoul et al., 2025; Baumgartner et al., 2021; Brooke, 1996; Demirkol & Şeneler, 2018; Kasperé et al., 2021; Şahin & Aybek, 2020).

FINDINGS

The analytical sample consisted of $N = 137$ undergraduate students studying in the Translation and Interpreting department who stated that they used DeepL and answered all 10 items of the System Usability Scale (SUS). The reliability and dimensionality of the SUS scale were evaluated using this sample. The internal consistency coefficient of the SUS was calculated as Cronbach's $\alpha = .80$ ($\alpha = .795$). This value indicates that the scale has a good level of internal consistency in the target sample. When item-total correlations were examined, it was observed that removing any item from the scale did not significantly increase the α value; therefore, all 10 items (negatively formulated items were reverse-coded) continued to be used together

(Borsci et al., 2009, p. 194). Cronbach's α , KMO, Bartlett test, and summary statistics of explained variances are presented in Table 2.

Table 2

Exploratory factor analysis indices for the System Usability Scale (n = 137)

Sampling adequacy and sphericity			
Measure			Value
Cronbach's α (10 items)			.80
Kaiser–Meyer–Olkin (KMO)			.814
Bartlett's test of sphericity $\chi^2(45)$, p			387.65, $p < .001$
Factor eigenvalues and explained variance			
Factor	Eigenvalue	% of variance	Cumulative %
1	3.67	36.66	36.66
2	1.73	17.28	53.94

As summarized in Table 2, the factor structure of SUS items was analyzed using principal axis factoring and oblimin rotation with EFA. A KMO value of .81 and a significant Bartlett test indicate that the data is suitable for factor analysis (see Table 2). Eigenvalue and scree analysis supported a two-factor solution; these two factors together explained approximately 54% of the total variance (see Table 2).

Factor loadings and common variances are reported in Table 3, and the findings are evaluated within the context of the dimensionality of SUS discussed in the literature.

Table 3

Factor loadings and communalities for the System Usability Scale (N = 137)

Item	Factor 1 (General Usability)	Factor 2 (Learnability)	Communality (h)
SUS1	.625	–	.549
SUS2_reverse	.450	–	.342
SUS3	.585	–	.410
SUS4_reverse	–	.517	.364
SUS5	.653	–	.515
SUS6_reverse	.586	–	.348
SUS7	.517	–	.317
SUS8_reverse	.665	–	.471
SUS9	.689	–	.550
SUS10_reverse	–	.519	.397

In the oblimin rotational EFA ($|load| < .40$ suppressed), the majority of items clustered in the first factor, representing the General Usability dimension (see Table 3). Two items, needing technical support and requiring extensive learning before using the system, clustered in a smaller second factor, reflecting the Learnability dimension. The scree plot showed a significant break after the first factor, indicating that the structure was predominantly driven by a dominant general factor.

Therefore, in subsequent analyses, a single composite SUS_100 score was used as an indicator of general usability; this approach supported a holistic assessment of DeepL's perceived usability within the context of RQ1.

Usability of DeepL (SUS_100)

For the analytical sample ($N = 137$), the total System Usability Scale (SUS) score for DeepL was transformed into a normalized value ranging from 0 to 100, following the standard scoring procedure established by Brooke (1996). This resulting metric, SUS_100, represents the tool's overall perceived usability. Descriptive statistics are as follows: $M = 75.69$, $SD = 14.83$; minimum = 20, maximum = 100. These descriptive statistics, including the means and standard deviations for all SUS items, are presented in Table 4.

Table 4

Mean and standard deviation of SUS item scores for DeepL ($N = 137$)

Item	SUS Statement	M	SD
1	I think I will want to use this system frequently.	3.71	1.06
2	I found this system unnecessarily complex.	4.45	0.68
3	I thought this system was easy to use.	4.42	0.76
4	Technical support may be needed to use this system.	4.20	0.94
5	I found the various functions in this system well integrated.	3.59	1.03
6	I thought there were too many inconsistencies in this system.	3.80	1.09
7	I think many people will learn to use this system very quickly.	4.37	0.83
8	I found this system very inconvenient to use.	4.24	1.03
9	I felt very confident using this system.	3.44	1.27
10	I thought I would need to learn a lot of things before I could do anything with this system.	4.06	1.16

In light of standard SUS benchmarks, where scores above 70 are typically deemed acceptable, DeepL's composite perceived usability falls into an "above-average" classification. This result suggests that translation students generally regard DeepL as a functional, practical, and satisfactory tool.

As detailed in Table 4, the mean scores for the vast majority of scale items exceed 3.5, indicating a common perception among students that DeepL is user-friendly, intuitive, and functionally consistent. Slightly lower averages for specific items, particularly concerning the integration of functions within the system (item 5) and the user's confidence during operation (item 9), nevertheless highlight discrete aspects of usability that could be enhanced.

Collectively, the score distribution affirms the view of DeepL as a practical and accessible resource in translation training. The range of scores, from a minimum of 20 to a maximum of 100, underscores considerable individual variation while also demonstrating a clear potential for high user satisfaction. These findings provide the quantitative evidence sought by RQ1, confirming the general positive perception of DeepL's usability and supporting the validity of the SUS within this specific research context.

Effect of translation-technology training on SUS scores

The analytical sample was divided into two groups based on whether or not they took the Translation Technology/Tools course: the group taking the course ($n = 76$, $M = 78.98$, $SD = 12.64$) and the group not taking the course ($n = 61$, $M = 71.60$, $SD = 16.37$). Descriptive statistics for these groups are presented in 5, and the t-test results are summarized both in the text and in the table footnote. The distribution of SUS_100 scores for the groups is also visualized in Figure 1 using a boxplot. Figure 1 shows that the SUS_100 scores of the group taking the course exhibit a higher central tendency (median) and a relatively tighter distribution (lower variance); while in the group not taking the course, the scores are distributed over a wider range and have a few outliers at the lower end.

Assumptions were checked before making intergroup comparisons using independent samples t-test. Levene equivalence test supported the assumption of equal variance with $F = 3.24$, $p = .074$. The main t-test analysis result was $(t(135)) = 2.98$, $p = .003$; the mean difference between groups was found in favor of the group taking the course with a SUS score of 7.38 (95% confidence interval [2.48, 12.28]). The effect size was Cohen's $d = 0.51$ (95% CI [0.17, 0.85]), indicating a moderate difference. The descriptive and statistical summary of this difference is given in Table 5. The results show that students who took the translation technology/tools course perceived DeepL as statistically significantly more usable.

Table 5

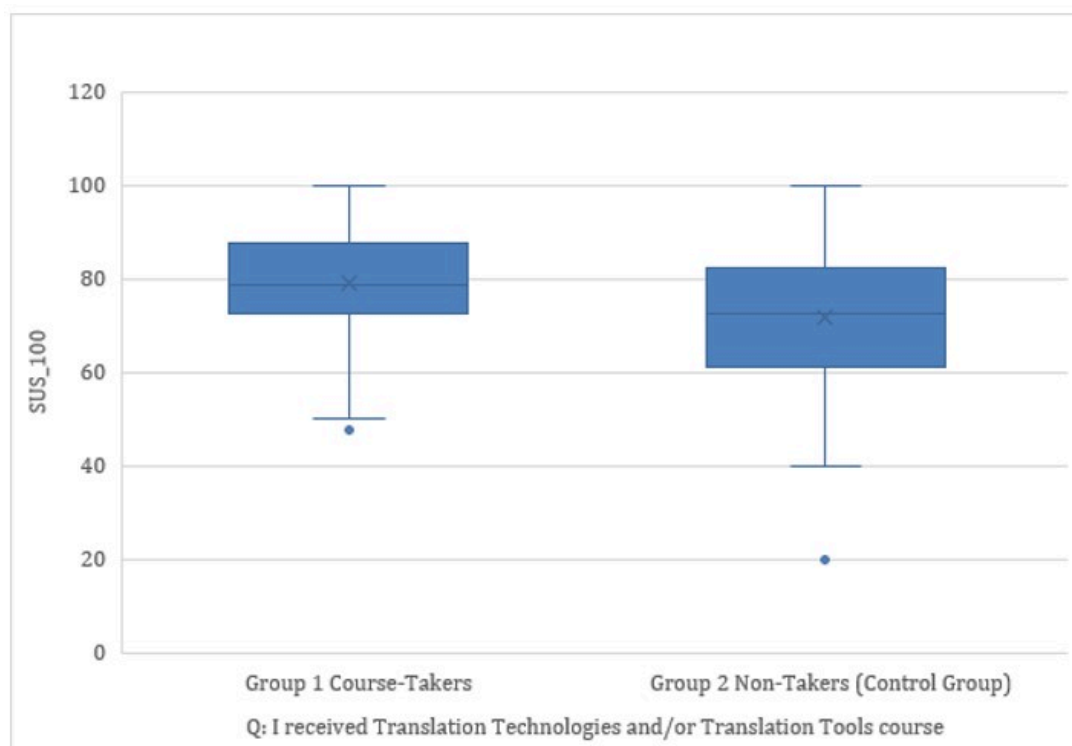
SUS_100 scores depending on translation technology course taking

Group	n	SUS_100 Median	SS
Course takers	76	78.98	12.64
Non-takers -control group-	61	71.60	16.37

Although normality tests and distribution analyses allowed the use of parametric tests, a complementary Mann-Whitney U test was also applied to assess the robustness of the results. The mean ranks were found to be 77.34 for the group taking the course and 58.61 for the group not taking the course; $U = 1684.00$, $Z = -2.75$, $p = .006$ (two-sided). The effect size was calculated as $r \approx .23$ (small-moderate). This non-parametric finding provides converging evidence confirming the t-test result and supports the interpretation that translation technology training makes a significant difference in the perceived usability of DeepL.

Figure 1

Distribution of DeepL SUS_100 scores based on translation technology course



Therefore, in the context of RQ2, students who have taken a translation technology/tools course perceive DeepL as more usable, both statistically and pedagogically significantly (medium effect size). These findings lend empirical support to the positive role of structured education in fostering technology adoption (Yang & Wang, 2019, p. 118). They further underscore how systematic training in translation technology actively shapes student perceptions of AI-based tools (Su & Li, 2023, pp. 2–3). The results reaffirm the critical impor-

tance of integrating AI literacy and technological competencies into translator education. This evidence provides a foundation for developing informed pedagogical approaches and institutional strategies to ensure the sustainable and effective use of AI in professional translation practice (Alotaibi & Salamah, 2023; Kasperé et al., 2021).

DISCUSSION AND CONCLUSION

The average SUS_100 score obtained for DeepL (≈ 75.7) is in a mid-to-high position. Studies show that average SUS scores mostly cluster in the 68–77 range. For example, values around 69 have been reported, and in the 70–76 range for some learning management system (LMS) solutions. In this context, the use of DeepL offers a level of usability comparable to, or even higher than, many LMS and similar ed-tech tools in the eyes of students (Ghasemi et al., 2025; Harrati et al., 2016; Orfanou et al., 2015; Vlachogianni & Tselios, 2023).

In the study, the fact that most items are loaded onto the general usability factor, while items related to learning needs and technical support are loaded onto the secondary factor, reflects this structural separation previously reported in ed-tech environments (Harrati et al., 2016; Orfanou et al., 2015). From this perspective, DeepL is positioned in a competitive place among widespread educational technologies not only with its position in the field of translation but also with its usability profile

The fact that translation technology/tools course takers perceive DeepL at a higher level of usability ($d \approx 0.51$) indicates that education plays a significant role in technology acceptance and user experience. In structured courses, students gain regular, guided experience with tools like DeepL through designed tasks, post-editing exercises, and error analysis. With regard to the technology acceptance model, it has been shown that repeated experiences increase perceived ease of use and perceived usefulness, which directly affects the behavioral intention to use such tools (Yang & Wang, 2019, p. 117).

Translation technology self-efficacy and computer self-efficacy are also variables that reinforce this effect. Students' beliefs about their own technical competence increase both the use of tools like DeepL and the satisfaction and comfort felt while working with these tools (Li et al., 2024; Preciado & Vallejo, 2025; Wang et al., 2024). In this context, it can be said that the obtained moderate effect size is not only statistically significant but also pedagogically significant.

These results can be directly related to the concept of AI literacy. Translation technology courses should not only be technical trainings answering the question of “how to use it?”; they should also offer a critical framework that discusses the strengths and weaknesses of AI-based tools, error types, bias risks, and ethical boundaries (Alzghoul et al., 2025; Chen, 2024; Zhang & Doherty, 2025). The quantitative results indicate that structured training positively recalibrates student perceptions of DeepL, moving them toward a more favorable and grounded assessment. Furthermore, these findings substantiate a strong correlation between AI literacy and technology acceptance within the field of translation.

The results strongly suggest that technology courses must be repositioned as a core component of the translation curriculum, rather than an elective supplement. Tools like DeepL should be methodologically embedded within comparative translation analysis, post-editing workshops, and quality assessment tasks (Alzghoul et al., 2025; Chen, 2024). This approach enables students to gain operational proficiency while also critically evaluating the limitations of machine output, recognizing common error patterns, and understanding the enduring necessity of human judgment.

Translation curricula must evolve to include dedicated modules on AI literacy, data literacy, and the emerging role of the translator as an AI mediator or manager (Krüger, 2022, 2023; Zhang & Doherty, 2025). Relatedly, education needs to raise student awareness of associated ethical and practical concerns, such as data privacy, copyright, and academic integrity (Alotaibi & Salamah, 2023; Kasperé et al., 2021). At an

institutional level, this requires developing clear, pedagogically informed policies for AI tool use. A strategic, ethically guided framework that regulates use is preferable to a simple prohibition, as it prepares students for the complex, technology-driven workflows of the modern language industry (Alzghoul et al., 2025; Su & Li, 2023).

The findings of this study should be evaluated considering some limitations. First, the sample consists only of students enrolled in translation and interpreting programs in Turkey and is based on a convenience sampling approach. While the sample size ($n = 137$) is sufficient for SUS and basic comparison analyses, caution should be exercised when generalizing to the whole country or different educational contexts. Relatedly, data collection took place during the summer term (June–September 2025), largely outside the active academic semester; this timing may have constrained participation and should be taken into account when interpreting the sample size. Secondly, the measurement tool used is self-report-based; therefore, the relationship between perceived usability and actual performance or task achievement is outside the scope of this study. Thirdly, the study focuses solely on DeepL and does not include other AI/CAT tools (e.g., Google Translate, Memsource, Trados-based NMT integrations). This leads to a concentration of findings within a specific tool context, preventing comparisons between different tools. Finally, the cross-sectional design limits the possibility of causally interpreting the relationship between translation technology training and perceived usability; a design that directly tracks pre- and post-lesson change was not preferred.

Future research could examine both cross-instrument comparisons and cross-cultural differences by working with broader and more diverse samples across different university types and countries. Longitudinal designs, comparing SUS scores and performance metrics before and after translation technology courses, could more clearly reveal the impact of training over time. Triangulating SUS findings with qualitative data collection techniques (e.g., think-aloud protocols, focus group interviews) will provide a deeper understanding of how students experience DeepL and similar tools. Finally, developing institutional guidelines and assessment frameworks for the ethical and responsible use of neural machine translation tools in education will remain a key agenda item at both the research and policy levels (Ko et al., 2025, pp. 6–7).

This study is original research that quantitatively evaluates the usability of the neural machine translation tool DeepL in the context of translation and interpreting education using the System Usability Scale. The findings from a sample of 137 translation and interpreting students show that the overall mean SUS_100 score of DeepL is 75.69, which is above the acceptable/good usability level (Harrati et al., 2016; Orfanou et al., 2015). The internal consistency of the SUS (Cronbach's $\alpha \approx .79$) and its resultant factor structure confirm that the scale is primarily defined by a dominant, overarching usability factor, while also featuring a secondary and more limited learnability dimension (Brooke, 1996; Vlachogianni & Tselios, 2023). Regarding RQ1, students clearly position DeepL as a tool that provides a functional, accessible, and broadly positive user experience.

Addressing RQ2, a significant and moderate difference in SUS_100 scores was found between students who had completed a translation technology course ($n = 76$) and those who had not ($n = 61$), with an effect size (Cohen's d) of approximately 0.51. This quantitatively demonstrates that systematic training enhances perceived usability. The finding aligns with extensions of the Technology Acceptance Model and literature on translation technology self-efficacy, supporting the premise that experience and instruction bolster perceived ease of use, which in turn strengthens attitudes and behavioral intentions toward tool adoption. Thus, the study reveals that DeepL can function not merely as a technical aid but as a legitimate educational resource that facilitates learning when embedded within sound pedagogical design.

These results carry direct implications for curriculum development. Translation technology courses should be designed as foundational, core elements of the curriculum. Integrating tools like DeepL into structured activities, such as post-editing, comparative translation, and error analysis, can reduce cognitive

load while building students' self-efficacy and intention to use technology (Li et al., 2024; Yang & Wang, 2019). Crucially, such courses must transcend operational training to encompass AI literacy, including critical engagement with issues of bias, data privacy, academic integrity, and ethical application (Alzghoul et al., 2025; Chen, 2024; Zhang & Doherty, 2025). At an institutional level, creating clear policy and evaluation frameworks for AI tool use will better prepare graduates for the AI-integrated workflows defining the modern language industry (Kasper et al., 2021; Su & Li, 2023). Considering studies examining the potential of AI-based systems in foreign language learning and findings regarding the acceptance of models such as ChatGPT as translation tools, it remains important to investigate the role of DeepL and similar tools in the translation training ecosystem in greater depth, along with their ethical and pedagogical dimensions (Sugiyama & Yamanaka, 2023; Wang et al., 2024). This study provides a quantitative and systematic foundation for evaluating AI-based tools in translator training through the critical lens of usability.

Data Availability Statement The anonymized raw dataset supporting the findings of this study (N = 233 participants) is openly available in the Zenodo repository at <https://doi.org/10.5281/zenodo.20511024>, under a CC BY-NC 4.0 license.



Peer Review	Externally peer-reviewed.
Author Contributions	
Conflict of Interest	The author has no conflict of interest to declare.
Grant Support	The author declared that this study has received no financial support.

Hakem Değerlendirmesi	Dış bağımsız.
Yazar Katkısı	
Çıkar Çatışması	Yazar çıkar çatışması bildirmemiştir.
Finansal Destek	Yazar bu çalışma için finansal destek almadığını beyan etmiştir.

Author Details	Evren Barut (Asst. Prof. Dr.)
Yazar Bilgileri	¹ Afyon Kocatepe University, Faculty of Science and Literature and Department of Translation Studies, Afyonkarahisar, Türkiye, 0000-0002-0915-9603 ebarut@aku.edu.tr

References

- Alotaibi, H. M., & Salamah, D. (2023). The impact of translation apps on translation students' performance. *Education and Information Technologies*, 28(8), 10709–10729. <https://doi.org/10.1007/s10639-023-11578-y>
- Alzghoul, M. A., Abualasal, J., Alazzam, T. S., Al-Ghammaz, S. A. D., & Alzghoul, R. M. (2025). The Status Quo of Translation Technology Tools in Translator Training Programs at Jordanian Universities. *World Journal of English Language*, 15(7), 398. <https://doi.org/10.5430/wjel.v15n7p398>
- Bangor, A., Kortum, P., & Miller, J. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Barut, E. (2026). *DeepL usability — System Usability Scale (SUS) survey raw data* [Data set]. Zenodo. <http://dx.doi.org/10.5281/zenodo.20511024>
- Baumgartner, J., Ruettgers, N., Hasler, A., Sonderegger, A., & Sauer, J. (2021). Questionnaire experience and the hybrid System Usability Scale: Using a novel concept to evaluate a new instrument. *International Journal of Human Computer Studies*, 147. <https://doi.org/10.1016/j.ijhcs.2020.102575>
- Birdsell, B. (2022). Student Writings with DeepL: Teacher Evaluations and Implications for Teaching. *JALT Postconference Publication*, 2021(1), 117. <https://doi.org/10.37546/jaltpcp2021-14>



- Borsci, S., Federici, S., & Lauriola, M. (2009). On the dimensionality of the System Usability Scale: A test of alternative measurement models. *Cognitive Processing*, 10(3), 193–197. <https://doi.org/10.1007/s10339-009-0268-9>
- Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, Thomas B., McClelland Ian Lyall, & Weerdmeester (Eds.), *Usability Evaluation in Industry* (Number 194, pp. 189–194). Taylor&Francis. <https://www.researchgate.net/publication/228593520>
- Chen, P. (2024). The Impact of Generative AI on the Role of Translators and Its Implications for Translation Education. In *Education Insights Educ. Insights* (Vol. 1, Number 2). <https://soapubs.com/index.php/EI>
- Clark, N., Dabkowski, M., Driscoll, P. J., Kennedy, D., Kloo, I., & Shi, H. (2021). Empirical Decision Rules for Improving the Uncertainty Reporting of Small Sample System Usability Scale Scores. *International Journal of Human-Computer Interaction*, 37(13), 1191–1206. <https://doi.org/10.1080/10447318.2020.1870831>
- Demirkol, D., & Şeneler, Ç. (2018). Sistem Kullanılabilirlik Ölçeğinin Türkçeye Çevirisi: SUS-TR. *Usak University Journal of Social Sciences*, 11(Eylül-2018), 238–253. <https://doi.org/10.29217/uujss.495>
- Derisma, D. (2020). The Usability Analysis Online Learning Site for Supporting Computer programming Course Using System Usability Scale (SUS) in a University. *International Journal of Interactive Mobile Technologies (IJIM)*, 14(9), 182. <https://doi.org/10.3991/ijim.v14i09.13123>
- Fini, A. (2009). The Technological Dimension of a Massive Open Online Course: The Case of the CCK08 Course Tools. *The International Review of Research in Open and Distributed Learning*, 10(5). <https://doi.org/10.19173/irrodl.v10i5.643>
- Gao, M., Kortum, P., & Oswald, F. L. (2020). Multi-Language Toolkit for the System Usability Scale. *International Journal of Human-Computer Interaction*, 36(20), 1883–1901. <https://doi.org/10.1080/10447318.2020.1801173>
- Ghasemi, M. S., Aghajafari, R., Abolaghasemi, J., Danesh, M. K., & Garosi, E. (2025). Development and validation of a virtual learning management system usability questionnaire: A case study. *Work*, 80(1), 75–86. <https://doi.org/10.3233/WOR-230290>
- Harrati, N., Bouchrika, I., Tari, A., & Ladjailia, A. (2016). Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Computers in Human Behavior*, 61, 463–471. <https://doi.org/10.1016/j.chb.2016.03.051>
- Jiang, R. (2022). How does artificial intelligence empower EFL teaching and learning nowadays? A review on artificial intelligence in the EFL context. *Frontiers in Psychology*, 13, 1–8. <https://doi.org/10.3389/fpsyg.2022.1049401>
- Kasperė, R., Horbačauskienė, J., Motiejūnienė, J., Liubinienė, V., Patašienė, I., & Patašius, M. (2021). Towards sustainable use of machine translation: Usability and perceived quality from the end-user perspective. *Sustainability (Switzerland)*, 13(23). <https://doi.org/10.3390/su132313430>
- Kenny, D. (2019). Technology and translator training. In *Routledge eBooks* (pp. 498–515). Informa. <https://doi.org/10.4324/9781315311258-30>
- Ko, U. H., Hartley, K., & Hayak, M. (2025). Exploring AI in Education: Preservice Teacher Perspectives, Usage, and Considerations. *Journal of Information Technology Education: Research*, 24. <https://doi.org/10.28945/5592>
- Krüger, R. (2022). Integrating professional machine translation literacy and data literacy. *Lebende Sprachen*, 67(2), 247–282. <https://doi.org/10.1515/les-2022-1022>
- Krüger, R. (2023). Some Reflections on the Interface between Professional Machine Translation Literacy and Data Literacy. *Journal of Data Mining and Digital Humanities*, 1. <https://doi.org/10.46298/jdmdh.9045>
- Krüger, R. (2024). Outline of an Artificial Intelligence Literacy Framework for Translation, Interpreting and Specialised Communication. *Lublin Studies in Modern Languages and Literature*, 48(3), 11–23. <https://doi.org/10.17951/lsmll.2024.48.3.11-23>
- Kusuma Nata Laksana, & Cahya Komara. (2024). Indonesian EFL Students' Perceptions of DeepL Machine Translation Tool: Utilization, Advantages, and Disadvantages. *Journal of Language and Literature Studies*, 4(2), 256–276. <https://doi.org/10.36312/jolls.v4i2.1931>
- Lewis, J. R., & Sauro, J. (2009). The Factor Structure of the System Usability Scale. In *Lecture notes in computer science* (pp. 94–103). Springer Science+Business Media. https://doi.org/10.1007/978-3-642-02806-9_12
- Lewis, J. R., & Sauro, J. (2017). Revisiting the Factor Structure of the System Usability Scale. *Journal of Usability Studies Archive*, 12(4), 183–192. <https://doi.org/10.5555/3190867.3190870>
- Li, F. (2023). A Survey of Translation Learners' Uses and Perceptions of Neural Machine Translation. *Theory and Practice in Language Studies*, 13(11), 3039–3048. <https://doi.org/10.17507/tpls.1311.34>
- Li, L. (2024). Artificial Intelligence Translator DeepL Translation Quality Control. *Procedia Computer Science*, 247(C), 710–717. <https://doi.org/10.1016/j.procs.2024.10.086>
- Li, X., Gao, Z., & Liao, H. (2024). An empirical investigation of college students' acceptance of translation technologies. *PLoS ONE*, 19(2). <https://doi.org/10.1371/journal.pone.0297297>
- Liang, L. (2025). The impact of using DeepL Translator on Chinese EFL students' story writing. *Journal of China Computer-Assisted Language Learning*, 5(1), 56–93. <https://doi.org/10.1515/jccall-2024-0009>

- Long, Y. (2025). Reimagining Translation Education: Fostering Student Engagement With AI Technologies in Cultural Translation Assessment. *Journal of Language Teaching and Research*, 16(5), 1506–1514. <https://doi.org/10.17507/jltr.1605.07>
- Lu Jie, Schmidt, M., & Shin Jinnie. (2025). *Beyond Technological Usability: Exploratory Factor Analysis of the Comprehensive Assessment of Usability Scale for Learning Technologies (CAUSLT)* Jie Lu**1*.
- Mohammed, S., Samad, M. S., & Mahdi, S. (2020). The Attitudes of Professional Translators and Translation Students towards Computer-Assisted Translation Tools in Yemen. *Journal of Language and Linguistic Studies*, 16(2), 1084–1095. <https://doi.org/10.17263/jlls.759371>
- Munawwarah, M., & Martriwati, ; (2024). *STUDENTS' PERCEPTION OF USING DEEPL AS MACHINE TRANSLATION IN ENGLISH LEARNING*. 5(2), 284–295. <https://doi.org/10.22236/ellter.v5i2.15813.g4779>
- Orfanou, K., Tselios, N., & Katsanos, C. (2015). Perceived usability evaluation of learning management systems: Empirical evaluation of the System Usability Scale. *The International Review of Research in Open and Distributed Learning*, 16(2). <https://doi.org/10.19173/irrodl.v16i2.1955>
- Özmat, D., & Akkoyunlu, B. (2024). Artificial Intelligence-Assisted Translation in Education: Academic Perspectives and Student Approaches. *Participatory Educational Research*, 11(Gift-Issue), 151–167. <https://doi.org/10.17275/per.24.99.11.6>
- Pietrzak, P. (2020). Inside and Outside the Translation Classroom. *Research in Language*, 18(2), 109–117. <https://doi.org/10.18778/1731-7533.18.2.07>
- Polakova, P., & Klimova, B. (2023). Using DeepL translator in learning English as an applied foreign language – An empirical pilot study. *Heliyon*, 9(8). <https://doi.org/10.1016/j.heliyon.2023.e18595>
- Prado, M. C. de A. P. (2024). Promoting AI Literacy among English Language and Literature and Translation Majors. *Innovative Teaching and Learning*, 6(1), 7–25. <https://doi.org/10.4208/itl.20240103>
- Preciado, J. S. A., & Vallejo, R. G. (2025). Self-Efficacy and Perception in the Use of GAI Tools in the Translation Classroom. *Seminars in Medical Writing and Education*, 4(526). <https://doi.org/10.56294/mw2025526>
- Rapa, A. A., Saja, I., & Azmi, A. A. A. (2024). The Use of Artificial Intelligence (AI) Translation Tools: Implications for Third Language Proficiency. *International Journal of Research and Innovation in Social Science*, VIII(IX), 1952–1960. <https://doi.org/10.47772/IJRISS>
- Şahin, M., & Aybek, E. (2020). Jamovi: An Easy to Use Statistical Software for the Social Scientists. *International Journal of Assessment Tools in Education*, 6(4), 670–692. <https://doi.org/10.21449/ijate.661803>
- Salinas, M. J. V., & Burbat, R. (2023). Google Translate and DeepL: breaking taboos in translator training. Observational study and analysis. *Iberica*, 2023(45), 243–266. <https://doi.org/10.17398/2340-2784.45.243>
- Son, J., & Kim, B. (2023). Translation performance from the user's perspective of large language models and neural machine translation systems. *Information*, 14(10), 574. <https://doi.org/10.3390/info14100574>
- Su, W., & Li, D. (2023). The effectiveness of translation technology training: a mixed methods study. *Humanities and Social Sciences Communications*, 10(1). <https://doi.org/10.1057/s41599-023-02066-2>
- Sugiyama, K., & Yamanaka, T. (2023). Proposals and Methods for Foreign Language Learning Using Machine Translation and Large Language Model. *Procedia Computer Science*, 225, 4750–4757. <https://doi.org/10.1016/j.procs.2023.10.474>
- Tian, S., Jia, L., & Zhang, Z. (2023). Investigating students' attitudes towards translation technology: The status quo and structural relations with translation mindsets and future work self. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1122612>
- Vlachogianni, P., & Tselios, N. (2023). Perceived Usability Evaluation of Educational Technology Using the Post-Study System Usability Questionnaire (PSSUQ): A Systematic Review. *Sustainability*, 15(17), 12954. <https://doi.org/10.3390/su151712954>
- Wang, Y. X., Chen, L. P., & Han, J. Y. (2024). Exploring factors influencing students' willingness to use translation technology. *Education and Information Technologies*, 29(13), 17097–17118. <https://doi.org/10.1007/s10639-024-12511-7>
- Yang, Y., & Wang, X. (2019). Modeling the intention to use machine translation for student translators: An extension of Technology Acceptance Model. *Computers and Education*, 133, 116–126. <https://doi.org/10.1016/j.compedu.2019.01.015>
- Zaghlool, Z. D. M., & Khasawneh, M. A. S. (2023). Aligning Translation Curricula with Technological Advancements; Insights from Artificial Intelligence Researchers and Language Educators. *Studies in Media and Communication*, 12(1). <https://doi.org/10.11114/smc>
- Zhang, J., & Doherty, S. (2025). Investigating novice translation students' AI literacy in translation education. *The Interpreter and Translator Trainer*, 19(3–4), 234–253. <https://doi.org/10.1080/1750399X.2025.2541478>