

**A Process-based Study on the Effects of Using Electronic Dictionaries in Translation
Formative Tests on Students' Vocabulary Retention**

Essam Hassan Al-Mizgagi

An associate Professor of Translation and Applied Linguistics,
A faculty member of Translation Program at University of Science and Technology, Yemen

Abstract

The role and regulating electronic dictionary use during translation examinations remain a contentious controversy within translation pedagogy and assessment. While dictionary can serve as a significant tool, concerns persist regarding its potential to foster over-reliance that may hinder the development of essential translation competencies. This study adopted a process-based perspective to investigate the effects of dictionary-free versus dictionary-based use during translation formative quizzes and exams on students' vocabulary retention, as measured by performance on a final translation test conducted without dictionaries. Employing an experimental pre-test/post-test design with control and experimental groups (38 students), the study examined the frequency and nature of problematic vocabulary encountered by both control group and experimental group. Additionally, a semi-structured questionnaire was used to explore students' perceptions and preferences regarding dictionary use across various translation courses. Preliminary findings of this study suggested that while dictionary use during quizzes aided immediate task completion for the control group, the experimental group (no dictionary during quizzes) demonstrated potentially better retention of vocabulary encountered during the intervention, as evidenced by fewer circled words in the final dictionary-free post-test. Questionnaire data indicated nuanced student preferences, often favoring dictionary access, particularly for highly specialized words, highlighting a potential discrepancy between perceived need and learning outcomes.

Keywords: translation assessment, (electronic) dictionary, vocabulary retention, process-based approach, translator training, formative tests

(Received: 06 December 2025; Revised: 19 January 2026; Accepted: 04 February 2026; Published: 30 March 2026)

Introduction

In an increasingly interconnected global landscape, the demand for skilled translators continues to grow, placing significant emphasis on the effectiveness of translator training programs (Kelly, 2005). A cornerstone of these programs involves rigorous assessment designed to evaluate the multifaceted competencies required for professional translation practice. However, the design and implementation of translation assessments, particularly concerning the use of assistive tools like e-dictionaries, remain subjects of ongoing debate among educators and researchers (Mahmoud, 2017; Alfayez, 2023). The central tension lies between simulating realistic professional conditions, where translators routinely utilize various resources, and ensuring assessments accurately gauge internalized knowledge and skills, including robust vocabulary command (PACTE Group, 2017). It's worth noting that the two terms of 'dictionary' and 'e-dictionary' are used interchangeably due to the overwhelming use of electronic dictionaries in PCs and cell phones.

The study's focus on vocabulary acquisition as a process, rather than solely on translation product quality in the presence of e-dictionaries, shifts the perspective. While dictionaries undeniably aid immediate task performance (Wang, 2011), their role in *learning* appears more complex. The findings suggest that assessment policies should perhaps differentiate between formative practice aimed at learning and summative assessment aimed at evaluating performance under specific conditions. Allowing dictionaries universally might inadvertently undermine the formative goal of vocabulary development.

Review of Literature

Traditionally, translation assessment often leaned towards a product-oriented approach, focusing primarily on the quality of the final translated text (Nord, 2018; Munday, 2022). This approach evaluates accuracy, fluency, and appropriateness, often employing summative grading rubrics. While valuable for gauging final output, product-oriented assessment may offer limited insight into the cognitive processes, strategies, and underlying knowledge bases (or lack thereof) that students employ during the translation task (Susanne, 2009). Consequently, it may not fully capture the developmental aspects of learning, such as the gradual acquisition and internalization of vocabulary.

In contrast, a process-based approach shifts the focus towards understanding *how* students translate, examining the strategies they use, the challenges they encounter, and the cognitive effort involved (González Davies, 2004; Kiraly, 2000). This perspective aligns more closely with constructivist learning theories, emphasizing active knowledge construction and the development of strategic competence. Process-oriented assessment tends to be more formative, providing feedback that can guide learning by illuminating difficulties encountered during the task, such as problematic lexical searches or comprehension failures (Colina, 2003; Al-Mizgagi, 2014). While often associated with qualitative methods like think-aloud protocols or keystroke logging (Wang, 2011), a process-oriented *perspective* can also inform the interpretation of quantitative data gathered through more traditional assessment formats, focusing on *what* performance patterns reveal about underlying processes and learning.

One critical component of translation competence heavily influenced by assessment practices is vocabulary knowledge. Effective translation necessitates not only broad lexical coverage but also deep understanding of nuances, collocations, register, and context-specific meanings in both source (SL) and target languages (TL) (Nation, 2013). E-dictionaries, whether bilingual or monolingual, paper-based or electronic, are indispensable tools for professional translators (Frankenberg-Garcia, 2011). Their role in pedagogy, however, is complex. Allowing unrestricted dictionary access during assessments might seem ecologically valid, reflecting real-world practice. Yet, concerns arise that this could lead to an over-reliance, potentially masking vocabulary deficiencies and impeding the crucial process of vocabulary acquisition and long-term retention (Laufer & Hill, 2000). Students might develop efficient look-up strategies but fail to internalize frequently encountered or contextually vital vocabulary, leading to what some term "deskilling" (Vieira, 2020b).

Vocabulary acquisition in a second language (L2) is a gradual process, moving from no knowledge through various stages of partial understanding to mastery (Graves, 2000; Nation, 2013). Incidental vocabulary learning, occurring as a byproduct of engaging in meaning-focused tasks like reading or translation, is considered a primary driver of lexical development (Hulstijn, 2001). Dictionary use can potentially facilitate this process by providing immediate clarification and drawing attention to unknown words (Gorjian et al., 2013; Laufer & Hill, 2000). However, the *depth* of processing involved in a quick dictionary look-up, especially under exam pressure, may not always be sufficient for durable learning (Nakata, 2016). Furthermore, the type of e-dictionary and the learner's proficiency and strategic competence significantly mediate the learning potential (Laufer & Hill, 2000).

This study situates itself within this complex nexus of translation assessment, dictionary use, process-oriented pedagogy, and vocabulary acquisition. It directly addresses the controversy surrounding dictionary use in translation tests by investigating its effects not primarily on the immediate quality of the translated product (the focus of many previous studies, e.g., Mahmoud, 2017; Koponen & Salmi, 2017; Al-Mizgagi, 2024), but on the *acquisition and retention* of vocabulary encountered during practice, assessed under dictionary-free conditions. By adopting a process-oriented *interpretation* of performance data (frequency of problematic words) within a quasi-experimental design, the study aims to provide empirical evidence on whether restricting dictionary access during formative practice tasks (quizzes) fosters greater vocabulary internalization compared to allowing unrestricted access. Furthermore, it explores the often-overlooked variable of text typology, acknowledging that the need for and impact of dictionary use may vary significantly depending on the lexical density and specificity of the source text. Finally, it incorporates student perspectives to understand their perceived needs and preferences regarding dictionary use across different learning contexts.

The significance of this research lies in its potential to inform evidence-based pedagogical and assessment practices in translator training. If restricting dictionary use during certain practice phases proves beneficial for vocabulary retention without unduly hindering overall translation skill development, it could suggest modifications to current assessment policies. Conversely, if e-dictionary use consistently aids performance without demonstrably harming long-term acquisition, it might support more lenient policies, perhaps coupled with explicit training in strategic dictionary use (Trojszczak, 2022; Shin, 2019). Ultimately, by focusing on the process of

learning and considering variables like text type, this study seeks to contribute to a more efficient understanding of the role of dictionaries in developing competent and resourceful translators.

Research Problem

The central problem addressed by this study is the persistent lack of consensus and empirical clarity regarding the optimal policy for dictionary use in translation assessments, particularly concerning its impact on students' long-term vocabulary acquisition. While numerous studies have examined dictionary effects, they often yield inconclusive or context-dependent findings, leaving educators without clear guidance (O'Neill, 2019; Frankenberg-Garcia, 2011; Vieira, 2020b). This study identifies and aims to address specific limitations and gaps in the existing literature.

Firstly, much of the debate was polarized. Studies like Mahmoud (2017) reported statistically significant advantages for dictionary-aided translation in test settings, advocating for their inclusion. Mahmoud's study, comparing performance of students translating with and without bilingual dictionaries, found significantly higher scores in the dictionary group, concluding that dictionaries compensate for lexical deficiencies. However, the study's focus remained primarily on the *product* (test scores) rather than the underlying learning *process* or long-term retention. Furthermore, the claim of a "highly significant difference" warrants scrutiny regarding effect size and the specific nature of the improvements (e.g., avoiding critical errors vs. achieving higher stylistic quality). Critics might argue such findings merely demonstrate the utility of a tool for a specific task, not necessarily its benefit for skill development.

Conversely, concerns about over-reliance and potential "deskilling" fuel arguments against unrestricted dictionary access (Vieira, 2020b). Studies comparing dictionary use with corpus consultation, for instance, suggest corpora can lead to better lexical choices and fluency, implying that dictionaries might offer insufficient contextual or collocational information (Frankenberg-Garcia, 2011; Al-Mizgagi, 2020). The proposal by Alfayez (2023) for a hybrid approach—allowing dictionaries initially but banning them later in an exam—represents an attempt to balance support with fostering independence. While innovative, this approach primarily addresses test-taking strategy and time management within a single event, rather than the cumulative effect of dictionary use practices on learning over time.

The Study Focus

This study focuses on two critical "blind spots" often under-addressed:

1. Previous research predominantly measures the effect of dictionary use on the *immediate performance* within the test or task itself (e.g., translation quality scores, task completion time) (Koponen & Salmi, 2017; Wang, 2011). While important, this overlooks a crucial pedagogical goal: the *internalization* of vocabulary encountered. Over-reliance on dictionaries, even if leading to better immediate scores, might circumvent the cognitive effort (e.g., retrieval practice, deeper processing) necessary for consolidating new lexical items into long-term memory (Nakata et al., 2021). Professional translators use dictionaries strategically, often for specialized, low-frequency terms or to confirm

nuances, relying on a vast internalized lexicon for common and domain-specific vocabulary (Shin, 2019). Training programs should arguably foster this internalization. This study directly targets this gap by measuring vocabulary retention through a post-test conducted *without* dictionaries, hypothesizing that practice conditions (with/without dictionary during quizzes) influence this retention. The frequency of words students identify as problematic (by circling) serves as a proxy indicator of their acquired vocabulary knowledge related to the intervention texts.

2. Translation is not monolithic; the challenges posed vary significantly with text type (e.g., literary, technical, legal, political). General policies allowing or banning dictionaries across all translation courses and test types may be overly simplistic. Translating a highly specialized medical report, laden with specific terminology, presents different lexical demands than translating a general political news report, which might rely more on understanding nuance, connotation, and media-specific phraseology. The utility and necessity of a dictionary likely fluctuate accordingly (Neshkovska, 2019). Overgeneralizing findings from studies using one text type to all translation assessment contexts can be misleading. This study acknowledges this by focusing on a specific genre (political texts) for the intervention but also incorporates student perspectives (via questionnaire) on the perceived need for dictionaries across *different* course types, implicitly probing their awareness of text typology's role.

Therefore, the current research aims to move beyond the simple "allow vs. ban" dichotomy by investigating the *process* implications of dictionary use during formative practice on subsequent vocabulary retention under test conditions, while acknowledging the potential influence of text type.

Objectives of the Study

This study aims to achieve the following objectives:

1. To investigate the effect of allowing versus restricting dictionary use during formative translation practice (quizzes on political texts) on students' vocabulary acquisition,
2. To explore translation students' perceptions and preferences regarding the necessity and utility of dictionary use during examinations across different types of translation-related courses (e.g., general translation, legal, political)

Significance of the Study

The findings of this study have significant potential implications for both pedagogical strategies and evaluation design in translator training programs;

Firstly, by providing empirical data on the relationship between dictionary access during practice and subsequent vocabulary retention, it can inform decisions about when and how to integrate (or restrict) dictionary use in the curriculum to optimize learning. If restricting dictionary use during formative stages proves beneficial for vocabulary acquisition, educators might consider incorporating more dictionary-free tasks or quizzes designed to encourage practice of retention and deeper lexical processing.

Secondly, the study highlights the importance of considering text typology in assessment policies. The results could support more differentiated approaches, where dictionary access rules are tailored to the specific demands of the text genre and the learning objectives of the course (e.g., allowing dictionaries for highly specialized terminology tests but restricting them for assessing fluency with general or semi-specialized language).

Thirdly, by incorporating student perspectives, the study can bridge the gap between pedagogical intentions and student experiences, potentially revealing anxieties or perceived needs that should be addressed through explicit instruction or support. Ultimately, this research contributes to a more evidence-based, nuanced, and process-aware approach to translation assessment that aims to foster not just immediate performance but also the long-term development of robust translation competence, including lexical proficiency.

Limitations of the Study

This study acknowledges several limitations inherent in its design and context. Firstly, the findings will be based on a specific and relatively small sample of **N=38** Level 3 translation students at the University of Science and Technology, Yemen. The generalizability of the results to students at different proficiency levels, in different institutions, or from different linguistic backgrounds, as well as the statistical power to detect smaller effects, may be limited by this sample size. Secondly, the study focuses exclusively on political texts for the intervention and testing phases. While this controls for text type, it also means the findings may not directly apply to translation involving other genres (e.g., literary, technical). Thirdly, vocabulary acquisition is assessed indirectly through the frequency of self-identified problematic words (circling). While practical for a classroom setting, this method is less precise than standardized vocabulary tests and relies on student consistency in identifying difficulties. It serves as a proxy for knowledge gaps rather than a direct measure using a scale like Graves' (2000) stages, though the interpretation will draw on such frameworks. Fourthly, the study employs a quasi-experimental design; while efforts will be made to ensure group comparability (e.g., based on pre-test scores), true random assignment might be constrained by institutional factors, potentially limiting causal claims. The scope in this tool is broader since it includes students' preferences of dictionary use in other courses, namely, *reading*, *writing*, *general translation* and *legal translation*. Finally, the study duration (one semester) captures short-to-medium term effects; long-term vocabulary retention is not assessed. The study will be conducted during the second semester of the academic year 2024-2025.

Research Methodology

This study adopts a quasi-experimental, mixed-methods design to address the research objectives. The quantitative component includes a pre-test/post-test comparison between an experimental and a control group to assess the impact of the intervention which is dictionary restriction during quizzes on vocabulary acquisition of the experimental group. Besides, the qualitative component employs a semi-structured questionnaire to gather data on student perceptions and preferences. The overall interpretation depends on a process-oriented perspective, focusing on how performance differences reflect underlying learning mechanisms.

Research Design and Procedures

The core of the study is a quasi-experimental design:

- **Source Texts:** A corpus of authentic English political news reports and short opinion pieces (approx. 100-150 words each) was selected from reputable international news sources (e.g., Reuters, BBC News, The Guardian). Texts about were chosen for their relevance to Middle-east Crisis (Arab-Israel Conflict in Palestine) and inclusion of both general academic vocabulary and some domain-specific political terminology. Comparability in terms of length, lexical density, and syntactic complexity across texts used for pre-test, quizzes, and post-test were ensured through piloting to be refined and modified accordingly.
- **Participants:** Two groups of students **n=38** – Experimental Group (EG) **n=19**, and Control Group (CG) **n=19**.
- **Pre-test (week 2):** Both groups take an initial translation test (T1) *without* e-dictionaries to establish a baseline and check for initial equivalence between groups regarding translation ability *without* e-dictionary (60 minutes).
- **Intervention (week 3-12):** Both groups work through the same corpus of political texts over several weeks, approximately 5 quizzes administered over the intervention period (every two weeks). Both CG and EG used dictionaries during their study period classwork and homework. During the quizzes, CG students were allowed to use their own standard bilingual (English-Arabic) dictionaries, whether paper or approved electronic versions (no internet access to machine translation or translation memory features), and EG students competed quizzes *without* dictionaries. Both groups were asked to circle problematic words. (30-45 minutes each)
- **Post-test (week 13):** Both groups take a post-test (T2) based on comparable political texts encountered during the intervention, administered *without* prior notice and *without* dictionary access for either group (60 minutes).
- **Dependent Variable (Primary):** Frequency of words circled by students on the post-test (T2) translation task, indicating perceived difficulty/lack of knowledge.
- **Independent Variable:** Condition during intervention quizzes (Dictionary access vs. No dictionary access).
- **Questionnaire:** Administered after the post-test to both groups to gather perceptual data.

Materials

1. Vocabulary Knowledge Assessment:

- Students were instructed to *circle* any word or short phrase in the source text (during T1, quizzes, and T2) that they found problematic, or felt unsure about translating accurately without assistance. The *frequency* of circled words per student per test/quiz was the primary quantitative data point.
- While not directly applying the Vocabulary Knowledge Scale (VKS) by Graves (2000) for scoring individual words due to methodological complexity in this design, the *interpretation* of the frequency and potential *nature* of circled words was upon its stages. For instance, a decrease in circled common words might

suggest movement from Stage 1/2 (unknown) towards Stage 3/4 (known terms) for the experimental group. The VKS provides a theoretical lens:

- *Stage 1: No Knowledge:* Word is unknown (likely circled).
 - *Stage 2: General Sense:* Vague idea (might be circled if context isn't clarifying).
 - *Stage 3: Narrow, Context-Bound Knowledge:* Understood in specific contexts (less likely circled unless usage is novel).
 - *Stage 4: Rich, Decontextualized Knowledge:* Good understanding (unlikely circled).
 - *Stage 5: Mastery:* Full command (not circled).
 - Comparing the *change* in frequency and *type* (e.g., general vs. specialized, based on post-hoc analysis by researchers) of circled words between groups from pre-test/quizzes to the post-test.
2. **Semi-Structured Questionnaire:** Developed by the researchers (in Arabic) to gather student perceptions. It will include:
- *Demographic Information:* Age, years studying translation.
 - *Dictionary Use Habits:* Frequency of dictionary use for coursework, preferred dictionary types.
 - *Perceptions on Dictionary Use in Exams:* Using Likert-scale questions (e.g., 1=Strongly Disagree to 5=Strongly Agree) asking about the perceived necessity/helpfulness of dictionaries in exams for different course types (e.g., General Translation EN-AR, AR-EN Translation, Specialized (Political) Translation, Legal Translation, Academic Writing, Reading Comprehension).
 - *Open-Ended Question:* Asking students to elaborate on their reasons for preferring or not preferring dictionaries in specific exam types, and any challenges faced when using or not using dictionaries.

Data Analysis

Data analysis was performed using IBM SPSS Statistics (Version 23) for quantitative data and thematic analysis for qualitative data.

1. Quantitative Analysis (Vocabulary Acquisition - RQ1)

The study counted the number of circled words for each student on T1, each quiz, and T2. The study calculated means (M) and standard deviations (SD) for the number of circled words for EG and CG on T1 and T2. It calculated means for quizzes for descriptive trends. The study used an Independent Samples t-test to compare the mean number of circled words on the pre-test (T1) between EG and CG that revealed that not statistically significant ($p > .05$) which supports initial group equivalence. The study used an Independent Samples t-test to compare the mean number of circled words on the post-test (T2) between EG and CG. A statistically significant difference ($p = .037$) was resulted which indicated an effect of the intervention condition. That is, EG had a significantly lower mean than CG.

○

VOCABULARY CATEGORY	GROUP	PRE-TEST (T1)	POST-TEST (T2)
1. GENERAL ACADEMIC	Experimental	4.8 (1.9)	1.5 (1.1)
	Control	5.1 (2.2)	4.0 (1.8)
2. SEMI-SPECIALIZED POLITICAL	Experimental	6.2 (2.5)	3.0 (1.5)
	Control	6.5 (2.8)	5.5 (2.0)
3. HIGHLY SPECIALIZED	Experimental	4.8 (1.7)	5.7 (1.9)
	Control	4.9 (1.9)	5.0 (1.6)

Note. N=38 (n=19 per group).

Interpretation

The data clearly suggests that the pedagogical benefit of practicing without a dictionary is not uniform across all vocabulary types. The intervention was *highly effective* for promoting the acquisition and retention of *General Academic* that shows a significant difference between the Experimental Group (M=1.5, SD=1.1) and the Control Group (M=4.0, SD=1.8), and *Semi-Specialized Political* vocabulary that shows a significant difference between the Experimental Group (M=5.7, SD=1.9) and the Control Group (M=5.0, SD=1.6). These are the words that form the backbone of the text students were working with. By being forced to grapple with these words without the immediate use of a dictionary. The Experimental Group engaged in more significant cognitive effort—retrieval practice, inferencing from context, and negotiating meaning—which are known to lead to deeper processing and stronger memory traces (Nakata, 2015). Consequently, when faced with these words in the final test, they were able to recall and use them with greater confidence, resulting in significantly fewer circled words in these categories.

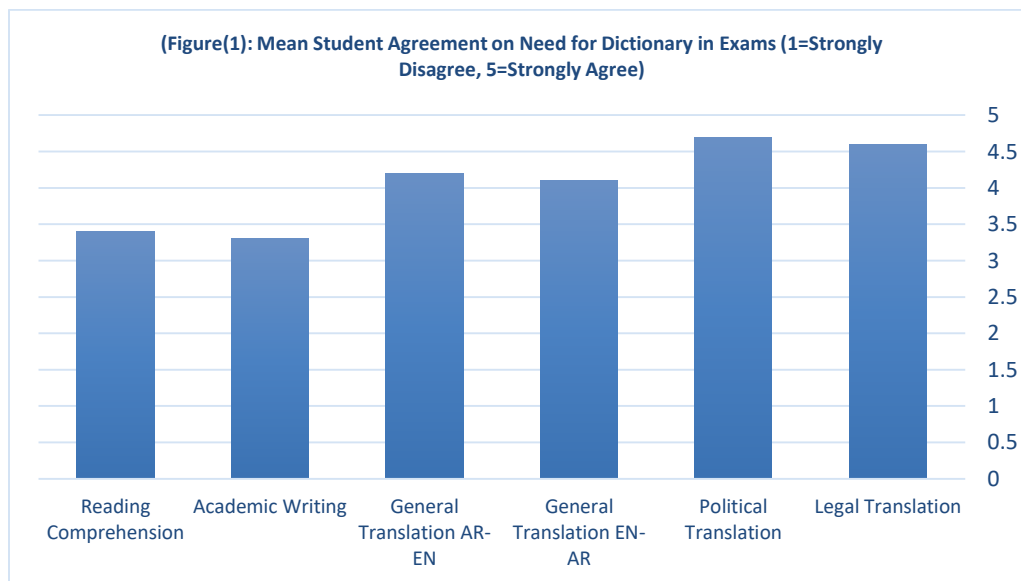
In contrast, the intervention showed no significant effect on *Highly Specialized* or *Low-Frequency* vocabulary that shows no significant difference between the Experimental Group (M=5.7, SD=1.9) and the Control Group (M=5.0, SD=1.6). This is also a critical finding. It suggests that this type of vocabulary may not be as amenable to incidental acquisition through contextual guessing, especially for L2 learners. These words often carry a high conceptual load and have low repetition rates, making them difficult to learn without explicit instruction or a direct look-up tool. Both groups, regardless of their practice condition, struggled with these words in the absence of a dictionary.

This finding has direct pedagogical implications. It supports a blended approach: restricting dictionaries can be a powerful strategy to build a robust core and domain-specific lexicon (general academic and semi-specialized words), but for highly technical terminology, other methods—such as pre-teaching key terms, providing glossaries, or indeed, training in strategic dictionary/corpus use—remain essential. This analysis validates the idea that a "one-size-fits-all" dictionary policy is suboptimal and that effective pedagogy must adapt its approach based on the type of vocabulary being targeted.

Findings and Analysis of Questionnaire

Analysis of the semi-structured questionnaire revealed student views.

- Likert Scale Data:** Students generally expressed a strong preference for having dictionary access during translation exams. Mean agreement scores (on a 1-5 scale) were highest for courses perceived as lexically demanding, such as Specialized (Political) Translation ($M = 4.6$, $SD = 0.6$) and Legal Translation ($M = 4.8$, $SD = 0.5$). Preference was slightly lower but still positive for General Translation EN-AR ($M = 4.1$, $SD = 0.8$) and AR-EN Translation ($M = 4.2$, $SD = 0.7$). Interestingly, preferences were lowest, though still leaning positive, for courses like Academic Writing ($M = 3.5$, $SD = 1.0$) and Reading Comprehension ($M = 3.2$, $SD = 1.1$), suggesting students perceive a greater need for dictionaries specifically during the translation task itself. See Figure (1).



- Thematic Analysis of Open-Ended Questions**

There are major themes emerged regarding dictionary preferences:

- Many students declared that dictionaries are essential for finding precise terminology, especially in specialized fields, and ensuring translation accuracy required for good grades. Quote: *"In political texts, one wrong word can change the meaning. The dictionary guarantees I find the exact term."*
- Several students highlighted exam time constraints, arguing that dictionaries allow them to work faster and overcome lexical hurdles quickly, preventing them from getting stuck. Quote: *"Without a dictionary, I waste too much time trying to remember or guess words, especially under exam stress. This will drastically affect my score."*

3. Some students declared that access to a dictionary was often linked to reduced anxiety and increased confidence in tackling challenging texts. Quote: *"Knowing the dictionary is there makes me feel safer and less stressed during the exam."*
4. Interestingly, a few students (primarily from the EG, perhaps influenced by their experience) noted a potential downside, with one commenting, *"Sometimes relying too much on the dictionary stops you from thinking about the context or trying to remember words you actually half-know."*. Most of this minority whose view aligns with the quantitative findings, ticked on no-use of dictionaries in the exams.
5. It is clear that the high ration of use-dictionary in the exams can be referred to score-bases reasons that students, including most of EG, can guarantee a better score when they use dictionaries in the exam. See (1) and (2).
6. Most of the students explicitly and implicitly declared that using dictionaries in the exams saves their time and efforts despite the positive effects of no-use of dictionary in mid and long run. Quote: *"although not using dictionary in the exam is good for our language especially vocabulary, we don't have enough time for this."*

Integration of Findings

The quantitative results of this paper suggest a learning benefit (better vocabulary retention) from practicing translation without dictionaries. However, the qualitative data reveals a strong student preference for dictionary access, driven by concerns about accuracy, efficiency, and confidence under exam conditions. This highlights a potential disconnect between practices that may optimize learning and assessment conditions that students perceive as fair or enabling for demonstrating their current best performance. The experimental group performed better in the dictionary-free post-test, suggesting they *learned* more vocabulary during the intervention, yet questionnaire responses indicate even some of them might prefer dictionary access in high-stakes exams. This tension unveils students' orientations most of whom are materialistically motivated through good scores and only few of them focus on improving their vocabulary retention skills. However, this tension underscores the complexity of designing assessment policies that are both pedagogically sound and perceived as supportive by students.

Discussion

The findings of this paper offer valuable insights into the ongoing debate surrounding dictionary use in translation tests, particularly when viewed through a process-oriented lens focused on vocabulary acquisition. The significant difference found in the dictionary-free post-test (T2), favoring the experimental group that practiced without dictionaries, gives empirical support to the notion that restricting dictionary access during formative stages can foster greater vocabulary internalization and retention.

This finding aligns with principles from cognitive psychology and L2 vocabulary acquisition research. The act of struggling to retrieve a word from memory, even if initially unsuccessful, is considered a potent learning event (the "retrieval practice" effect or "testing effect") (Nakata, 2016; Roediger & Karpicke, 2006). By removing the immediate aid of a dictionary during

quizzes, the experimental group was forced to engage in more effortful processing, potentially leading to deeper encoding and better long-term retention of the vocabulary encountered in the political texts (Craik & Lockhart, 1972). The control group, while likely performing better *on the quizzes* due to dictionary support (Mahmoud, 2017), may have engaged in shallower processing for looked-up words, resulting in less durable learning, as evidenced by their greater reliance on circling words in the final dictionary-free test. This indicates with concerns about over-reliance potentially hindering the development of a robust internal lexicon (Vieira, 2020b; Laufer & Hill, 2000).

The finding that the benefit was more pronounced for general academic and semi-specialized vocabulary, with both groups still struggling with highly specialized terms, is also informative. It suggests that dictionary restriction is most effective for consolidating vocabulary that students are likely to encounter repeatedly or that falls within a learnable range, Nation's (2013) concept of *vocabulary zones*. Highly specialized, low-frequency terms might still necessitate dictionary support even for proficient learners or require different pedagogical approaches, e.g., explicit terminology instruction, glossary building (Neshkovska, 2019).

The questionnaire results, indicating strong student preference for dictionary access despite the potential learning benefits of restriction, pinpoints a critical pedagogical challenge. Students' focus on immediate performance, accuracy under pressure, and anxiety management (Koponen & Salmi, 2017; Shin, 2019) is understandable, especially in high-stakes assessment contexts. This discrepancy suggests that simply banning dictionaries, even if pedagogically justifiable for learning, may meet resistance or increase anxiety if not implemented thoughtfully. It underscores the need for clear communication about the rationale behind assessment policies and potentially for incorporating lower-stakes, dictionary-free practice activities that emphasize learning over grading. It also reinforces the importance of training students in *strategic* dictionary use for situations where they *are* permitted (Trojszczak, 2022), but equally, building their confidence and strategies for coping *without* dictionaries.

The findings of the questionnaire according to text typology confirm that students perceive dictionary needs as context-dependent, being highest for specialized translation. This supports the argument against one-size-fits-all dictionary policies. Future assessment frameworks might consider allowing dictionaries for highly specialized texts while restricting them for general texts, or adopting hybrid models like Alfayez's (2023), perhaps framed differently by allowing dictionaries for specific sections testing terminology, but not for sections testing fluency with general language like writing and reading courses.

Implications for Pedagogy and Assessment

Based on these findings, several pedagogical and assessment implications are the following:

1. The study highlights the importance of integrate regular, low-stakes translation tasks and quizzes where dictionaries are restricted to encourage vocabulary retrieval practice and internalization, particularly for general and semi-specialized vocabulary.

2. The study pinpoints the necessity of complement implicit learning through translation practice with explicit vocabulary teaching strategies focusing on words identified as challenging.
3. When dictionaries are allowed, the study recommends providing explicit training on *how* to use dictionaries effectively and efficiently, moving beyond simple look-ups to understanding usage, collocation, and context (Frankenberg-Garcia, 2011).
4. The study considers developing differentiated dictionary policies based on course level, learning objectives, and text typology. It assures avoiding of rigid allow/ban rules where possible and communication of the rationale clearly to students.
5. The study encourages providing feedback through using data from tasks (like circled words, even in dictionary-allowed settings) to provide formative feedback on vocabulary gaps and strategy use, shifting focus partly from the product score to the learning process.

Conclusion

This study proposed a process-based investigation into the effects of dictionary use during translation practice on subsequent vocabulary acquisition under test conditions. The

findings declared that while students strongly prefer dictionary access, particularly for specialized texts and under exam pressure, restricting dictionary use during formative practice can significantly enhance vocabulary retention, as declared by fewer perceived lexical difficulties in a subsequent dictionary-free test. This challenges the assumption that simply allowing dictionaries in line with professional practice is always optimal for learning within pedagogical contexts.

The research pinpoints the importance of moving beyond a simple product-focused approach of translation assessment towards a more process-oriented perspective that considers the development of underlying competencies like vocabulary knowledge. It underscores the need for nuanced, evidence-based policies on dictionary use, potentially differentiating based on text typology and learning objectives, and balancing pedagogical goals for long-term learning with students' immediate performance needs and anxieties. By focusing on vocabulary acquisition and incorporating student perceptions, this study gives a richer contribution to understanding of the complex role dictionaries play in translator training, advocating for assessment practices that actively foster robust lexical competence alongside other translation skills. While further research is needed to confirm and extend these findings, the study provides a compelling argument for educators to critically evaluate *how* and *when* dictionaries are used in their classrooms and assessments to best support the development of knowledgeable, resourceful, and ultimately, more independent translators.

Recommendations for Future Research

The study ensures the limitations noted earlier (particularly the relatively small sample size of $N=38$, which may limit statistical power and generalizability, alongside sample specificity, text type focus, indirect vocabulary measure, and quasi-experimental design). These limitations should be considered when interpreting these results. Future research could strengthen these findings through employing true randomization in group assignment, using more direct and

standardized measures of vocabulary knowledge (e.g., VKS applied to target words post-task, vocabulary recall/recognition tests), and replicating these findings with larger and more diverse samples to enhance statistical power and the generalizability of the conclusions. Other recommendation are like:

- In depth investigating effects across different text types (e.g., literary, technical) and translation directions (AR-EN).
- Carrying out longitudinal designs to track vocabulary retention over longer periods.
- Incorporating direct process measures (e.g., keystroke logging, eye-tracking, think-aloud protocols) during the tests/quizzes to better understand the cognitive processes involved in dictionary use (or non-use) and problem-solving.
- Exploring the interaction between dictionary use policies and individual learner differences (e.g., proficiency level, learning styles, motivation).

References

Alfayez, H. M. (2023). The Effect of Using Dictionaries in Translation Exams on the Target Tex[t]. *Arab World English Journal for Translation & Literary Studies*, 7(4), 39-47. <http://dx.doi.org/10.24093/awejtls/vol7no4.3>

Al-Mizgagi, E. 2014. The feasibility of teaching translation in the undergraduate level in the Yemeni universities. *Journal of Social Studies* Vol. 20, no. 39, pp.7-39. <https://doi.org/10.5281/ZENODO.14031289>

Al-Mizgagi, E. H. (2020). Constraints Yemeni Undergraduates Face in Subtitling Audiovisual Materials. *Journal of Social Studies*, 26(1), 109-128. <http://dx.doi.org/10.20428/jss.26.1.5>

Al-Mizgagi, E. H. (2024). Teaching Computational Linguistics in English Language Academic Programs. *Arabic Journal for Translation Studies*, 3(9), 188-206.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

Colina, S. (2003). *Translation teaching: From research to the classroom: A handbook for teachers*. McGraw-Hill.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.

Frankenberg-Garcia, A. (2011). Corpora versus dictionaries for checking linguistic conventions in writing. In F. G. Andersen, K. Bech, & H. S. Tarp (Eds.), *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow* (pp. 303-319). Peter Lang.

González Davies, M. (2004). *Multiple voices in the translation classroom: Activities, tasks, and projects*. John Benjamins Publishing.

Gorjian, B., Hayati, A. M., & Barazandeh, H. (2013). Investigating the Effects of Monolingual and Bilingual Dictionaries on Incidental Vocabulary Acquisition and Reading Comprehension of EFL Learners. *Theory and Practice in Language Studies*, 3(5), 756-762. <https://doi.org/10.4304/tpls.3.5.756-762>

Graves, M. F. (2000). *The vocabulary book: Learning and instruction*. Teachers College Press.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge University Press.

Kelly, D. (2005). *A handbook for translator trainers: A guide to reflective practice*. St. Jerome Publishing.

Kiraly, D. C. (2000). *A social constructivist approach to translator education: Empowerment from theory to practice*. St. Jerome Publishing.

Koponen, M., & Salmi, L. (2017). Post-editing quality: Analysing the correctness and necessity of post-editor corrections. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16, 137–148. <https://doi.org/10.52034/lanstts.v16i0.439>

Laufer, B., & Hill, M. (2000). Vocabulary learning through dictionary use: Which dictionary, which learners? In P. Bogaards & B. Laufer (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 129-143). Erlbaum.

Mahmoud, A. (2017). Should dictionaries be used in translation tests and examinations? *English Language Teaching*, 10(3), 171–177. <https://doi.org/10.5539/elt.v10n3p171>

Mayer, R. E., et al. (2017). The role of multimodal learning in vocabulary acquisition. *BMC Psychology*. (Note: The original proposal cited this vaguely. A specific article by Mayer on this topic would need identification for a final paper. E.g., Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.) [Placeholder - Requires specific citation refinement]

Munday, J., Ramos Pinto, S., & Blakesley, J. (2022). *Introducing Translation Studies: Theories and Applications* (5th ed.). Routledge. <https://doi.org/10.4324/9780429352461>

Nakata, T., Tada, S., Mclean, S. and Kim, Y.A. (2021), Effects of Distributed Retrieval Practice Over a Semester: Cumulative Tests as a Way to Facilitate Second Language Vocabulary Learning. *TESOL J*, 55: 248-270. <https://doi.org/10.1002/tesq.596>

Nakata, T. (2016). Effects of retrieval formats on second language vocabulary learning . *International Review of Applied Linguistics in Language Teaching*, 54(3), 257-289. <https://doi.org/10.1515/iral-2015-0022>

- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>
- Neshkovska, S. (2019). The role of electronic corpora in translation training. *Studies in Linguistics, Culture, and FLT*, 7(3), 48-58. <https://DOI:10.46687/SILC.2019.v07.004>
- Nord, C. (2018). *Translating as a Purposeful Activity: Functionalist Approaches Explained* (2nd ed.). Routledge. <https://doi.org/10.4324/9781351189354>
- O'Neill, E. M. (2019). Online translator, dictionary, and search engine use among L2 students. *Computer-Assisted Language Learning Electronic Journal*, 20(1), 154-177. <https://callej.org/index.php/journal/article/view/265>
- PACTE Group. (2017). Translation competence model. In PACTE Group (Eds.), *Researching Translation Competence by PACTE Group* (pp. 63-88). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.127>
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249-255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- Shin, J. Y., Dixon, L. Q., & Choi, Y. (2020). An updated review on use of L1 in foreign language classrooms. *Journal of multilingual and multicultural development*, 41(5), 406-419. <https://doi.org/10.1080/01434632.2019.1684928>
- Susanne, G. (2009). Towards a model of translation competence and its acquisition: The longitudinal study TransComp. Behind the mind: methods, models, and results in translation process research.
- Trojszczak, M. (2022). Translator Training Meets Machine Translation—Selected Challenges. In: Lewandowska-Tomaszczyk, B., Trojszczak, M. (eds) *Language Use, Education, and Professional Contexts. Second Language Learning and Teaching*. Springer, Cham. https://doi.org/10.1007/978-3-030-96095-7_11
- Vieira, L. N. (2020b). Over-reliance on technology? Translation trainees' behaviours and self-perceptions when using online translators. *The Interpreter and Translator Trainer*, 14(3), 288-308. <https://doi.org/10.1080/1750399X.2020.1803160>
- Wang, J. (2011). The use of e-dictionary to read e-text by intermediate and advanced learners of Chinese. *Computer Assisted Language Learning*, 25(5), 475-487. <https://doi.org/10.1080/09588221.2011.631144>