

DEVELOPING A TRANSLATION COURSEBOOK EVALUATION CHECKLIST: EXPERT APPRAISAL OF A PRELIMINARY FRAMEWORK

TỔNG THỊ MỸ LIÊN

ĐỖ HƯƠNG GIANG

Faculty of English Language and Culture,
VNU University of Languages and International Studies

Received: 30 July 2026; Revised: 30 August 2026; Accepted for publication: 10 September 2026.

Tóm tắt

Số với số lượng lớn các khung đánh giá và các nghiên cứu thực nghiệm về khung đánh giá tài liệu giảng dạy tiếng Anh, lĩnh vực nghiên cứu dành riêng cho khung đánh giá chuyên biệt trong đào tạo dịch thuật còn khá sơ khai. Do đó, nghiên cứu này được thực hiện nhằm xây dựng một bộ tiêu chí đánh giá giáo trình dịch thuật. Bộ tiêu chí sơ bộ gồm 50 tiêu chí, được thiết kế dựa trên khung đánh giá giáo trình của Cunningsworth (1995), mô hình năng lực dịch thuật của PACTE (2017) và quy trình xây dựng bộ tiêu chí của Tomlinson (2023). Sau đó, bộ tiêu chí này đã được đánh giá về mức độ phù hợp, rõ ràng và khả thi bởi 14 giảng viên dịch thuật ở bốn trường đại học công lập tại Hà Nội. Kết quả cho thấy bộ tiêu chí nhận được các đánh giá tích cực từ các chuyên gia với mức độ đồng thuận cao, trong đó có năm tiêu chí cho thấy mức độ khác biệt tương đối lớn trong phản hồi của các chuyên gia. Nghiên cứu này đã đưa ra một bộ tiêu chí sơ bộ được chuyên gia thẩm định để đánh giá giáo trình dịch thuật; tuy nhiên, cần có thêm các nghiên cứu tiếp theo để xác trị và đánh giá tính hữu ích trong thực tiễn của bộ tiêu chí này trong đào tạo biên-phiên dịch tại các trường đại học.

Từ khóa: Đánh giá giáo trình dịch thuật, , đào tạo dịch thuật, năng lực dịch thuật.

1. Introduction

Coursebooks are a crucial component of the teaching process, serving as an important means for teachers to organize and deliver learning content systematically (Satriani & Resmini, 2021). Based on coursebook guidelines and suggestions, they can save time and effort in designing and implementing teaching activities effectively (Çelenk & Er, 2025). Therefore, evaluation for coursebook selection or improvement is an essential scientific task. In the field of English Language Teaching (ELT), various material evaluation theories and frameworks have been established to assess the relevance, effectiveness, and compatibility of materials. The most salient are those by Cunningsworth (1995), McDonough et al. (2013), and Tomlinson (2023). Additionally, there have been various empirical studies on the development of ELT material evaluation checklists, such as Litz (2001), Islam (2013), Vy (2025), Mukundan et al. (2011), and Zohrabi (2011).

When it comes to translation and interpreting training in higher education, the program should not only establish theoretical foundations for students but also provide opportunities for practicing translation skills (Rijoly, 2023). Beyond learning content, a good translation coursebook should also create a simulated working environment, helping students enhance their critical thinking, creativity, and professional translation competences. This is especially important now, as the rapid growth of AI has brought significant changes to the translation and interpreting fields and their training. Compared with the existing variety of evaluation frameworks and empirical studies on evaluation checklists for ELT materials, the number of specialized ones dedicated to translation training, especially in the context of AI-assisted translation, remains modest, and these existing frameworks have not

Email: lienttm@vnu.edu.vn

comprehensively included essential knowledge and skills in translator training in the new era. For example, 23 evaluative questions in Poulis and Ali's (2020) study focused on general principles, such as the alignment of the material to students' needs, the balance between theory and practice, the quality and diversity of texts, and the support in the self-study process. Ho's (2017) empirical research employed questionnaires to evaluate a coursebook of translation and interpretation, yet these were relatively limited as they targeted only two major criterion groups. While content-focused criteria encompassed level appropriateness, the appropriateness and distribution of themes, and pedagogical usefulness in developing learners' knowledge, pronunciation, interpretation skills, vocabulary, and knowledge of sentence structures, the other criteria regarding teaching activities involved the role of group discussion and interpretation activities in facilitating pre-translation preparation, text comprehension, skills of listening, interpreting, and summarizing, and learners' motivation. Given that technological competence had received little attention in translation textbook evaluation, Li (2022) adopted Wang and Qin's (2018) Translators' Technological Competence (TTC) to evaluate the effectiveness of translation textbooks in standardizing learners, following four components: technical knowledge, instrumental ability, information literacy, and technical thinking. In general, existing translation coursebook evaluation instruments contain general material-evaluation principles such as alignment with learning objectives and learner engagement promotion, alongside some translation-specific criteria. However, these instruments seem to overlook the growing prominence of AI and machine translation as well as increasing demands for post-editing skills.

As a result, the need to develop an evaluation checklist for translation coursebooks with a solid theoretical foundation that reflects requirements for translation competence in the modern age has emerged. Such a checklist will both help evaluate the relevance of a translation coursebook to the training objectives and assist training units in selecting or improving coursebooks. That was why the present study was conducted to develop a preliminary checklist for evaluating translation coursebooks and obtain expert appraisal of its criteria.

2. Theoretical Frameworks

One of the theoretical frameworks of this study is Cunningsworth's (1995) checklist, which is built on four core principles and could be specified and applied flexibly. Cunningsworth (1995) states that coursebooks should meet learners' needs by matching them with the aims and objectives of the language-learning programme. Besides, coursebooks should equip learners with the necessary knowledge and skills for their own language-using purposes. Coursebooks should also offer well-organized exercises and activities to support learners across their different proficiency levels and learning styles. At the same time, coursebooks should provide great support in students' self-study and teachers' lesson preparation (Cunningsworth, 1995). The present study combined this comprehensive model and a translation competence model introduced in the next section to devise the preliminary evaluation checklist.

Another theoretical framework utilized is the translation competence model designed by the PACTE research group. This model includes five main translation sub-competences (bilingual, extra-linguistic, knowledge of translation, instrumental, and strategic) and monitoring components. Specifically, the bilingual sub-competence involves procedural knowledge translators use to communicate in two languages, including "pragmatic, socio-linguistic, textual, grammatical, and lexical knowledge" in both languages (PACTE, 2017, p.39). The extra-linguistic sub-competence is characterized by translators' knowledge about the world, including cultures in the source language (SL) and target language (TL), general knowledge, and specialized subject areas. The knowledge of translation sub-competence is measured by the translators' understanding of what translation is and how it works, including methods, processes, common problems, and professional aspects such as the market and target audiences. The instrumental sub-competence refers to skills to use tools and resources for translation, such as dictionaries, search engines, and parallel texts. Meanwhile, the strategic sub-competence concerns the ability to manage and control the translation process, from planning, evaluating, identifying, and

solving problems to coordinating and activating all other sub-competences. These sub-competences integrate and activate a series of psycho-physiological mechanisms, comprised of cognitive characteristics, personal traits, and skills that support translation, such as memory, perseverance, and creativity. The present study was partly based on this model for the content section of the preliminary evaluation checklist, as it is one of the most recent and well-researched.

In addition, according to Tomlinson (2023), developing a set of criteria should be systematic and adaptable, beginning with a general set of criteria and subsequently refining it into a context-specific one. Specifically, this process typically starts with brainstorming a list of universal criteria grounded in the principles of language acquisition, findings from classroom-based research, and the evaluators' professional experience. Importantly, these criteria should be formed as specific assessment questions that explore the extent to which the materials affect learners, rather than merely describing their features. Subsequently, some of the criteria should be subdivided, and the set of universal criteria should be monitored and revised. This stage aims to ensure that each question is evaluative, clear, focused on a single matter, answerable, free of dogma, and sufficiently reliable to be interpreted consistently by different evaluators. After being checked and refined, the list is categorized. It is also crucial to trial the criteria before applying them in practice. Last but not least, the set should be utilized to evaluate a material or a typical unit for each level.

3. Methodology

3.1. Design of the Preliminary Evaluation Checklist

The preliminary evaluation checklist was designed in accordance with the procedures suggested by Tomlinson (2023). First, relevant theories and frameworks of materials evaluation were reviewed, particularly established frameworks such as Cunningsworth (1995), to identify the commonly recognized qualities and criteria for evaluating coursebooks. Based on this review, an initial pool of general criteria applicable to coursebook evaluation was compiled. Moreover, given the specific nature of translation coursebooks, literature on translation competence frameworks was subsequently reviewed, with particular reference to the PACTE (2017) model, to identify and incorporate criteria related to translation-specific competences.

The initial pool of criteria was then reviewed and refined by subdividing broader criteria where necessary and adjusting their wording in accordance with the essential features of a criterion proposed by Tomlinson (2023), particularly that a criterion should be evaluative, answerable, and specific. Criteria addressing similar aspects were consolidated where appropriate, while criteria considered insufficiently relevant to the evaluation of translation coursebooks were excluded. The remaining criteria were then organized into major categories according to the aspects of coursebook quality and translation competence they addressed. Following this process of identification, refinement, consolidation, and categorization, a preliminary evaluation checklist consisting of 50 criteria was produced for subsequent expert appraisal.

3.2. Expert Appraisal

To obtain expert appraisal of the preliminary checklist, translation lecturers were recruited through purposive and snowball sampling. The former helps researchers to select appropriate subjects and appears to be more cost-effective and easier to implement than random sampling; meanwhile, the latter is cost-effective and suitable for small samples, especially in studies with a limited number of relevant subjects (Dubey & Kothari, 2022). The participants were selected based on their involvement in translation teaching and their professional experience in translation education, with experience in translation coursebook development considered an additional indicator of relevant expertise rather than a mandatory requirement. This approach was adopted in consideration of the research context. In Vietnamese higher education, not many institutions offer translation training programs, and the formal development of translation coursebooks is not widespread across institutions. Consequently, restricting the expert sample exclusively to lecturers with substantial experience in developing translation coursebooks would have considerably limited the pool of eligible participants.

Specifically, the participants were 14 lecturers teaching translation at four public universities in Hanoi: seven from VNU University of Languages and International Studies, two from Hanoi University, four from Thuongmai University, and one from the School of Languages and Tourism at the University of Industry. Their experience in translation teaching varied: two had 3-4 years of experience, four had 5-10 years, seven had 11-20 years, and one had more than 20 years. Their experience in translation coursebook development also varied: five had no experience in coursebook development, seven had authored one or two coursebooks, and two had developed three to five coursebooks.

The preliminary checklist was used as a survey questionnaire, in which the lecturers were asked to evaluate each criterion across three aspects: relevance, clarity, and answerability on a 5-level Likert scale (1 = “very low” and 5 = “very high”). In this study, relevance refers to the extent to which a criterion is appropriate and pertinent to the evaluation of translation coursebooks and reflects an important aspect of their quality. Clarity refers to the extent to which a criterion is clearly and unambiguously worded, allowing its meaning to be readily understood by users of the checklist. Answerability refers to the extent to which a criterion can be meaningfully evaluated based on observable or identifiable information available in the coursebook. These aspects are common in assessing coursebook evaluation criteria, as introduced by Tomlinson (2023). The 1-5 scale not only facilitates experts’ evaluations but also supports quantitative data analysis, calculating the mean score and standard deviation (SD) for each criterion.

This survey was sent to the lecturers through Cognito Forms with a two-week deadline. The data collected were analyzed with descriptive statistics. SPSS Statistics was used to calculate the mean and SD for each criterion across three aspects: relevance, clarity, and answerability. To facilitate the interpretation of the experts’ ratings, explicit decision thresholds were established for the three evaluation aspects. Given that the questionnaire employed a five-point Likert scale, a mean score of 4.00 was set as the minimum threshold for a favorable evaluation, indicating that the average rating of a criterion was at least “high”. In addition, a standard deviation of 1.00 was used as an indicator of the degree of variation in the experts’ responses. Additionally, tables were utilized to present data visibly and systematically wherever needed.

Finally, for those criteria that saw noticeable variations in expert appraisal, further discussion with the involved experts was conducted to explore the underlying reasons. It is important to note that all responses were presented anonymously: L01 for Lecturer 01, L02 for Lecturer 02, and so on.

4. Results

4.1. The Preliminary Evaluation Checklist

As seen from Table 1, the preliminary evaluation checklist included 50 criteria, categorized into four groups: six in Group A: Objectives and approaches, seven in Group B: Teaching methods, assessment, and testing, four in Group C: Design and organization, and 33 in Group D: Content. The last group was subdivided into six sub-groups, with D.1 (General attributes about topics, source texts (ST), and translation tasks) including 11, D.2 (Bilingual competence) including seven, D.3 (Extra-linguistic competence) including three, D.4 (Translation knowledge competence) including four, D.5 (Instrumental competence) including three, and D.6 (Strategic competence) including five.

Table 1: The preliminary evaluation checklist

No.	Assessment criteria
	<i>Group A: Objectives and Approaches</i>
C01	The material has clear objectives.
C02	The material’s objectives align with those of the syllabus.
C03	The material’s objectives align with learners’ needs.
C04	The material aligns with the educational context.
C05	The material is helpful to teachers in designing activities and to students in improving their translation skills.

C06	The material is flexible and can accommodate a variety of different teaching and studying styles.
	<i>Group B: Teaching methods, assessment, and testing</i>
C07	The material incorporates teaching and learning approaches aligning with the specific educational context.
C08	The material includes the tasks that encourage learners' active engagement.
C09	The material includes the tasks that encourage learner autonomy.
C10	The material includes the tasks that encourage learners to self-evaluate their translation competence.
C11	The material includes the tasks that allocate the translation workload in proportion to the actual course duration.
C12	The material includes the tasks that can be completed flexibly in different formats: individually, in pairs, or in groups.
C13	The material includes tasks that provide the knowledge and skills required in assessment tests and have the same format as those in assessment and testing.
	<i>Group C: Design and organization</i>
C14	The content of the material is logically organized.
C15	The structure of each unit follows an ascending order of difficulty level.
C16	The material provides references and materials for self-study.
C17	The organization of the material is consistent, logical, clear, and easy to follow.
	<i>Group D: Content</i>
	<i>D.1. General attributes about topics, STs, and translation tasks</i>
C18	The topics of the STs are practical and align with real-world translation contexts.
C19	The topics of the STs are diverse and cover fields in accordance with the course's expected learning outcomes.
C20	The topics of the STs are new and updated.
C21	The topics of the STs are relevant to the cultural context of SL.
C22	Vocabulary, sentence structures, and linguistic features are suitable for the learners' level.
C23	The STs are diverse in text types.
C24	The STs are authentic, closely reflecting the real-world language use.
C25	The tasks are aligned with the material's objectives.
C26	The tasks are diverse, enabling learners to engage with a variety of exercise types and develop different skills.
C27	The tasks' difficulty level is suitable for the learners' level.
C28	The instructions for the tasks are clear and easy to understand.
	<i>D.2. Bilingual competence</i>
C29	The material includes the tasks requiring learners to use pragmatic conventions in both SL and TL to perform appropriate speech acts in specific communicative contexts.
C30	The material includes tasks addressing different registers in both the SL and TL.
C31	The material includes tasks addressing different dialects in both the SL and TL.
C32	The material includes the tasks related to cohesion and coherence in both the SL and TL.
C33	The material includes tasks addressing the structures, linguistic features, and conventions of different text types in both the SL and TL.
C34	The material includes the tasks related to syntactic structures in both the SL and TL.
C35	The material improves learners' vocabulary in both the SL and TL.
	<i>D.3. Extra-linguistic competence</i>
C36	The material provides cultural knowledge of both the SL and TL.
C37	The material provides foundational knowledge of general topics.
C38	The material provides specialized knowledge across various fields.
	<i>D.4. Translation knowledge competence</i>
C39	The material provides knowledge of different levels of translation (word, phrase, sentence, and text).
C40	The material provides knowledge of the stages in the translation process.
C41	The material provides knowledge of translation strategies and techniques.
C42	The material provides knowledge of common translation problems.
	<i>D.5. Instrumental competence</i>
C43	The material includes tasks requiring learners to use available resources to support their translation, such as dictionaries, encyclopedias, and parallel texts.

C44	The material includes tasks that involve using modern technology to support translation, such as search engines, machine translation tools, and AI.
C45	The material includes practice tasks that involve post-editing machine translation and AI outputs.
	<i>D.6. Strategic competence</i>
C46	The material includes practice tasks following each stage of the translation process.
C47	The material includes tasks that involve selecting and practicing translation strategies and techniques.
C48	The material includes tasks that encourage the evaluation of the translation process and the quality of the draft translation.
C49	The material includes tasks that require the flexible integration of knowledge, skills, resources, and translation tools.
C50	The material includes tasks that involve identifying and resolving translation problems.

4.2. Results from the Expert Appraisal

4.2.1. Group A: Objectives and Approaches

Table 2 demonstrates that the experts highly rated all criteria in Group A across three dimensions, with mean scores ranging from 4.00 to 4.86. Most notably, C02 and C05 were the most positively evaluated by the experts for both relevance (similar mean scores of 4.86) and clarity ($M = 4.71$ and $M = 4.64$, respectively). Regarding answerability, all the criteria in this group had slightly lower mean scores, especially C03 and C04, which shared the lowest mean scores of 4.00. Moreover, with SDs ranging from approximately 0.30 to 0.90, the criteria were shown to receive a high level of expert consensus.

Table 2: Assessment of criteria in Group A

Criteria	Relevance		Clarity		Answerability	
	M	SD	M	SD	M	SD
C01	4.71	0.47	4.50	0.86	4.57	0.51
C02	4.86	0.36	4.71	0.61	4.29	0.73
C03	4.64	0.50	4.29	0.73	4.00	0.88
C04	4.57	0.65	4.14	0.66	4.00	0.88
C05	4.86	0.36	4.64	0.63	4.43	0.65
C06	4.79	0.43	4.57	0.51	4.29	0.73

4.2.2. Group B: Teaching methods, Assessment, and Testing

Similar to those in Group A, all seven criteria in Group B received consistently positive evaluations from the experts, with C10 having the lowest mean score of 4.00 for answerability and C07 having the highest SD of 0.80 for clarity (see Table 3). The relevance ratings of these criteria were particularly high, with all mean scores being from 4.57 onwards. Regarding clarity and answerability, mean scores for the criteria in this group were slightly less robust, with C11 and C12 being the clearest ($M = 4.71$) and C12 being the most answerable ($M = 4.64$).

Table 3: Assessment of criteria in Group B

Criteria	Relevance		Clarity		Answerability	
	M	SD	M	SD	M	SD
C07	4.64	0.50	4.21	0.80	4.29	0.61
C08	4.64	0.63	4.50	0.65	4.14	0.77
C09	4.57	0.65	4.29	0.73	4.07	0.73
C10	4.57	0.65	4.36	0.63	4.00	0.78
C11	4.71	0.47	4.71	0.47	4.57	0.51
C12	4.64	0.63	4.71	0.61	4.64	0.63
C13	4.79	0.58	4.57	0.76	4.36	0.75

4.2.3. Group C: Design and Organization

It is apparent from Table 4 that, like Groups A and B, Group C was positively evaluated by the experts with a high level of consensus (all SDs consistently remaining under 0.64). Specifically, mean scores for relevance were extremely high, ranging from 4.79 to 5.00, with C14 achieving the highest; meanwhile, clarity and answerability had a similar mean range of 4.64 - 4.86, with C15 having the highest for clarity and C17 for answerability.

Table 4: Assessment of criteria in Group C

Criteria	Relevance		Clarity		Answerability	
	M	SD	M	SD	M	SD
C14	5.00	0.00	4.64	0.50	4.64	0.50
C15	4.86	0.36	4.86	0.36	4.64	0.50
C16	4.79	0.58	4.64	0.63	4.64	0.63
C17	4.86	0.36	4.79	0.43	4.86	0.36

4.2.4. Group D.1: Content_General attributes about topics, STs, and translation tasks

Table 5 shows that all criteria in Group D.1, except C20, received overall positive ratings, with C19 obtaining the lowest mean score of 4.14 for answerability and C18 having the highest SD of 0.84 for answerability. C20 was the only criterion with a mean score below 4.00 and an SD above 1.00 on one aspect, indicating its lower answerability and relatively diverse perspectives among experts.

Table 5: Assessment of criteria in Group D.1

Criteria	Relevance		Clarity		Answerability	
	M	SD	M	SD	M	SD
C18	4.86	0.36	4.50	0.52	4.36	0.84
C19	4.79	0.58	4.57	0.65	4.29	0.83
C20	4.36	0.75	4.36	0.75	3.79	1.05
C21	4.50	0.76	4.50	0.65	4.36	0.63
C22	4.50	0.76	4.57	0.76	4.14	0.77
C23	4.64	0.63	4.64	0.63	4.43	0.76
C24	4.64	0.63	4.64	0.63	4.43	0.65
C25	4.79	0.43	4.79	0.43	4.71	0.47
C26	4.71	0.49	4.57	0.51	4.64	0.49
C27	4.57	0.65	4.57	0.51	4.21	0.70
C28	4.86	0.36	4.86	0.36	4.79	0.58

4.2.5. Group D.2: Content_Bilingual competence

As shown in Table 6, five of the criteria in this group were generally positively assessed, with C35 emerging as the highest-rated criterion with mean scores of 4.71, 4.50, and 4.43 on the three aspects (SDs around 0.85). Additionally, while C31's clarity evaluation was consistently good, this criterion was evaluated as a little less relevant and answerable, with a slightly wider spread of opinions among the experts (Relevance: M = 3.86, SD = 1.10; Answerability: M = 3.93, SD = 0.62). Moreover, despite achieving satisfactory ratings for clarity and answerability (M > 4.00, SD < 1.00), C32 had a slightly elevated standard deviation of 1.02 for relevance.

Table 6: Assessment of criteria in Group D.2

Criteria	Relevance		Clarity		Answerability	
	M	SD	M	SD	M	SD
C29	4.36	0.75	4.36	0.75	4.14	0.66
C30	4.36	0.75	4.50	0.65	4.14	0.66
C31	3.86	1.10	4.36	0.75	3.93	0.62
C32	4.43	1.02	4.14	0.95	4.14	0.66
C33	4.36	0.93	4.07	0.92	4.07	0.73
C34	4.50	0.94	4.36	0.93	4.29	0.73
C35	4.71	0.83	4.50	0.86	4.43	0.85

4.2.6. Group D.3: Content_Extra-linguistic competence

All the criteria in Group D.3 received satisfactory ratings (see Table 7). Relevance's mean scores were either 4.50 or 4.57. In terms of clarity, C38 achieved the highest mean score (M = 4.57). Regarding answerability, C37 recorded the highest mean score (M = 4.43). All these criteria's SDs were under 0.90 across the three dimensions, securing a relatively high level of consensus among the experts.

Table 7: Assessment of criteria in Group D.3

Criteria	Relevance		Clarity		Answerability	
	M	SD	M	SD	M	SD
C36	4.57	0.85	4.43	0.85	4.36	0.84
C37	4.50	0.65	4.43	0.65	4.43	0.65
C38	4.57	0.65	4.57	0.65	4.21	0.89

4.2.7. Group D.4: Content_ Translation knowledge competence

In Group D.4, C40 and C42 had satisfactory mean scores and SDs (see Table 8). By contrast, the other two saw a wider distribution of experts' opinions. Specifically, despite high mean scores for relevance (M = 4.50) and clarity (M = 4.43), C39's standard deviations exceeded 1.00. Similarly, in terms of relevance, C41 had a slightly high SD of more than 1.00.

Table 8: Assessment of criteria in Group D.4

Criteria	Relevance		Clarity		Answerability	
	M	SD	M	SD	M	SD
C39	4.50	1.09	4.43	1.09	4.57	0.85
C40	4.43	0.94	4.29	0.91	4.64	0.63
C41	4.36	1.01	4.43	0.94	4.43	0.76
C42	4.57	0.76	4.43	0.76	4.29	0.61

4.2.8. Group D.5: Content_ Instrumental competence

Table 9 reveals that all three criteria in Group D.5 were generally assessed positively by the experts, with C45 rated the highest on the three aspects (M = 4.64, 4.64, and 4.50). This group also demonstrated a high level of consensus among the experts, with SD values consistently below 1.00 across all three aspects.

Table 9: Assessment of criteria in Group D.5

Criteria	Relevance		Clarity		Answerability	
	M	SD	M	SD	M	SD
C43	4.43	0.94	4.36	0.93	4.29	0.83
C44	4.64	0.63	4.43	0.65	4.29	0.73
C45	4.64	0.63	4.64	0.63	4.50	0.65

4.2.9. Group D.6: Content_ Strategic competence

Like group D.5, the expert evaluations for Group D.6 were generally positive across the three dimensions (see Table 10). C48 received the highest overall ratings with the highest relevance and clarity mean scores (4.64 and 4.50, respectively), while the most answerable criterion in the group was C47 (M = 4.57). This criterion group also recorded a high level of consensus among the experts, with SDs staying below 1.0.

Table 10: Assessment of criteria in Group D.6

Criteria	Relevance		Clarity		Answerability	
	M	SD	M	SD	M	SD
C46	4.14	0.86	4.43	0.76	4.43	0.65
C47	4.36	0.84	4.36	0.75	4.57	0.65
C48	4.64	0.63	4.50	0.65	4.43	0.65
C49	4.57	0.85	4.43	0.94	4.29	0.83
C50	4.43	0.85	4.36	0.84	4.21	0.89

4.2.10. Summary of results from the expert appraisal and further discussion with the experts

In general, the experts gave positive ratings and showed a high level of consensus for the checklist. In total, 45 of 50 criteria had mean scores of 4.0 or higher and SDs of 1.0 or lower. The others, namely C20, C31, C32, C39, and C41, received several assessments that slightly diverged from the majority.

Further discussion with the involved experts revealed reasons for their low ratings for these criteria. As for C20, L06 suggested that it was unrealistic and unanswerable, as "the topics that are updated at the time of coursebook design will become outdated." This lecturer also gave low ratings for the relevance

of C31 and C32 since *“translation coursebooks should focus on the standard language, and dialects are not general with rather low applicability,”* and *“C32 mentions certain atypical factors depending much on the textual context.”* Meanwhile, L07 recommended revising C39 since *“sometimes translation doesn’t simply mean finding equivalents at the word or phrase level”*, while L06 stated that this criterion should be excluded from the checklist as it was associated with the contents of coursebooks for another course. L06 also referred to the same suggestion and reason for C41. Following further discussion with the experts, the researchers decided to retain C32 and C41 while removing the remaining criteria.

5. Discussion

Different from existing translation-coursebook evaluation instruments, which either focus on general principles of textbook evaluation like Poulis and Ali’s (2020) or incorporate translation-specific criteria related to the development of students’ interpretation, summarizing, and technological skills such as Ho’s (2017) and Li’s (2022), the preliminary 50-criterion checklist proposed in this study covers emerging skills required in the modern age such as AI use and post-editing skills.

The present study shares certain similarities with Khalid et al. (2009), Nimehchisalem and Mukundan (2015), and Zokaeieh et al. (2019) in data collection methods and procedures. Specifically, an initial set of criteria was generated through a review of existing literature before being refined through expert feedback. However, these studies differ in their theoretical foundations. While Khalid et al. (2009) developed their checklist mainly based on Garvin’s model, and the prototype of Nimehchisalem and Mukundan (2015) was designed based merely on related literature on ELT textbook evaluation, the present study drew on both previous findings and the PACTE model of translation competence.

In addition, Nimehchisalem and Mukundan (2015) and this research differed in the data collection. While the former relied heavily on qualitative feedback from focus groups and multiple rounds of expert review, the latter placed a greater emphasis on quantitative evidence. Besides, to collect the quantitative data, whereas Nimehchisalem and Mukundan (2015) asked ELT teachers to rate their agreement with each statement in the survey, experts in the present study were asked to rate the relevance, clarity, and answerability of each criterion. While the quantitative data in Nimehchisalem and Mukundan’s (2015) study were used to assess validity, reliability, impact, and practicality, this study sought expert consensus on the appropriateness and usability of the criteria.

Notably, in terms of data analysis procedures, different from this research, Zokaeieh et al. (2019) focused on establishing the construct validity of their critical textbook evaluation checklist through Exploratory Factor Analysis (EFA) and Partial Confirmatory Factor Analysis (PCFA). These differences reflect the distinct research purposes. Whereas Zokaeieh et al. (2019) partly aimed to validate the dimensional structure of an existing instrument, the current study prioritized content validity to ensure that the checklist items were relevant, comprehensible, and applicable to intended users.

6. Conclusion and implications

In conclusion, the study developed an evaluation checklist for translation coursebooks. A preliminary 50-criterion checklist was developed based on Cunningsworth’s (1995) checklist, PACTE’s (2017) translation competence model, and Tomlinson’s (2023) checklist development procedures. The checklist received generally favorable ratings from the experts, while five criteria showed comparatively greater variation in expert responses and warrant further examination. Further discussions with the experts resulted in the retention of two out of five disputed criteria. These findings provide preliminary evidence supporting the potential usefulness of the checklist for translation coursebook evaluation. However, further empirical validation and testing with actual translation coursebooks and evaluators are needed to establish its validity and practical usefulness in university translation training.

The checklist proposed in this study not only addresses traditional textbook qualities, such as content organization and learning objectives, but also incorporates criteria related to learners’ AI literacy and technology-related competences. Therefore, it has the potential to inform the selection and development of teaching materials that better prepare learners for translation work in the AI age,

where human translators need to be able to use these tools effectively, develop their post-editing skills, and understand technology-related ethical issues in translation.

However, this study has some limitations that warrant future research. Firstly, as the checklist was developed based on expert judgements and has not been piloted, its usability, effectiveness, and practicality in real-world educational settings remain unexplored. Therefore, further research should be conducted to validate the checklist. In addition, although the present research evaluated the relevance of each criterion to the translation textbook assessment, as well as its comprehensibility and applicability, the extent to which the criteria cluster into the intended dimensions of the checklist remains unverified. Hence, future studies should employ EFA, followed by PCFA, to determine whether the existing grouping of criteria adequately represents the construct of translation coursebook quality in the era of AI.

Acknowledgement: This research was funded by VNU University of Languages and International Studies (VNU-ULIS) in the project No. N.25.26.

References

- Çelenk, S., & Er, G. (2025). An EFL coursebook evaluation: Teachers' perspectives on the 3rd grade English coursebook. *Eurasian Journal of Language Teaching & Linguistic Studies (EAJLTLS)*, 5(1), 581-607. <https://eajltls.com/index.php/home/article/view/109>
- Cunningsworth, A. (1995). *Choosing your coursebook*. Heinemann.
- Dubey, U. K. B., & Kothari, D. P. (2022). *Research methodology: Techniques and trends*. CRC Press.
- Garvin, D. A. (1988). *Managing quality: The strategic and competitive advantage*. Free Press.
- Ho, P. V. P. (2017). An evaluation on the course-book of translation and interpretation. *Van Hien University Journal of Science*, 5(2), 4-13. <https://doi.org/10.58810/vhujs.5.2.2017.5227>
- Islam, M. M. (2013). An evaluation of English language textbook from teachers' perspectives. *International Journal of English and Education*, 2(3), 117-129.
- Khalid, M., Muhammad, Z. I., & Muhammad, S. (2009). Textbook evaluation through quality indicators: The case of Pakistan. *Bulletin of Education and Research*, 31(2), 1-27. <https://pu.edu.pk/images/journal/pesr/PDF-FILES/1-Textbook%20evaluation%20article%20%28improved%29.pdf?utm>
- Li, H. (2022). Evaluation of translation technology textbooks from the perspective of translator's technological competence. *Journal of Language Teaching*, 2(5), 1-5. <https://doi.org/10.54475/jlt.2022.005>
- Litz, D. (2001). Textbook evaluation and ELT management: A South Korean case study. *Asian EFL Journal*, 48, 1-53.
- McDonough, J., Shaw, C., & Masuhara, H. (2013). *Materials and methods in ELT: A teacher's guide* (3rd ed.). Blackwell.
- Mukundan, J., Nimehchisalem, V., & Hajimohammadi, R. (2011). Developing an English language textbook evaluation checklist: A focus group study. *International Journal of Humanities and Social Science*, 1(12), 100-106.
- Nimehchisalem, V., & Mukundan, J. (2015). Refinement of the English language teaching textbook evaluation checklist. *Pertanika Journal of Social Sciences & Humanities*, 23(4), 761-780.
- PACTE (2017). PACTE translation competence model. In A. H. Albir (Ed.), *Researching translation competence by PACTE group* (pp. 35-41). John Benjamins.
- Poulis, V. M., & Ali, W. I. (2020). Students' assessment of the translation material in the Departments of English in Kurdistan region-Iraq. *Journal of University of Raparin*, 7(4), 138-155.
- Rijoly, H. M. (2023). Evaluating and developing textbook for translation class based on students' satisfaction and need analysis. *Journal of Language Education Research*, 5(2), 45-58.
- Satriani, I., & Resmini, S. (2021). ESP coursebook evaluation: Students' perception on activities and tasks. *ELT in Focus*, 4(1), 20-25.
- Tomlinson, B. (2023). Materials evaluation. In B. Tomlinson (Ed.), *Developing materials for language teaching* (3rd ed., pp. 25-57). Continuum.
- Vy, V. L. (2025). Examining Vietnamese EFL teachers' perceived impact of the "Global Success" textbooks on their teaching practices and learners' language proficiency in public secondary schools: A cross-case study. *Canadian Journal for New Scholars in Education/ Revue canadienne des jeunes chercheurs et chercheurs en éducation*, 16(1), 86-104.
- Wang, S., & Qin, J. (2018). Constructing technological competence system for translators in the context of big data. *Technology Enhanced Foreign Language Education*, (1), 90-96.
- Zohrabi, M. (2011). Coursebook development and evaluation for English for general purposes course. *English Language Teaching*, 4(2), 213-222. <https://doi.org/10.5539/elt.v4n2p213>
- Zokaeieh, S., Karimi, S. N., Nouri, F., & Hakimzadeh, A. (2019). Developing a critical checklist for textbook evaluation. *Journal of English Language Pedagogy and Practice*, 12(24), 202-225.