The effect of teacher experience and training on translation assessment: Results from the EACT project

José Tomás Conde Ruano
University of the Basque Country, Spain
tomas.conde@ehu.eus

Christian Olalla-Soler
Universitat Autònoma de Barcelona, Spain
christian.olalla@uab.cat

Abstract: This paper presents the results of a study that explores the relationship between the background of translator trainers in Spain and their attitudes towards and practices in assessment. A sample of 97 translator trainers from five Spanish universities completed an online questionnaire designed to gather information about such attitudes and practices. Several questionnaire responses that provided information about the participants' background characteristics and the attitudes and practices in question were analysed using tests of association. The data indicate the impact of two factors: having received training in translation teaching and having experience as a professional translator. Both seemed related to greater awareness of the importance of assessment within translator training and to a willingness to improve and adapt assessment practices over time. Even so, the relatively few statistically significant associations observed, combined with generally small-to-moderate effect sizes, suggest substantial individual and contextual variability, potentially driven by unmeasured internal factors such as preferences, intuition, prior experiences, as well as external conditions such as language pairs and cohort sizes. At the same time, given the study's limited power to detect small effects, we cannot exclude more subtle relationships that may emerge in larger samples or more fine-grained designs. Establishing whether such background factors are associated with assessment is relevant for designing targeted educator development and for improving the consistency and professional alignment of assessment in translator education.

Keywords: assessment, assessment practices, attitudes towards assessment, translator training.

1. Introduction

In this paper, we present the results of a study on the relationship between the background of translator trainers in Spain and their attitudes towards and practices in assessment. Assessment is inherently subjective (Conde Ruano, 2008), as "there are almost as many translation criteria as translation teachers" (Li, 2006: 84), and as many ways to assess as there are evaluators. Despite extensive research on translation evaluation, there remains a significant gap in understanding whether assessment approaches are associated with personal and professional traits. This paper aims to address this gap.

Our study is part of a broader investigation into the evaluation of translation practices in Spanish translator training programmes: the Evaluation in the Acquisition of Translation Competence (EACT) research project,[1] which aims to: (1) describe the assessment procedures used in BA translation courses in Spain; and (2) establish assessment procedures for different levels of translation competence acquisition in the English/French/German to Basque/Catalan/Galician/Spanish language pairs. These procedures are based on the translation competence level framework developed in the NACT project.[2] This framework defines descriptors for five competences – language, extralinguistic, instrumental, service provision and translation problem solving – and five acquisition levels (A1 and A2: introduction to translation; B1 and B2: introduction to specialised translation and C: specialised translation) (PACTE, 2018; Hurtado Albir and Rodríguez-Inés, 2022). The EACT project has designed assessment procedures for levels A1 to B1.

Its primary goal, describing the assessment procedures used in BA translation courses in Spain, is divided into four specific objectives:

1. To explore the current conception of assessment held by translator trainers (see Huertas-Barros and Domínguez Araújo, 2024).
2. To describe the assessment procedures adopted in practical translation courses: assessment criteria, task types, translation grading scales, rubrics (see Gay Punzano and Hurtado Albir, 2025).
3. To determine whether assessment procedures vary depending on the language pair and the linguistic and cultural context involved.
4. To explore whether there is an association between the background characteristics of translator trainers, defined as lecturers who teach in BA programmes in Translation and Interpreting in Spain, and their attitudes towards and practices in assessment (this paper).

The present study focuses on the fourth objective, examining factors such as age, professional experience, training in assessment, and whether these are associated with trainers' views on the importance of assessment, satisfaction with assessment methods, use of rubrics, and adaptation of tasks and procedures.

This paper is structured as follows. Section 2 provides an overview of studies on assessment practices and their main results. Section 3 presents the methodological framework (questionnaire design, sampling procedure, sample characteristics, statistical analysis, and ethical considerations). Section 4 presents and discusses the findings. Section 5 outlines the most relevant conclusions and describes the study's limitations.

2. Background

Translation evaluation has historically evolved alongside the concept of quality, with both sharing common roots. In Translation Studies, two works are often highlighted as marking the beginning of reflection on translation evaluation

(McAlester, 2000): *Möglichkeiten und Grenzen der Übersetzungskritik* (Reiss, 1971) and *Towards a Model of Translation Quality Assessment* (House, 1977, 1997 2nd ed.). Reiβ focuses on establishing categories and criteria for a systematic evaluation of translations, emphasizing the importance of text types and functions. House's model provides a comprehensive framework for translation quality assessment based on detailed textual and contextual analysis, highlighting the significance of both linguistic and situational factors.

Subsequent work shifted toward the practice of assessment, focusing on how evaluators apply criteria and areas where inconsistencies arise. This represented a move away from mere model-building toward studying assessment-in-use. Early studies documented recurrent issues (Arango-Keeth and Koby, 2003) such as the absence of a proper terminology, the coexistence of different procedures depending on the theoretical approach chosen, and persistent conceptual divergences regarding what 'competence' comprises and how it is broken down into sub-competences and their interrelationships. Other researchers considered subjectivity (Brunette, 2000; Delisle, 2001) and the lack of universal standards and models (Bowker, 2000; Martínez and Hurtado Albir, 2001; Hovy et al., 2002; Rodríguez Verbo, 2004) as the most significant problems.

For some authors (Künzli, 2001; Conde Ruano, 2009), this body of work was marked by a lack of scientific rigor. Up to that point in the early 21st century, only approaches based on the author's subjectivity had been proposed (Bowker, 2000). These approaches were primarily prescriptive or theoretical in nature, lacking empirical evidence and scientific methodology. Subsequent studies began to examine the phenomenon of translation evaluation with a more empirical approach.

Below, we review several empirical studies on translation competence assessment. So far, these approaches have been limited in number and confined to narrow institutional or methodological scopes, despite the crucial role of assessment procedures in measuring the process of developing translation competence in higher education.

The empirical studies reviewed can be organised around five non-exclusive foci: (i) how assessment supports specific learning outcomes in translator training; (ii) contrasts between product-oriented and process-oriented evaluation; (iii) the materials and instruments used in assessment (e.g., texts, rubrics); (iv) participants' perceptions of assessment procedures; and (v) the evolution of evaluative practices over time and their alignment with professional standards. As part of the first focus, Domínguez Araújo (2015) conducted research on the assessment practices of conference interpreter trainers in Spain using interviews, questionnaires, and direct observation. The results of her study, in which both trainers and students participated, demonstrated the relevance of assessment in the development of students' metacognition, self-regulation and learning autonomy. The study also brought to light the need to foster communication between trainers and students to achieve agreement on learning objectives and assessment needs and procedures. Additionally, it pointed to differences in the conception of good assessment, where trainers focused on the importance of fostering students' self-regulation and motivation and encouraging self-assessment, while students believed that good assessment should be ipsative and objective.

Haro-Soler's (2018) study also belongs to the first focus. She investigated the self-efficacy beliefs of students of the Universidad de Granada's BA in translation by means of questionnaires and focus groups. The main result was the positive effect on the students' self-efficacy beliefs of group projects carried out outside the classroom. Constructive feedback and the use of rubrics and

clear assessment criteria also had a positive influence on the students' self-efficacy beliefs.

Another contribution to this focus is Gardy (2016), who focused on the Canadian academic context and studied translation assessment practices from 2000 to 2015 by means of semi-structured interviews and questionnaires that aimed to collect data from both trainers and students. The data also unveiled that assessment practices had a strong effect on the students' stress levels, leading the author to conclude that such practices should be adapted to current assessment trends and the profession.

Shifting to the second focus (process- vs. product-oriented evaluation), Gardy (2016) found that most of the trainers conceived of assessment as product-oriented and as being geared towards monitoring the students' learning progression.

Also within the second focus, complementing Gardy's findings, Huertas-Barros and Vine (2018) explored the assessment practices used in 27 translation and interpreting master level programmes in the United Kingdom through questionnaires and curriculum analysis. They observed a sector-level tilt toward process-oriented assessment, which they link to programmes' use of translation-competence models as curricular foundations.

In his research on the academic assessment of translations, Conde Ruano (2009) used a questionnaire and analysed the marked-up translations produced by the participants, including the changes, comments and annotations they made while evaluating the texts. He gathered data from 88 participants of different profiles: potential users of translated products, professional translators, translator trainers, and trainee translators. His findings showed that text characteristics affected assessment conception and that participants with different profiles approached assessment differently. Conde Ruano's (2009) study revealed that when participants evaluated several translations consecutively, they seemed to experience fatigue, leading them to conserve their efforts. He also found that they seemed to memorise the original text as they progressed. Interestingly, despite leaving fewer corrections and comments on successive texts, this did not result in noticeable variation in the overall perceived quality of the translations.

Shifting to the third focus, within the empirical literature the largest number of studies concentrates on identifying the materials (mostly texts), criteria, and tools used to assess translations, rather than analysing the actual behaviour of evaluators. An early attempt to explore assessment practices was carried out by Martínez Melis (1996), preceding the move toward more systematic empirical research. In her study, Martínez Melis (1996) surveyed translator trainers at the Faculty of Translation and Interpreting of the Universitat Autònoma de Barcelona to examine how translation courses were being assessed. She found that the most important assessment criteria related to the linguistic quality of translations, the ability to solve translation problems, and the ability to accurately transfer source text information.

A later work, similar in participant profiles and objectives, is that of Waddington (2001), who conducted a survey among translator trainers from European and Canadian universities. The primary assessment task was translating a text, with only half of the trainers including other complementary tests. Regarding criteria for correcting translation tasks, the findings showed that trainers from the same institution did not share rubrics or other assessment tools, with most focusing on identifying translation errors and/or using a holistic, impression-based method to assess translations.

More recently, Pavani (2016) investigated assessment practices in BA translation programmes in Spain and Italy through a questionnaire for translator

trainers. The results uncovered that assessments were based on the translation of one or many texts in class, complemented by tasks such as revising a translation, sight translating, creating terminological resources, and comparing translations. Assessment scales were very frequently used to categorise translation errors. Additionally, Pavani assessed the efficacy of a multi-dimensional assessment system for BA level translation courses.

In relation to participants' perceptions of assessment, an early contribution was Martínez Melis (1996), noted above. Her results showed that trainers were satisfied with their assessment practices, which focused on more than merely identifying and penalizing translation errors.

In Huertas-Barros and Vine (2016), both trainers' and students' perceptions of assessment practices in the MA in translation at the University of Westminster were examined through a questionnaire. The evidence appeared to confirm a mismatch between the confidence the two groups had in the assessment tools employed.

With respect to the fifth focus, relating to the evolution of evaluative practices, Huertas-Barros and Vine (2018) found significant variability in assessment practices across UK universities, leading to inconsistencies in how translation quality was judged. They concluded that these universities would benefit from providing training in assessment practices to make trainers feel more confident in their judgement and ensure their assessment criteria corresponded to those used in professional practice.

Responding to this identified need, Huertas-Barros and Vine (2019a, 2019b) designed methods and strategies to facilitate the inclusion of training in assessment in academic programmes. They also investigated the effects of promoting dialogue between trainers and trainees for the collaborative definition of assessment criteria, concluding that this practice fostered clarity in what is to be assessed and how. Gardy (2016) found that assessment practices had remained largely unchanged over time and did not reflect the evolution of the profession.

In summary, previous research indicates recurring themes in the empirical study of translation evaluation, to wit the relevance of the activity, its evolution, the object of evaluation, the behaviour and opinions of evaluators, and the materials, criteria, and tools they use. However, we still do not know whether trainers' personal and professional backgrounds shape their assessment views and behaviours. This is important because, if background factors do structure attitudes and practices, programmes can target staff development and policy to improve consistency, transparency, and alignment with competence-based curricula and professional expectations. The next section addresses this gap.

3. Methodology

This section describes the methodology used. It includes information on the sample, the data collection tool, the selection of items for analysis, and the statistical analysis.

3.1. Sample
The sample comprised translator trainers working in the English/French/German to Basque/Catalan/Galician/Spanish language pairs in BA programmes at five Spanish universities: Universidad del País Vasco, Universidade de Vigo, Universitat Jaume I, Universitat Autònoma de Barcelona, and Universidad de Granada. These universities were selected because they are the only Spanish members of the research project's consortium. Currently, more than thirty

Spanish universities offer programmes in translation and interpreting. Ortega Arjonilla & Fuentes Morán (2017) identified thirty universities offering such programmes at the time of their study; since then, at least four additional programmes have been launched.

A total of 163 trainers were invited to complete an online questionnaire (see section 3.2). Non-probabilistic sampling methods, specifically convenience and purposive sampling, were used (Campbell et al. 2020). The curricula of each university were analysed to identify potential respondents. Eligible participants were required to teach written translation into or from any of Spain's official languages (Spanish, Catalan, Galician and Basque), be it direct or inverse translation, in any year of a four-year degree. Initial contact was made through the Dean's office or departmental administration staff, followed by a direct email to each potential respondent.

3.2. Data collection tool

To collect data on trainers' attitudes and practices in assessment, an online questionnaire was designed in three phases. The first phase involved drafting the initial version, with some items extracted from previous questionnaires (Conde Ruano, 2009; Domínguez Araújo, 2015; Huertas-Barros and Vine, 2016, 2018; Pavani, 2016, among others). New elements were included to address the unique aspects of this study, which compares five university environments, including bilingual and non-bilingual regions,[3] and institutions with varying elective course options, providing a more comprehensive picture of assessment practices across diverse contexts.

The second phase involved testing. Three tests were conducted: an internal pilot test in July 2020 by 11 research team members (all affiliated with the five universities in the consortium), an assessment of the scale by four external translator trainers from other Spanish universities in September 2020, and a pilot test in October 2020 with 12 translator trainers from a university outside the consortium. To mitigate potential conflicts of interest, all participants provided informed consent, their feedback was anonymized and was used exclusively to refine the questionnaire. These tests focused on (1) question types and possible answers; (2) question wording and order; (3) mandatory questions; (4) the adequacy of the introduction; (5) the usability of the online tool built using LimeSurvey; and (6) the estimated completion time.

The third phase involved obtaining ethical approval for the questionnaire and related materials prior to implementation. The questionnaire, informed consent form, and data anonymization and management protocol were submitted to the Research Ethics Committee of the Universitat Autònoma de Barcelona, which approved it on December 11, 2020. Ethical approval was sought from this committee because the Universitat Autònoma de Barcelona was the coordinating institution of the research project.

The final version[4], titled "Questionnaire on current assessment practices in translation degree programmes in Spain", consisted of four sections: (1) programme-level assessment practices in the practical written translation modules; (2) module-level assessment for each practical written translation module taught by the respondent; (3) final remarks; and (4) demographic data.

---

[3] Alongside Spanish, some autonomous communities have a co-official language: Catalan (Catalonia, Balearic Islands), Valencian (Valencian Community), Basque (Basque Country and Basque-speaking areas of Navarre), and Galician (Galicia).
[4] The questionnaire was developed and administered in Spanish (available at: https://ddd.uab.cat/record/272353?ln=en). For reporting purposes, the items are presented here in English translation.

Questions were nested, so not all participants answered the same number of questions, which also varied depending on the number of modules. The basic number of questions was 39 distributed as follows: 17 questions in section 1, followed by 9 in section 2, 2 in section 3, and 11 in section 4. The questionnaire was designed and implemented in Spanish.

Most questions were either multiple-choice (allowing single or multiple selections) or employed a four-point Likert scale. A "Don't know/Prefer not to answer" ("DK/DA") option was provided. This aligns with Adelson and McCoach (2010), who found that the absence of a midpoint did not compromise data integrity. Removing the midpoint can decrease social desirability bias, leading to more genuine responses (Garland, 1991). Including a "DK/DA" option addresses concerns about accommodating genuinely neutral options (Tsang 2012) and avoids ambiguity associated with midpoints (Raaijmakers et al., 2000). This approach aims to enhance response clarity while accommodating those who cannot or prefer not to choose from among the options provided, minimising biases and interpretative challenges. Open-ended questions were kept to a minimum and included solely to elicit additional information or clarifications.[5]

3.3. Statistical analysis

Data analysis primarily involved computing descriptive statistics and, to a lesser extent, non-parametric group-difference tests and tests of association, depending on the data type. All tests were carried out using *jamovi* (version 2.4). For descriptive statistics (mainly reported in Table 1), we computed the mean, median, maximum, minimum, and standard deviation for ordinal data (e.g., age). For data from four-point Likert scale items (DK/DA responses treated as missing data), we computed the mean, median, and standard deviation. For categorical data, we used frequencies and percentages.

For inferential statistics, the procedure depended on the data type in the variables being cross-tabulated (Pardo and Ruiz, 2012; Pardo and San Martín, 2010; Pardo, Ruiz and San Martín, 2009). When comparing age groups (ordinal, treated as groups) with items answered on a four-point Likert scale, we used the Kruskal-Wallis H test (omnibus) as the omnibus test of distributional differences across groups; the effect size reported was ε^2. Planned post-hoc pairwise comparisons were to be conducted only if the omnibus test was significant. No Kruskal-Wallis test reached significance, so no post-hoc tests were run.

When the data being addressed were dichotomous, we assessed associations using Fisher's exact test, reporting phi as the effect size. Fisher's exact test is valid for all sample sizes and is preferable when expected cell counts are small (e.g., >20% of expected frequencies <5, or any expected frequency <1), conditions that applied in several of our tables (Kim, 2017). For categorical data with more than two categories in at least one item, we performed a chi-square test of independence (with continuity correction if needed) to assess associations, using Cramér's V as the effect size measure. Effect sizes (ε^2, phi, V) are interpreted descriptively to gauge magnitude of association, not causality.

We set the significance threshold at $\alpha = 0.05$, following standard practice in exploratory research (Maier and Lakens, 2022). Given the exploratory nature of our investigation, which seeks to identify potential relationships between translator trainers' background characteristics and their assessment practices rather than to test pre-registered hypotheses, further reducing α would unduly

---

[5] The questionnaire can be accessed at: https://ddd.uab.cat/record/272353?ln=en.

increase the risk of Type II errors, particularly with our sample size of 97 participants. Following Maier and Lakens' (2022) recommendation to balance Type I and Type II errors as a function of sample size, a sample of this size would in fact warrant an optimal α slightly above 0.05 (around 0.068). Given our sample size of 97 respondents, the cross-tabulation analyses have reasonable power to detect medium-sized associations (e.g. phi/Cramér's V around 0.30) at $\alpha = 0.05$, but more limited power to detect small effects (around 0.10) (Cohen, 1988). Non-significant results are therefore interpreted as suggesting the absence of strong, systematic associations in this sample, while acknowledging that smaller effects may have gone undetected.

In this context, we did not apply formal corrections for multiple comparisons (e.g. Bonferroni adjustments). In line with recommendations for exploratory work (Armstrong, 2014), we treated the analyses as hypothesis-generating rather than strictly confirmatory: strict control of the family-wise error rate is less critical here than avoiding an excessive inflation of Type II errors, especially given that many of the tested relationships are not independent and involve overlapping variables. In addition, our sample size provides only moderate statistical power, so further lowering the α level would substantially reduce our ability to detect small-to-moderate effects. For example, a Bonferroni adjustment for the 26 tests reported here would yield an effective α of approximately .002 (0.05/26), making it unlikely to identify even moderate associations that may nonetheless be of practical or theoretical interest. We therefore report exact p-values and effect sizes and interpret the overall pattern of results with appropriate caution, viewing the findings as a basis for future confirmatory studies in which more stringent error control procedures may be warranted. Given that effect sizes are important, we prioritise reporting and discussing both statistical significance and effect size measures following the threshold levels in Table 1.

Instead of applying formal corrections, we acknowledge the number of comparisons made and interpret results with caution. By maintaining an α of 0.05 and refraining from overly conservative corrections, we balance the need for statistical rigor with the exploratory nature of this study, ensuring that potentially meaningful relationships are not overlooked. Future research may build upon these findings with larger samples and pre-registered hypotheses, where confirmatory statistical approaches, including multiple comparison corrections, may be more appropriate.

Table 1: Interpretation of effect sizes

Effect	phi (based on Pearson's r; Cohen, 1988)	Cramér's V (Cohen, 1988)	ε^2 (Field, 2013)
Small	Up to 0.29	Up to 0.29	Up to 0.06
Medium	Up to 0.49	Up to 0.49	Up to 0.14
Large	0.5 or over	0.5 or over	0.15 or over

All statistical procedures were performed in *jamovi* (v. 2.4). All the two-way tables used to compute Fisher's exact and chi-squared tests can be found in Supplemental Material 2.

4. Results and analysis

Findings are reported as follows: First, we describe the characteristics of the sample. Second, we present the selection of elements considered for analysis,

with emphasis on participants' demographic information. Third, we provide results of cross-tabulating age with selected items related to participants' attitudes and practices. Fourth, we focus on relationships between attitudes and practices, and between those practices themselves. Fifth, we present results concerning the relationship between knowledge of the profession and participants' attitudes and practices. Sixth, we examine how participants' experience as translator trainers and professional translators relates to their attitudes and practices. The last subsection explores the relationship between the training received by trainers and their attitudes and practices.

4.1. Sample characteristics
Of the 163 trainers invited, 97 (59.9%) completed the questionnaire during the 11-day period it was accessible (January 19-30, 2021). The number of respondents per university ranged from 13 to 23, with completion rates between 43.2% and 82.4%. Among the respondents, 11.3% were involved in the EACT project.

Most respondents were female (69.1%) with a mean age of 48.2 (SD = 10.5). Most participants reported having a single mother tongue (i.e., the language first acquired at home or their primary language) (56.3%); 38.5% had two. In total, 69.8% identified as native Spanish speakers, 21.9% Catalan, 19.8% Galician, 8.3% English, 7.3% Basque, 6.3% German, and 5.2% French, while the remaining 11.5% had other mother tongues.

Nearly 60% of participants held tenured positions, with the most common category being tenured lecturer (36.1%), while 26.8% reported being part-time lecturers. Experience in translator training was categorised into three groups: up to 10 years, 11-20 years, and more than 20 years, reflecting early, mid, and late career stages. The sample was evenly divided: 35.1% of respondents had over 20 years' experience, 33% between 11 and 20 years, and 31.9% up to 10 years. Three quarters had received teacher training (over 40 hours in 74% of cases), 41.2% had training in teaching translation (over 40 hours in 65% of cases), 52.6% had training in assessment (up to 10 hours in 31.4% of cases), and 19.6% had training in assessment in translation teaching (over 40 hours in 52.6% of cases). In total, 42.3% had conducted research on translation didactics.

Regarding professional translation experience, 86.6% had worked as professional translators, while 48.5% were combining teaching and translating at the time of the questionnaire. Most respondents with professional experience were or had been part-time translators (66.7%), with 46.5% having up to 10 years of experience, 28.6% between 11 and 20 years, and 16.1% between 21 and 30 years. Of the 38.1% who were or had been full-time translators, 43.8% had up to 10 years of experience, 37.5% between 11 and 20 years, and 15.7% between 21 and 30 years. Most had translated for publishing houses (65.5%) while half had offered revision services (50%). Other areas respondents mentioned included legal and administrative translation (45.2%), technical translation (39.3%), audiovisual translation (31%), economic and financial translation (29.8%), officially certified translation (26.2%), and post-editing (13.1%), while almost 12% had also worked in other areas. A total of 36.1% said that they were members of a professional translator association.

In terms of academic background, 80.7% held a PhD, 69.2% of which were in the field of translation studies. Additionally, 21.6% had completed an MA, mostly related to translation studies (52.4%).

4.2. Selection of items for analysis
As stated in Section 1, the aim of this study was to explore the relationship between the background characteristics of translator trainers in Spain and their

attitudes towards and practices in assessment. The items selected refer to: (a) participants' background characteristics (age, training received, professional experience); (b) their assessment practices; and (c) their attitudes towards assessment. Items were extracted from Section 1, which focused on a general overview of participants' attitudes and practices, and Section 4, which looked at background characteristics. Tables 2 to 4 present the items identified and extracted for analysis and the descriptive results obtained. The tables are organised according to the type of information provided, e.g. attitudes, practices or background characteristics. Table 2 shows the results regarding attitudes.

Table 2: Items selected and descriptive results (attitudes)

Item	m	Mdn	SD	n	DK/DA
1.1. How important do you think assessment is in your teaching activity as a whole?	3.56	4	0.558	97	0
1.2. How important do you think assessing translations (the end-product) is within the assessment of students' level of translation competence?	3.51	4	0.615	96	1
1.3. How important do you think assessing students' translation process is within the assessment of their level of translation competence?	3.49	4	0.632	96	1
1.14. How satisfied are you with the assessment you carry out?	3.27	3	0.587	97	0

Table 3 presents the descriptive results regarding assessment practices.

Table 3: Items selected and descriptive results (practices)

Item	Yes	No	n	DK/DA
1.4. To determine grades, do you use different types of assessment tasks depending on the competence(s) being assessed?	82.1%	17.9%	95	2
1.5. To determine grades, do you use different types of assessment tasks depending on the level of the module?	64.9%	17.5%	95	2
1.6. Do you use formative assessment activities that do not count towards final grades (i.e. activities solely for identifying students' progress and weaknesses)?	87.5%	12.5%	96	1
1.7. Do you carry out blind assessment (i.e. do you try to assess work without knowing whose it is)?	52.6%	47.4%	97	0

Item	M	Mdn	SD	n	DK/DA
1.8. Do you base assessment procedures (type of tests, grading scale) on professional translation practice?	3.22	3	0.823	96	1
1.9.1. To what extent do you apply the following criteria when selecting texts for use in assessment tasks? They must be authentic (texts that professional translators might actually translate)	3.69	4	0.568	97	0
1.9.2. To what extent do you apply the following criteria when selecting texts for use in assessment tasks? They	3.74	4	0.485	97	0

must require the application of the competences being assessed					
1.9.3. To what extent do you apply the following criteria when selecting texts for use in assessment tasks? The difficulty they entail must be appropriate to the level	3.81	4	0.441	97	0

Item	Yes	Sometimes	No	n	DK/DA
1.10. When correcting translations, do you use a grading scale with a classification of possible translation errors?	69.5%	21.1%	9.5%	95	2
1.11. Do you use rubrics (i.e. marking guides that describe each aspect to be assessed and its performance levels) to assess tasks?	39.2%	29.9%	30.9%	97	0
1.12. Have you changed the assessment tasks you use in your onsite translation teaching in the last 5 years?	80.4%		19.6%	92	5
1.13. Have you changed the assessment criteria you use in your onsite translation teaching in the last 5 years?	40.2%		59.8%	92	5

Finally, Table 4 presents the results regarding background characteristics.

Table 4: Items selected and descriptive results regarding background characteristics

Item	Up to 10 years	11-20 years	+20 years	n	DK/DA
4.3. Years of translation teaching experience	31.9%	33%	35.1%	97	2

Item	Mean	Mdn	SD	min.	max.	n	DK/DA
4.5. Age	48.2	48	10.5	25	67	95	2

Item			Yes	No	n	DK/DA
4.8. Professional translation experience			86.6%	13.4%	97	0
4.9. Do you currently combine your work as a translation teacher with professional translation?			48.5%	51.5%	97	0
4.11.b. Have you received any training in translation teaching?			41.2%	58.8%	97	0
4.11.c. Have you received any training in assessment?			52.6%	47.4%	97	0

4.3. Age

This subsection examines whether age group (item 4.5) is associated with three indicators of assessment attitudes and practices. No omnibus test reached significance (See Table 5 below).

The findings indicate that age is not a determining factor in attitudes towards assessment and practices among translator trainers. Specifically, three observations emerge from the Kruskal–Wallis comparisons by age group: (1) ratings of the importance of assessment did not differ significantly across age groups; (2) the reported use of grading scales likewise did not differ by age; and (3) although rubric use varied across respondents, age-related differences were not statistically significant.

The absence of significant associations between age and assessment-related practices suggests that these practices may be widely adopted across all age groups and more strongly influenced by other variables, such as institutional

policies, professional training, or personal pedagogical preferences. For example, the frequent use of grading scales may reflect departmental norms or standardised procedures rather than differences attributable to age. Similarly, the variability in rubric use might reflect diverse institutional approaches or disciplinary conventions in translation education, rather than generational differences.

These findings align with previous research (Panadero et al., 2019) indicating that while age can influence teaching styles and openness to pedagogical innovation, core assessment beliefs and practices tend to be shaped by broader professional and institutional frameworks rather than individual characteristics.

Table 5: Results of cross-tabulating the selected items with age

1.1. How important do you think assessment is in your teaching activity as a whole?			
Response	Mean	Mdn	SD
1 (not important at all)	-		
2	54.7	53.0	6.7
3	45.6	45.0	10.86
4 (very important)	49.6	51.5	10.1
DK/DA	-		
$H = 4.127$; df = 2; $p = 0.127$; $\varepsilon^2 = 0.044$			
1.10. When correcting translations, do you use a grading scale with a classification of possible translation errors?			
Response	Mean	Mdn	SD
Yes	47.6	47.0	10.3
Sometimes	52.2	53.0	10.7
No	47.9	47.0	11.2
DK/DA	-		
$H = 1.324$; df = 2; $p = 0.516$; $\varepsilon^2 = 0.001$			
1.11. Do you use rubrics (i.e. marking guides that describe each aspect to be assessed and its performance levels) to assess tasks?			
Response	Mean	Mdn	SD
Yes	45.0	45.0	11.3
Sometimes	49.4	51.0	8.9
No	51.2	52.0	9.9
DK/DA	-		
$H = 5.534$; df = 2; $p = 0.063$; $\varepsilon^2 = 0.059$			

4.4. Attitudes towards and practices in assessment
To examine how participants' attitudes towards assessment relate to their assessment practices, we cross-tabulated items concerning their perceived importance of assessment with various assessment-related behaviours. Table 6 shows how the results of two tests achieved statistical significance, while the remaining tests did not provide statistically significant evidence of association.

One significant association was found between the perceived importance of assessment in teaching and the use of grading scales that classify possible translation errors (Fisher's exact test = 15.675; $p < 0.001$; phi = 0.584), with a moderate to strong effect size. The data indicate that participants who place greater emphasis on assessment in their teaching were more likely to report using grading scales, suggesting these tools guide students in recognising and understanding errors in their translations. This pattern is consistent with the idea that the structured nature of grading scales facilitates explicit connections between translation quality and assessment criteria, reinforcing a systematic approach to evaluation.

A second significant association was found between participants' satisfaction with their assessment methods and whether they had modified their

assessment tasks in face-to-face teaching in the past five years (Fisher's exact test = 7.958; p = 0.013; phi = 0.305), with a moderate effect size. In our data, respondents who reported having modified tasks also tended to report higher satisfaction with their methods. This should be interpreted as an association, not as evidence that task modification causes increased satisfaction.

Table 6: Results of cross-tabulating items related to the participants' attitudes towards assessment with items related to their assessment practices

Item related to attitudes	Associated item related to practices	Fisher's exact test	p	phi
1.1. How important do you think assessment is in your teaching activity as a whole?	1.10. When correcting translations, do you use a grading scale with a classification of possible translation errors?	15.675	< 0.001	0.584
1.2. How important do you think assessing translations (the end product) is within the assessment of students' level of translation competence?	1.5. To determine grades, do you use different types of assessment tasks depending on the level of the module?	0.141	1.000	0.036
1.3. How important do you think assessing students' translation process is within the assessment of their level of translation competence?	1.6. Do you use formative assessment activities that do not count towards final grades (i.e. activities solely for identifying students' progress and weaknesses)?	1.227	0.547	0.136
1.3. How important do you think assessing students' translation process is within the assessment of their level of translation competence?	1.11. Do you use rubrics (i.e. marking guides that describe each aspect to be assessed and its performance levels) to assess tasks?	5.874	0.202	0.259
1.14. How satisfied are you with the assessment you carry out?	1.12. Have you changed the assessment tasks you use in your onsite translation teaching in the last 5 years?	7.958	0.013	0.305

Although these two relationships were statistically significant, other cross-tabulations did not provide statistically significant evidence of association. The data showed no association between the perceived importance of assessing the translation product and the use of different types of assessment tasks by module level (p = 1.000; phi = 0.036), nor between the perceived importance of assessing the translation process and the use of formative assessment activities (p = 0.547; phi = 0.136). Similarly, there was no significant association between the perceived importance of assessing the translation process and the use of rubrics in assessment (p = 0.202; phi = 0.259). The lack of association in these cases may be explained by institutional norms or curricular constraints, that may limit the extent to which instructors are able to implement certain assessment methods, regardless of their perceived importance.

Beyond attitudes, we also examined the relationship between assessment task design and text selection criteria, as displayed in Table 7. The data revealed two significant associations, both suggesting a structured approach to assessment in translator training.

Table 7: Results of cross-tabulating items related to the participants' assessment practices

Item related to practices	Associated item related to practices	Fisher's exact test	p	phi
1.4. To determine grades, do you use different types of assessment tasks depending on the competence(s) being assessed?	1.9.2. To what extent do you apply the following criteria when selecting texts for use in assessment tasks? They must require the application of the competences being assessed	9.783	0.005	0.366
1.10. When correcting translations, do you use a grading scale with a classification of possible translation errors?	1.6. Do you use formative assessment activities that do not count towards final grades (i.e. activities solely for identifying students' progress and weaknesses)?	0.225	1.000	0.049
1.11. Do you use rubrics (i.e. marking guides that describe each aspect to be assessed and its performance levels) to assess tasks?	1.6. Do you use formative assessment activities that do not count towards final grades (i.e. activities solely for identifying students' progress and weaknesses)?	0.359	0.861	0.048

Item related to practices	Associated item related to practices	χ^2 with continuity correction	df	p	V
1.5. To determine grades, do you use different types of assessment tasks depending on the level of the module?	1.9.3. To what extent do you apply the following criteria when selecting texts for use in assessment tasks? The difficulty they entail must be appropriate to the level	2.945	1	0.050	0.236

The first significant association was between using different types of assessment tasks depending on the competences being assessed and selecting texts that require the application of those competences (Fisher's exact test = 9.783; $p = 0.005$; phi = 0.366). In our data, respondents who reported adapting tasks to specific competences also tended to report selecting texts aligned with those competences, a pattern consistent with alignment between assessment methods and intended learning outcomes.

A second significant association was found between using different types of assessment tasks depending on the level of a module and selecting texts whose difficulty is appropriate to that level (χ^2 with continuity correction = 2.945; df = 1; $p = 0.050$; $V = 0.236$). This near-threshold result suggests that

some instructors adjust both task type and text difficulty according to student proficiency, indicating a flexible, level-sensitive approach to evaluation.

While these associations point to a structured and adaptive approach to assessment, in some areas other comparisons did not provide statistically significant evidence of association. For example, there was no association between using grading scales to classify errors and implementing formative assessment activities ($p = 1.000$; phi = 0.049), nor between using rubrics to assess tasks and engaging in formative assessment ($p = 0.861$; phi = 0.048). These non-significant results may reflect contextual factors (e.g., institutional emphasis on summative assessment) or the possibility that formative strategies are used informally, without being tightly linked to predefined grading structures. Such hypotheses are, however, beyond the scope of the present analyses.

4.5. Knowledge of the profession

To explore whether participants report incorporating professional practice into assessment, we examined two key relationships: whether assessment methods aligned with professional translation standards, and whether text selection criteria reflected authenticity in professional practice.

We first examined whether trainers anchored assessment in professional practice. Item 1.8 asked "Do you base assessment procedures (type of tests, grading scale) on professional translation practice?" and responses were rated on a scale from 1 (not at all) to 4 (very much), with a mean of 3.22 (Mdn = 3; SD = 0.823). Responses to this item were cross-tabulated with responses to item 1.2, which asked how important participants considered assessing the translation product within the broader evaluation of translation competence (m = 3.51; Mdn = 4; SD = 0.615). The analysis did not provide statistically significant evidence of association (Fisher's exact test = 9.181; $p = 0.123$; phi = 0.332).

This result may indicate that perceiving translation product assessment as important does not necessarily reflect a purposeful alignment of assessment procedures with professional translation standards. This may be due to variability in assessment traditions across institutions, where some trainers prioritise pedagogical frameworks over industry practices.[6] While certain professional contexts do apply structured evaluation protocols, such as the European Union's Directorate-General for Translation, which uses predefined quality assurance procedures and error typologies (Moorkens et al., 2018), many other professional settings lack fixed grading criteria. For instance, freelance translators are often evaluated based on client satisfaction (Williams, 2009) or delivery speed rather than formal rubrics. Similarly, in fields such as audiovisual translation or localisation, quality is frequently judged in terms of usability and viewer experience (Tuominen, 2025), which may not correspond to academic grading standards. Additionally, trainers may emphasise other aspects of assessment, such as formative feedback and student progress, over strict professional alignment.

The second analysis examined whether text selection practices reflect professional authenticity. Item 1.9.1, which refers to authenticity in text selection ("They must be authentic (texts that professional translators might actually translate)"), had a mean response of 3.69 (Mdn = 4; SD = 0.568). This item was cross-tabulated with item 1.3, which asked about the perceived

---

[6] However, this interpretation is speculative, as the study did not include institutional-level comparisons.

importance of assessing the translation process (M = 3.49; Mdn = 4; SD = 0.632). The result was a statistically significant association (Fisher's exact test = 11.364; *p* = 0.012; phi = 0.346), indicating that participants who emphasised authentic texts also tended to assign greater importance to assessing the translation process. This aligns with contemporary competence-based models of translation assessment, which emphasise that translation quality involves not only the final product but also decision-making, problem-solving, and revision strategies used during the process (PACTE, 2020). The connection between authenticity and process-oriented assessment may stem from a pedagogical effort to simulate professional environments, where translators navigate real-world constraints, collaborate with clients, and refine their drafts iteratively.

4.6. Teaching experience

In all, 31 participants (31.9%) had up to 10 years of translation teaching experience, with 32 (33%) indicating that they had between 11 and 20 years' experience, and 34 (35.1%) more than 20 years. To examine the association between teaching experience (item 4.3) and assessment approaches, we investigated whether years of translation teaching were associated with participants' attitudes and practices in assessment. The results (Table 8) revealed one significant association related to teaching experience, alongside several non-significant findings.

Table 8: Results of cross tabulations involving the participants' translation teaching experience (item 4.3)

Years of translation teaching experience cross-tabulated with...	Fisher's exact test	*p*		phi
1.1. How important do you think assessment is in your teaching activity as a whole?	6.262	0.124		0.248
1.14. How satisfied are you with the assessment you carry out?	6.427	0.153		0.266
Years of translation teaching experience cross-tabulated with...	χ^2	df	*p*	*V*
1.12. Have you changed the assessment tasks you use in your onsite translation teaching in the last 5 years?	7.599	2	0.020	0.287
1.13. Have you changed the assessment criteria you use in your onsite translation teaching in the last 5 years?	3.924	2	0.141	0.207

The analysis of years of translation teaching experience revealed a significant association with having modified assessment tasks in the preceding five years (χ^2 = 7.599; df = 2; *p* = 0.020; *V* = 0.287). This implies that instructors with longer teaching experience more often reported having adjusted their assessment tasks over time, possibly due to pedagogical refinement, curricular changes, or increased confidence in their teaching strategies. Less experienced instructors may still be establishing their assessment framework, leading to fewer modifications. This finding emphasises that assessment strategies may evolve through experience, reflection, and exposure to diverse teaching contexts.

However, other comparisons did not provide statistically significant evidence of association: teaching experience and satisfaction with assessment (*p* = 0.153; phi = 0.266) and teaching experience and the perceived importance of assessment in teaching (*p* = 0.124; phi = 0.248). Within this sample, these results suggest that while experience co-occurs with reported task modification, it is not clearly related to satisfaction or perceived importance. Such patterns

may reflect institutional or programme-level factors that shape perceptions relatively uniformly across career stages, or the influence of training, departmental policies, or professional networks beyond years of teaching alone.

4.7. Professional translation experience

Of the 97 participants, 86.6% had experience as professional translators (item 4.8). To explore whether professional translation experience was associated with assessment approaches, we cross-tabulated experience as a translator with three key assessment-related variables.

First, we examined whether professional translation experience was associated with basing assessment procedures on professional practice. The result was not significant (Fisher's exact test = 5.443; $p = 0.101$; phi = 0.216), indicating that having worked as a professional translator was clearly not associated with a greater emphasis on industry-aligned assessment methods. This may suggest that translation assessment in academic settings follows pedagogical conventions rather than direct professional replication.

By contrast, we found a significant association between professional translation experience and currently combining teaching with professional translation (Fisher's exact test = 16.678; $p < 0.001$; phi = 0.416). Nearly half of the participants (48.5%) reported combining both professions, highlighting a strong link between continued engagement in professional practice and an active teaching role. This may reveal that ongoing professional involvement can enrich teaching by exposing trainers to evolving industry standards, technological developments, and real-world challenges in translation.

The outcomes show that having professional translation experience does not necessarily shape the way assessment is designed or lead to greater involvement in professional organisations. Similarly, teaching experience appears to be associated with adjusting assessment tasks over time, but it was not significantly associated with satisfaction with assessment practices or the perceived importance of assessment. These data highlight a potential disconnect between professional and academic assessment approaches, showing that translation pedagogy may be guided more by institutional and educational frameworks than by direct professional practice.

4.8. Training

Of the 97 respondents, 41.2% reported that they had received training in translation teaching, while 52.6% had received training in assessment. To examine how training in translation teaching and assessment influences assessment-related attitudes and practices, we cross-tabulated these variables with relevant assessment behaviours. Table 9 presents the results, revealing two significant associations related to training in translation teaching and no significant associations related to training in assessment.

Among the cross-tabulations involving training in translation teaching (item 4.11.b), two tests reached statistical significance. Participants who reported having had such training also reported rubric use more frequently ($\chi^2 = 6.159$; df = 2; $p = 0.041$; $V = 0.252$). They also reported having modified their assessment tasks in the past five years more often ($\chi^2 = 7.885$; df = 1; $p = 0.005$; $V = 0.293$). These results provide statistical evidence of non-random associations between training in translation teaching and both rubric use and recent task modification. These findings should be interpreted as associations rather than causal effects, even though the pattern is consistent with a more structured and reflective approach to assessment.

However, no statistically significant associations were observed between training in translation teaching and the other assessment-related behaviours

examined, such as the use of grading scales, formative assessment, or selecting texts by level of difficulty. This suggests that, in our data, training in translation teaching co-occurred with rubric use and task modification but was not clearly associated with the remaining behaviours. Several hypotheses might account for this. Firstly, many training offerings may emphasise course design and didactic methods, with assessment receiving less attention. Secondly, institutional policies and local norms may exert a stronger influence on choices such as grading scales or formative assessment than individual training. These explanations are speculative; evaluating them would require targeted studies of training content and institutional context.

Table 9: Results of cross-tabulations involving assessment and participants' training in translation teaching

Training in translation teaching (item 4.11.b)			
Item	Fisher's exact test	p	phi
1.1. How important do you think assessment is in your teaching activity as a whole?	3.628	0.184	0.192
1.8. Do you base assessment procedures (type of tests, grading scale) on professional translation practice?	0.967	0.885	0.102
1.9.3. To what extent do you apply the following criteria when selecting texts for use in assessment tasks? The difficulty they entail must be appropriate to the level	2.086	0.417	0.137
1.10. When correcting translations, do you use a grading scale with a classification of possible translation errors?	1.679	0.421	0.138
1.14. How satisfied are you with the assessment you carry out?	2.288	0.355	0.156

Item	χ^2 with continuity correction	df	p	V
1.6. Do you use formative assessment activities that do not count towards final grades (i.e. activities solely for identifying students' progress and weaknesses)?	0.392	1	0.531	0.064
1.11. Do you use rubrics (i.e. marking guides that describe each aspect to be assessed and its performance levels) to assess tasks?	6.159	2	0.041	0.252
1.12. Have you changed the assessment tasks you use in your onsite translation teaching in the last 5 years?	7.885	1	0.005	0.293
1.13. Have you changed the assessment criteria you use in your onsite translation teaching in the last 5 years?	0.593	1	0.441	0.103

Training in assessment (item 4.11.c)				
Item	χ^2 with continuity correction	df	p	V
1.7. Do you carry out blind assessment (i.e. do you try to assess work without knowing whose it is)?	0.233	1	0.629	0.049

In contrast, training in assessment (item 4.11.c) did not provide statistically significant evidence of association with blind assessment practices (χ^2 with continuity correction = 0.233; df = 1; p = 0.629; V = 0.049). This result should not be read as evidence of no effect, but rather as an absence of detectable association in our data. Several hypotheses could account for this pattern. It is possible that assessment training in some programmes may emphasise conceptual frameworks such as those focusing on validity and reliability over hands-on implementation, while offering limited guidance on procedures such as blind marking. Alternatively, institutional norms, department policies, or individual teaching philosophies may play a larger role in the adoption of blind assessment than training alone. These interpretations are tentative, as no institutional comparisons were conducted in this study, and would in fact require dedicated research on training content and local assessment policies.

The two significant associations related to training in translation teaching offer a tentative insight into how professional development may relate to assessment methods. One interpretation is that these associations reflect the content and emphasis of training available to translator trainers. Rubrics are a structured, widely promoted tool and are typically included as a core component of translation teacher training programmes. By contrast, formative assessment strategies or text-selection practices may be more individualised and less systematically covered in training curricula. These explanations are hypotheses, not causal claims, and would require targeted evidence drawn from training syllabi to be substantiated.

Another hypothesis is that training in translation teaching may coincide with a more reflective stance toward assessment, with instructors periodically refining and adjusting their evaluation methods. Two non-exclusive mechanisms could underlie this pattern: firstly, self-selection, whereby educators already engaged in continuous professional development are more likely to seek training and to experiment with new techniques (including task modification); and secondly, exposure effects, whereby participating in training prompts a review of existing practices and surfaces opportunities for incremental change. These explanations are speculative, since the present data cannot discriminate between them or establish causality.

In contrast, no statistically significant associations were observed between assessment training and the behaviours examined. This should not be interpreted as evidence of no effect, but rather as an absence of detectable association in this sample, also given the scope of the study. The pattern may indicate a gap between the conceptual emphasis of some assessment training and its practical implementation in translator education, and this possibility warrants targeted investigation in future studies.

5. Discussion and concluding remarks

This study set out to explore how translator trainers' background characteristics related to their attitudes towards and practices in assessment within five Spanish universities. While the field of translation assessment has traditionally been shaped by subjective, prescriptive approaches (Künzli, 2001; Reiss, 1971), recent research has called for a shift towards more empirical and pedagogically grounded evaluation models (Gardy, 2016; Huertas-Barros and Vine, 2018). This paper responds to that call by offering insights on how personal and professional variables are related to the ways in which assessment is understood and implemented in translator education.

Three major results emerged from the data. First, age does not appear to be associated with assessment practices or attitudes. This aligns with findings in broader educational research in Spain suggesting that institutional factors and training are more strongly associated with assessment strategies than demographic traits such as age (Panadero et al., 2019). The widespread use of grading scales and rubrics across age groups may point to shared departmental norms and possibly curricular expectations, rather than generational differences.

Second, we observed that professional translation experience (particularly when ongoing) was associated with the integration of more authentic materials and process-oriented evaluation methods. This aligns with findings by Domínguez Araújo (2015) and is further supported by Kiraly (2000, 2012), who advocates for project-based learning and the use of authentic tasks to foster learner autonomy and professional competence. Similar perspectives are echoed by Biel (2011) and Korda (2023), who emphasize the pedagogical value of professional realism and authentic projects in translator education. However, not all experienced professionals translated this engagement into assessment practices aligned with industry norms, which may suggest a disconnect between professional and pedagogical logics. This result reinforces conclusions from earlier studies (e.g., Gardy, 2016) highlighting the persistence of academic traditions over real-world practices in the conduct of evaluation.

Third, training in translation pedagogy (though not specifically in assessment) appears to be associated with more reflective and adaptive assessment behaviours, such as the use of rubrics and the periodic modification of tasks. In contrast, no statistically significant association was found for formal assessment training, echoing previous concerns about the limitations of training that remains theoretical rather than practical (Huertas-Barros and Vine, 2019a). Together, these two findings underline the need to rethink how assessment is covered in translator trainer education, emphasising actionable techniques and real-world application over abstract models.

Considered as a whole, the results point to a high degree of variability in assessment practices, with associations observed with professional involvement and teaching experience. Other factors not examined here – such as institutional policies, teaching loads, language combinations or the role of programme-level curriculum design – are also likely to contribute to this variability. Individual instructors appear to draw from their experience, intuition, and available training to inform their practices. This echoes earlier critiques of the field's lack of standardization and scientific grounding of assessment practices, with Bowker (2000) noting the absence of consistent criteria and Martínez and Hurtado Albir (2001) calling for empirically validated models. These concerns remain relevant today and affirm recent calls for improved training and greater collaboration between trainers and students in defining assessment criteria (Huertas-Barros and Vine, 2019b).

From a practical standpoint, our findings support the development of targeted training programmes for translator trainers, particularly those focused on bridging the gap between theory and practice. Such programmes should:

1. Include hands-on modules on assessment design, rubric creation, and process-oriented feedback (*cf.* Way, 2021).
2. Facilitate collaboration among trainers to reduce variability and foster shared best practices (*cf.* Sánchez Nieto et al., 2009).
3. Encourage dialogue with students to increase transparency, fairness, and learner autonomy in assessment (cf. Domínguez Araújo, 2015; Huertas-

Barros and Vine, 2019b), including formative activities that help students understand and use the rubrics.

This study also contributes to the ongoing EACT project, offering a detailed snapshot of the current state of assessment in five member institutions. While the results are informative, certain limitations must be acknowledged. The sample, though robust, is limited to five universities out of nearly thirty involved in translation education in Spain. The institutions involved in the study may differ from others in their curricular structures, teaching philosophies, or institutional support for assessment innovation, potentially affecting the generalizability of our findings. Moreover, the use of an online questionnaire, while efficient, may have introduced non-response bias. Trainers who responded might have had a stronger interest in assessment or more positive perceptions of their own practices. This possibility warrants further investigation through triangulated data collection methods, such as interviews, document analysis, or classroom observation. A small proportion of respondents (11.3%) were also members of the EACT project. Their inclusion reflects their dual role as university teachers with an interest in assessment, and their responses were collected and analysed using the same procedures applied to all participants. While their involvement may suggest a heightened awareness of assessment-related issues, we do not consider this a conflict of interest but acknowledge it as a potential source of positive bias that warrants transparency.

A further limitation concerns the way attitudes towards assessment were operationalised. In this study, attitudes were measured using a small number of global items capturing perceived importance and overall satisfaction. These indicators offer only a broadbrush approximation of trainers' underlying belief systems, which may include their views on fairness, transparency, workload, or alignment with learning outcomes. Some of the null or weak associations observed may therefore reflect the limited depth of attitudinal measurement rather than the absence of substantive relationships. Future research should complement the more detailed practice-related items with richer attitudinal scales or qualitative approaches that capture the complexity of trainers' beliefs.

Another limitation relates to statistical power and multiplicity. With N = 97 and several subgroup and cross-tabulation analyses, the study is adequately powered to detect medium-to-large associations but underpowered to detect small effects. In addition, we conducted 26 tests at $\alpha = 0.05$, meaning that one or two statistically significant results would be expected by chance alone, even if all null hypotheses were true. The predominantly non-significant tests should therefore not be taken as definitive evidence that background factors are unrelated to assessment attitudes and practices. Rather, they may indicate that any existing relationships were unlikely to be strong in this sample, while smaller associations may have gone undetected. The small number of significant associations that did emerge should be interpreted as provisional and hypothesis-generating rather than as conclusive evidence of strong, systematic relationships. Future confirmatory studies with larger samples, pre-registered hypotheses and appropriate procedures for controlling familywise or false-discovery error rates (FWDRs and FDRs) will be needed to establish which of these associations are robust.

Future research should continue to explore how institutional, linguistic, and curricular contexts mediate assessment practices. Follow-up studies could also investigate the impact of newly implemented train-the-trainer programmes, considering differences in their pedagogical focus, or assess the long-term effects of collaborative assessment design on student outcomes and trainer satisfaction.

In conclusion, this study helps fill a gap in the empirical literature on translation assessment by showing that, while professional experience and pedagogical training are associated with some aspects of assessment behaviour, there remains significant variability. Despite its limitations, the study highlights the value of systematically examining translator trainers' practices and opens the door to informed improvements in training, policy, and research. More consistent training, clearer institutional guidance, and collaborative frameworks could help move translation assessment closer to a balanced model that reflects both academic standards and professional realities.

References

Adelson, J.L. and McCoach, D.B. (2010). Measuring the mathematical attitudes of elementary students: The effects of a 4-point or 5-point Likert-type scale. *Educational and Psychological Measurement*, *70*(5), 796-807. https://psycnet.apa.org/doi/10.1177/0013164410366694

Arango-Keeth, F. and Koby, G.S. (2003). Assessing assessment: Translator training evaluation and the needs of industry quality assessment. In B. J. Baer and G.S. Koby (Eds) *Beyond the ivory tower: Rethinking translation pedagogy* (pp. 117–134). John Benjamins. https://doi.org/10.1075/ata.xii

Armstrong R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502-508. https://doi.org/10.1111/opo.12131

Biel, Ł. (2011). Professional realism in the legal translation classroom: Translation competence and translator competence. *Meta: Translators' Journal*, *56*(1), 162–178. https://doi.org/10.7202/1003515ar

Bowker, L. (2000). A corpus-based approach to evaluating student translations. *The Translator, 6*(2), 183-210. https://doi.org/10.1080/13556509.2000.10799065

Brunette, L. (2000). Towards a terminology for translation quality assessment. A comparison of TQA practices. *The Translator*, *6*(2), 169-182. https://doi.org/10.1080/13556509.2000.10799064

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., and Walter K. (2020). Purposive sampling: Complex or simple? Research case examples. *Journal of Research in Nursing*, *25*(8), 652-661. https://doi.org/10.1177/1744987120927206

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.

Conde Ruano, J. T. (2009). *Proceso y resultado de la evaluación de traducciones.* [Doctoral dissertation, Universidad de Granada]. Universidad de Granada. https://digibug.ugr.es/handle/10481/2309

Conde Ruano, J. T. (2008). La evaluación de traducciones, a examen. In Fernández Sánchez, M. M. and Muñoz Martín, R., *Aproximaciones cognitivas al estudio de la traducción y la interpretación*. Comares.

Delisle, J. (2001). L'évaluation des traductions par l'historien. *Meta: Translators' Journal*, *46*(2), 209-226.

Domínguez Araújo, L. (2015). *La evaluación para el aprendizaje de la interpretación de conferencias.* [Doctoral dissertation, Universitat Autònoma de Barcelona]. Universitat Autònoma de Barcelona. https://www.tdx.cat/handle/10803/317965

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. SAGE Publications.

Gardy, P. (2016). L'évaluation en didactique de la traduction : Un état des lieux. *The Journal of Specialised Translation*, 26, 20-49. https://doi.org/10.26034/cm.jostrans. 2016.276

Garland, R. (1991). The mid-point on a rating scale: Is it desirable? *Marketing Bulletin*, 2, 66-70.

Gay Punzano, L., and Hurtado Albir, A. (2025). Assessment procedures in translation degree programmes in Spain. Results of the EACT project. *Meta*, *69* (3), 732-757. https://doi.org/10.7202/1121667ar

Haro-Soler, M. M. (2018). *Las creencias de autoeficacia del estudiantado de traducción: una radiografía de su desarrollo.* [Doctoral dissertation, Universidad de Granada]. Universidad de Granada. https://digibug.ugr.es/handle/10481/53590

House, J. (1977). *A model for translation quality assessment*. Günter Narr.

House, J. (1997). *Translation quality assessment: A model revisited*. Günter Narr.

Hovy, E., King, M. and Popescu-Belis, A. (2002). Principles of Context-Based Machine Translation Evaluation. *Machine Translation, 17*, 43-75.

Huertas-Barros, E. and Domínguez Araújo, L. (2024). Translation tutors' assessment perceptions and practices in Spain: Results from the EACT project. *Hermeneus, 26*, 253–282. https://doi.org/10.24197/her.26.2024.253-282

Huertas-Barros, E. and Vine, J. (2016). Translator trainers' perceptions of assessment: An empirical study. In M. Thelen, G.-W. van Egdom, D. Verbeeck, Ł. Bogucki and B. Lewandowska-Tomaszczyk (Eds.), *Translation and meaning: New Series, 41* (pp. 29-39). Peter Lang. https://doi.org/10.3726/978-3-653-05892-5

Huertas-Barros, E. and Vine, J. (2018). Current trends on MA translation courses in the UK: Changing assessment practices on core translation modules. *The Interpreter and Translator Trainer*, *12*(1), 5-24. https://doi.org/10.1080/1750399X.2017.1400365

Huertas-Barros, E. and Vine, J. (2019a). Constructing standards in communities: Tutors' and students' perceptions of assessment practices on an MA translation course. In E. Huertas-Barros, S. Vandepitte and E. Iglesias-Fernández (Eds.), *Quality assurance and assessment practices in translation and interpreting* (pp. 245-269). IGI Global. https://doi.org/10.4018/978-1-5225-5225-3

Huertas-Barros, E., and Vine, J. (2019b). Training the trainers in embedding assessment literacy into module design: a case study of a collaborative transcreation project. *The Interpreter and Translator Trainer*, *13*(3), 271–291. https://doi.org/10.1080/1750399X.2019.1658958

Hurtado Albir, A., and Rodríguez-Inés, P. (2022). Research perspectives. *MonTi Monografías De Traducción E Interpretación*, 204–209. https://doi.org/10.6035/MonTI.2022.ne7.06

Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*, *42*(2), 152-155. https://doi.org/10.5395/rde.2017.42.2.152

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

Kiraly, D. (2012). Project-based learning: A case for situated translation. In M. Olohan (Ed.), *The translator as mediator of cultures* (pp. 83-117). John Benjamins.

Korda, D. (2023). Translation training: The use of authentic projects. *CTTL Journal*, *1*(10) 302-337. https://doi.org/10.51287/cttl202310

Künzli, A. (2001). Experts versus novices: L'utilisation de sources d'information pendant le processus de traduction. *Meta: Translators' Journal*, *46*(3), 507-523. https://doi.org/10.7202/003363ar

Li, D. (2006). Translator Training: What translation students have to say. *Meta: Translators' Journal*, *47*(4), 513-531. https://doi.org/10.7202/008034ar

Maier, M. and Lakens, D. (2022). Justify your alpha: A primer on two practical approaches. *Advances in Methods and Practices in Psychological Science*, *5*(2). https://doi.org/10.1177/25152459221080396

Martínez Melis, N. (1996). Évaluation et traduction. In *Actes I Congrés International sobre Traducció* (pp. 485-492). Servei de publicacions de la Universitat Autònoma de Barcelona.

Martínez, N. and Hurtado, A. (2001). Assessment in translation studies: Research Needs. *Meta: Translators' journal, 46*(2), 272-287. https://doi.org/10.7202/003624ar

McAlester, G. (2000). The evaluation of translation into a foreign language. In C. Schäffner and B. Adab (Eds.), *Developing translation competence* (pp. 229-242). John Benjamins. https://doi.org/10.1075/btl.38

Moorkens, J., Castilho, S., Gaspari, F. and Doherty, S. (Eds.) (2018). *Translation quality assessment: From principles to practice*. Springer.

Ortega Arjonilla, E. and Fuentes Morán, M. T. (2017). Esbozo de una cartografía de los estudios de traducción e interpretación en la universidad española (versión 2.0 - datos de 2017). In E. Ortega Arjonilla, A. B. Martínez López and F. García Luque (Eds.), *Cartografía de la traducción, la interpretación y las industrias de la lengua. Mundo profesional y formación académica: interrogantes y desafíos* (pp. 25-69). Comares.

PACTE (2018). Competence levels in translation: Working towards a European framework. *The Interpreter and Translator Trainer*, *12*(2): 111-131. https://doi.org/10.1080/1750399X.2018.1466093

PACTE (2020). Translation competence acquisition. Design and results of the PACTE group's experimental research. *The Interpreter and Translator Trainer*, *14*(2), 95-233. https://doi.org/10.1080/1750399X.2020.1732601

Panadero, E., Fraile, J., Fernández-Ruiz, J., Castilla-Estévez, D. and Ruiz-Díaz, M. (2019). Spanish university assessment practices. *Assessment and Evaluation in Higher Education*, *44*(3), 379-397. https://doi.org/10.1080/02602938.2018.1512553

Pardo, A. and Ruiz, M. Á. (2012). *Análisis de datos en ciencias sociales y de la salud III*. Editorial Síntesis.

Pardo, A. and San Martín, R. (2010). *Análisis de datos en ciencias sociales y de la salud II*. Editorial Síntesis.

Pardo, A., Ruiz, M. Á. and San Martín, R. (2009). *Análisis de datos en ciencias sociales y de la salud I*. Editorial Síntesis.

Pavani, S. (2016). *La evaluación sumativa en la didáctica de la traducción. Un estudio empírico en la enseñanza de la traducción español-italiano.* [Doctoral dissertation, Universitat Autònoma de Barcelona]. http://hdl.handle.net/10803/400016

Raaijmakers, Q.A.W., van Hoof, A., Hart, H., Verbogt, T.F.M.A., Vollebergh, W.A.M. (2000). Adolescents' midpoint responses on Likert-type scale items: Neutral or missing values? *International Journal of Public Opinion Research*, *12* (2), 209-217. https://doi.org/10.1093/ijpor/12.2.209

Reiss, K. (1971). *Möglichkeiten und Grenzen der Übersetzungskritik*. Max Hueber.

Rodríguez Verbo, B. (2004). Towards a methodology for assessing literary translation. *Traduçao e Comunicaçao*, *13*, 35-54.

Sánchez Nieto, M. T., Fraile Vicente, E., Frechoso Remiro, J. C., Bachiller Martínez, J. M., Carrera Fernández, J., and Zarandona Fernández, J. M. (2009). DoCENTIC: Experiencias de colaboración interdisciplinar en Traducción e Interpretación. *@tic. revista d'innovació educativa*, *3*, 44–51. https://ojs.uv.es/index.php/attic/article/download/111/1131

Tsang, K.K. (2012). The use of midpoint on Likert Scale: The implications for educational research. *Hong Kong teachers' centre journal, 11*, 121-130.

Tuominen, T. (2025). User-centred subtitling: Usability as an assessment and process framework for interlingual subtitles. *Journal of Audiovisual Translation*, *8*(1), 1–25. https://www.jatjournal.org/index.php/jat/article/view/379

Waddington C. (2001). Different Methods of Evaluating Student Translations: The Question of Validity. *Meta: Translators' Journal*, *46*(2), 311–325. https://doi.org/10.7202/004583ar

Way, C. (2021). Developing manageable individualised formative assessment of translator trainees through rubrics. *Research in Language*, *19*(2), 135–154. https://czasopisma.uni.lodz.pl/research/article/download/12322/11891

Williams, M. (2009). Translation quality assessment. *Mutatis Mutandis*, *2*(1), 3–23.