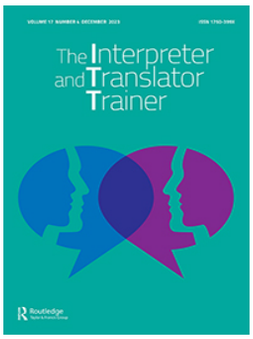

The Interpreter and Translator Trainer

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ritt20

The impact of collaborative processes on target text quality in translator training

Gys-Walt Van Egdom, Iris Schrijver, Heidi Verplaetse & Winibert Segers

To cite this article: Gys-Walt Van Egdom, Iris Schrijver, Heidi Verplaetse & Winibert Segers (24 Feb 2024): The impact of collaborative processes on target text quality in translator training, The Interpreter and Translator Trainer, DOI: 10.1080/1750399X.2024.2320512

To link to this article: https://doi.org/10.1080/1750399X.2024.2320512

Published online: 24 Feb 2024.

THE INTERPRETER AND TRANSLATOR TRAINER
https://doi.org/10.1080/1750399X.2024.2320512

Routledge
Taylor & Francis Group

The impact of collaborative processes on target text quality in translator training

Gys-Walt Van Egdom[a], Iris Schrijver[b], Heidi Verplaetse[c] and Winibert Segers[c]

[a]Translation, Intercultural Communication and Education, Utrecht University, Utrecht, The Netherlands; [b]Dept Translation and Interpreting, University of Antwerp, Antwerp, Belgium; [c]Translation and Interpreting Studies, KU Leuven, Belgium

ABSTRACT

This article explores the impact of collaboration on target text quality in translator training. By comparing team translations with those by individual peers, and analysing the highest and lowest scoring teams, the authors aimed to understand the impact of collaboration on quality. The comparison indicates that translations in a skills lab setting marginally outperform individual translations, but that carelessness is more likely to be observed in the intermediate individual target texts produced in a collaborative setting.

ARTICLE HISTORY
Received 27 January 2023
Accepted 15 February 2024

KEYWORDS
Translation evaluation; collaborative learning; PIE method; translation competence

Introduction

Although translation quality is an elusive concept, the notion has always been ubiquitous in the language industry, in translator training and in Translation Studies (Moorkens et al. 2018). In the past, textual factors were the focus of research on quality (Hönig and Kussmaul 1982). In recent years, the notion has expanded significantly, reaching far beyond textual aspects. Attention is now paid to translation process quality, translation service provision, general ergonomics, quality in relation to price, etc (see Koby et al. 2014; Van Egdom et al. 2018).

In translator training, the debate on quality has centred increasingly on competence acquisition (e.g. PACTE 2020). Translation scholars have intensified research on competence-based learning in response to calls for professionalisation, a major motivation behind initiatives to improve graduate employability (see Huertas Barros and Vine 2016). Research in this domain has sought to define the knowledge, skills and attitudes that are required in the language industry, to enhance the curriculum, clarify learning outcomes, and ensure graduates deliver high-quality services. Competence-based learning has given rise to the publication of competence models (e.g. PACTE 2003; Göpferich 2009; EMT Board 2022). These frameworks form the backbone of many translation programmes throughout Europe.

CONTACT Gys-Walt Van Egdom g.m.w.vanegdom@uu.nl

The road towards high-quality competence-based learning is fraught with pitfalls: not only does it seem impossible to lodge relevant competences of the diversified translation industry into an MA programme, at a more fundamental level, but it is also observed that competence models fail to operationalise subcompetences in a well-considered way (see Galán-Mañas and Hurtado Albir 2015). Furthermore, it remains questionable whether adding modules to improve specific skills such as technology proficiency, interpersonal abilities, and service provision could negatively affect core translation competence.

In this contribution, quality and how to achieve it in the context of translator training will occupy a central position. One possible way trainees might be able to reach higher quality is by working collaboratively on translation tasks (Thelen 2016). This contribution seeks to answer the following research question: *What are the effects of a collaborative approach to translator training on target text (TT) quality*? To this end, a comparison has been made between the TT quality of translations produced by trainee teams and that of output produced by peers who worked individually. At the outset, the study hypothesised that 'skills labs' would enhance TT quality through co-creation (see Thelen 2016). As this study is firmly embedded in research on competence acquisition, an attempt was made to identify behavioural patterns indicative of competent collaborative behaviour.

Theoretical framework

Translation evaluation

Despite the wealth of publications aimed at broadening the perspective on translation quality, monographs on quality (assessment) indicate that finding reliable, valid, and objective methods to measure TT quality remains an active pursuit (see Van Egdom et al. 2018). The term 'reliability', considered a necessary precondition of validity and objectivity, can refer to both interrater and intrarater reliability. The latter can be defined as the ability of a rater to produce similar outcomes under the same conditions. The former refers to consistent judgement across evaluators. Recent efforts have developed methods to enhance the reliability of TT testing, addressing previous issues with test design. Validity is ensured simply when a test measures what it purports to measure. The issue of validity has been proven much thornier than that of reliability. As intimated by Adab (2000), the objective of a translation assignment is to measure the complex behavioural construct of translator competence, but assessment is generally based on a textual product. While it is true that the TT is the outcome of trainee behaviour, the tangible actions that lead to a (in)correct solution in it have become irretraceable: identical solutions across students' versions can be indicative of different forms of (in)competence. There seems to be a broad consensus that objectivity is unattainable. Still, attempts have been made to 'neutralise' subjectivity, by excluding personal preferences, and guaranteeing repeatability and reproducibility. A type of assessment method that holds great promise is corpus-based assessment, where the correctness of solutions depends on their occurrence in a corpus (see Bowker 2001, 2003).

Among the methods that have been proposed, three basic types of assessment can be distinguished. The holistic approach (cf. Tijtgat and Segers 2019) bases judgements on the overall impression of the text as a whole. Although some scholars

have tried to craft cases in favour of a holistic approach, it is often discredited as subjective (Garant 2009; Waddington 2004). Another approach is the analytical approach, which 'is a way to assess the quality of a translation by looking at segments of the text [...] and awarding or deducting points for the overall score [...] based on whether each text unit meets certain criteria' (Mariana, Cox, and Melby 2015, 155; Verplaetse and Lambrechts 2022). Despite their promise of objectivity, analytical methods also face criticism. Analytical testing is often time-consuming and criteria tend to overlap and leave room for interpretation (Saldanha and O'Brien 2014, 101–102). Attempts to overcome the shortcomings of holistic and analytical approaches have resulted in the rubricated approach (see Colina 2009). The rubrics are comparable to error categories in analytical testing, but instead of awarding or deducting points for text units a score is produced per rubric on the basis of the evaluator's impression/knowledge (see also Hurtado Albir and Taylor 2015, 274–275). Rubric scores are aggregated to obtain a total score. This approach appears to combine the best of the holistic and analytical approaches, but some rubrics seem to leave too much interpretative room.

Another critical distinction that can be made in TT testing touches upon the notion of validity. In TT testing, a critical distinction in validity is made between criterion-based and norm-referenced test designs (cf. Tijtgat and Segers 2019). Assessments, ideally reflecting learning outcomes, often rely on specific criteria, primarily gauging translation performance by error count. However, this type of assessment overlooks broader learning processes, as shared errors might be indicative of instructional gaps. Norm-referenced testing, in contrast, compares individual performance against group norms, deriving criteria posterior to translation evaluation. In norm-referenced TT testing, individual performance is compared to that of peers. These tests use statistical docimology to ensure appropriate test difficulty and discern student proficiency levels, offering insights into performance in a learning community. However, the construct validity is uncertain (Van Egdom et al. 2018), since norm-referenced testing obfuscates the relationship between a test construct and what the test purports to measure, the achievement of specific learning outcomes. In recent years, translation scholars sought ways to reap the benefits of criterion-based and norm-referenced testing, aiming to bring (outcome-oriented) learning and assessment in line, resulting in the development of methods like the Preselected Items Evaluation (PIE).

Collaboration

The concept of 'collaboration', which has been studied in various educational contexts (Bruffee 1973; Dillenbourg 1999), is viewed as a key driver of study success: collaborative approaches are student-centred and instil a sense of ownership and empowerment (see Konttinen 2022; Van Egdom and Segers 2019). In translator training, Donald Kiraly was the first to highlight this: he argued in favour 'authentic experiential learning', i.e. re-creating actual professional translation situations in classrooms, where students act as service providers (the translator, project manager, reviser, etc.), and texts are commissioned as real translations (2000). Hanna Risku has also been an important advocate of a collaborative approach; her didactic concept of 'situated learning' revolves around the principle of co-construction or co-creation: she believes that 'cognitive processes linked

to the acquisition of knowledge do not occur solely in the brain but are instead shared with the individual's interaction with others and the environment' (Risku 2016, 15).

The pedagogical rationale for implementing collaborative learning modules chimes in with initiatives for professionalisation of translation curricula. The main idea behind these initiatives was that graduate employability would be enhanced if learning outcomes were to reflect the requirements for the language industry. In the EMT Competence Framework (2022), a gold standard for translator training, personal and interpersonal skills, as well as translation service provision have gradually been given more weight. Collaborative learning methods seem to offer the best opportunity to develop those subcompetences.

Since the early 1990s, the 'skills lab' has been a notable example of collaborative learning (see Gouadec 2007; INSTB 2017; Thelen 2006). It is considered ideal for testing and assessing translator trainees and services, in accordance with professional standards (ISO 17,100 2015; Thelen 2015, p. 96). In this simulated environment, students run a translation agency, mirroring professional practice, and gain practical knowledge and skills through role-playing and project management (Buysschaert et al. 2018). In its most elaborate form, a skills lab can be considered a 'holistic' method in translator training: every aspect of professional life, from writing a business plan through processing a translation project to carrying out project post-mortems, is covered in a functional manner.

Despite being praised as a leading-edge method for acquiring translation skills and claimed to ensure objective testing (Thelen 2015), the effectiveness of skills labs in delivering high-quality translations remains unproven. But there is reason to assume that it does: in its workflow outline, ISO 17,100 stipulates that collaboration is instrumental in generating high-quality translations (2015), as compliance with the 'more-eyes principle' yields better results. Still, to date no conclusive evidence is available on the question whether TT quality benefits from collaboration in translator training.

Method

Participants and materials

The data for the current study consists of two learner translation corpora, collected in two subsequent phases. In the first phase, data were collected from individual student translations, viz. individual TTs (ITTs), produced by trainee translators from two institutions. In total, 61 students from these institutions participated in this project: 54 KU Leuven students (45 BA3 students in Applied Linguistics and 9 students of the Master in Translation), 7 Zuyd University students (BA1 students in Translation). This resulted in a learner translation corpus with mixed diversification. All students were asked to produce a Dutch translation of an excerpt of 'Cancer patients with insurance still face huge costs' (332 words, retrieved from Futurity.org).[1] This yielded a learner translation corpus of 22,284 words. The subcorpora per training level represent 11.3%, 72.7% and 16% of the full first part of the corpus for 1st, 3rd and 4th year students respectively.

In a second phase of the study, data from students' collaborative translations were collected, viz. collaborative TTs (or CTTs). These CTTs of the same ST as used in the first part of the study were produced in a skills lab, with 39 students of the

MA Translation at Universiteit Utrecht.[2] All students had two weeks' experience working in a skills lab. The students in the skills lab were assigned distinct roles, contributing to a collaborative effort broadly in line with the ISO-17100 standard (ISO 2015):

(1) Project Manager-Proofreader: Responsible for overseeing the workflow and processing of the assignment. The project manager also assumed the role of proofreader, which included making minor corrections.
(2) Translator: Responsible for generating the TT1; the translator was advised to conduct a self-revision ('check') of their work.
(3) Reviser: Tasked with examining the TT against the source language content. They focused on identifying errors and other issues in the TT1. Their role also involved assessing suitability of the content for its intended purpose.[3]

Although the students progressed through each step of the collaborative process individually, there was oral communication between team members at the beginning and end of each step. No recordings were made of these discussions.

The data collected during this phase consists of 13 draft translations (TT[1]s), 13 revised translations (TT[2]s) and 13 CTTs of the same ST. This yielded a 'collaborative' corpus of 14,021 words. In addition, students' work processes during the different stages were logged, using the keystroke logging software Inputlog (Leijten and Van Waes 2013; Leijten et al. 2014). Eleven logfiles of TT[1]s, 11 log files of TT[2]s, and 11 log files CTTs were collected. In two cases, log files of team processes were not submitted by the students. Process data were collected to gain insight in general process behaviour, more specifically during drafting and proofreading. In order to gain insight into the nature of the changes which were implemented during the students' revision processes as part of the collaborative translations, the LISA QA form was used, a quality assurance form developed by the Localization Industry Standards Organization.[4] Twelve QA forms were collected; one form was not submitted.

Both for 'individual' translations and 'collaborative' work, the process of co-creation typically also includes interaction with the client who commissions the translation. For the current study no interaction with a client was possible within the scope of the experiments. Client interaction was limited to the brief, in which translators and teams were instructed to translate the article for the quality news magazine *Elsevier* in the Dutch context and the quality newspaper *De Standaard* in the Belgian context, as part of a series of articles on health care systems.

Product evaluation

For this experiment, translation quality was assessed using the Preselected Items Evaluation (or PIE) method (H. Kockaert and Segers 2014; H. J. Kockaert and Segers 2017). With this analytical method, assessment is based on preselected elements (PIs) in the ST (punctuation marks, words, clauses . . .) that are likely to present difficulties to (aspiring) translators, because of a lack of fit between the goal of the TT and the (linguistic and extra-linguistic) means at the translator's disposal (Nord 1994, 2002, 2011). This method was chosen primarily for its time efficiency and reliability, providing

a strong foundation for text comparison across a diversified learner corpus, rather than for being an infallible quality measure (see Colman, Segers, and Verplaetse 2021; Van Egdom et al. 2018).

In the first stage of the analysis, we selected a total number of 26 ST items (see Appendix 1), covering various types of translation problems. These covered not just linguistic issues (e.g. grammar, spelling), but also stylistic (register), pragmatic and structural challenges, considering the broader translation context including commissioners and end users. The added value of the PIE method lies in its emphasis on dichotomous scoring: as opposed to other analytical evaluation methods, PIE only distinguishes between 'correct' and 'incorrect'. Dichotomous scoring ensures high intra- and interrater reliability (Van Egdom et al. 2018; see also Segers and van Egdom 2018, 30, 57, 59).

In the second phase of the evaluation process, a dichotomous selection of translation solutions was made: the evaluators[5] discussed the PIs and listed solutions that were good or passable. In this phase, the list of adequate solutions is reliant on the accumulated experience of evaluators. In the third phase, the solutions in the learner corpus that corresponded with the PIs were scrutinised by the two evaluators. In this phase, the list of correct solutions was extended with solutions that were not foreseen in the second phase. In case of doubt or a low degree of agreement among evaluators, judgement on the correctness was passed based on the frequency of terms and collocations in parallel texts. As soon as a list of satisfactory solutions was drawn up, a cross-check was performed once more to ensure consistency, repeatability, reproducibility and stability, as well as transparency.

The following step in the evaluation was the calculation of the 'raw' PIE score, i.e. the fraction of correctly solved ST items per translation (x/26), for each ITT and for the team versions (excluding TT[2]). Despite its common use in translation evaluation, it is noted that the 'raw score' may not accurately represent PI scores, as it relies largely on evaluators' intuition (Tijtgat and Segers 2019, 320–321). For this reason, we refined our test by calculating the docimological p value and the d index of each item. It is important to note that the p value represents a different metric than the p value calculated for statistical significance: the docimological p value signifies item difficulty, indicating the percentage of candidates within a group who answered a PI correctly. The d index reflects the discriminatory power of an item by capturing differences in performance between a bottom group and a top group. By eliminating items with poor p values and d index rates, we introduced the element of norm-reference, essentially considering group performance in the refinement process. This approach aligns with the insights from Matlock-Hetzel (1997) and Wiersma and Jurs (1990), who provided a foundation for docimological item justification. They indicate that good items have a p value between 0.27 and 0.73 and a d index above 0.29.

Finally, we compared the raw and docimologically justified PIE scores of the ITTs to those of the CTTs. In the latter case, we also focused on TT[1]s and final CTTs, since the TT[1] scores served as an ideal basis for comparison, because they were produced under similar circumstances as the ITTs.

Translation processes

In the second phase of the study, findings on output quality were enriched with findings on collaborative translation processes. Earlier tentative studies linking TT quality to student translation behaviour suggest that there is no single yardstick for successful translator behaviour. Poor translations in the study are due to an uncoordinated translation style. Translators who performed well worked in an organised manner, but employed different methods. In other words, different approaches appear to lead to success (Verplaetse et al. 2018). This study investigated whether this is also true for collaborative processes.

As mentioned above, process data was gathered with Inputlog (7 for Windows), a tool that registers keystrokes, mouse movements and clicks with their corresponding timestamps (Leijten et al. 2014). Inputlog data were used to retrace the decision-making steps that were considered crucial in the development of each translation, particularly for the ST items selected in the PIE evaluation.

Due to project limitations, a full analysis of the process files was unfeasible; focus was instead on the development of TT[1]s CTTs. We adopted both a macro and a micro perspective. A broader view paid attention to process phases (time spent on the orientation, translation and revision phases, as per Schrijver, Van Vaerenbergh, and Van Waes 2012), types of external resources used and time spent on consulting resources, pausing behaviour (pause time, total number of pauses >2 seconds, average and median pause duration). Additionally, process data were retraced to reconstruct the solutions for PIs, considering that each item was associated with a translation problem, and that translation problems are likely to place heavier demands on translator competence. The general, source and revision analyses, as well as the process graphs, generated by Inputlog, proved to be very helpful to reconstruct the development of ITT, TT[1] and CTT versions. Revision processes in the collaborative setting were monitored with the LISA QA forms that detailed revisers' impressions, identified errors, error types, weightings, and suggested revisions (see Appendix 2).

Results

Product quality

As mentioned above, the two evaluators first calculated the 'raw' item scores of the ITTs, TT[1]s and CTTs. Since the learner corpus consists of translations from a heterogeneous group, mean ITT scores were calculated distinguishing between study programme phases. This distinction was made as the learner corpus serves as a 'monitor' corpus to check if student progress (or improved TT quality) is evident throughout translation curricula. In this respect, the results were reassuring.

As shown in Table 1, the individual 'raw' mean scores show improved product quality: students in the first year of the undergraduate programme obtained a mean score of 14.2; students in the final stages of the bachelor's programme obtained considerably higher marks (18.4 on average). Students in the master's programme performed slightly better than their peers in the final year of the undergraduate programme, with a mean score of 19.3. Looking at the mean scores of all ITTs, we see that 69.4% of items (18/26) were solved correctly.

Table 1. Means of raw PIE scores in ITTs.

Individual Assignment	Mean Raw Scores (/26)
BA1 (7)	14.2
BA3 (45)	18.4
MA (9)	19.3
ALL (61)	18

Table 2. Means of raw PIE scores TT[1]s and CTTs.

Team Assignment	Mean Raw Scores (/26)
TT[1] (13)	17.6
CTT (13)	19

When comparing the raw PIE scores of ITTs to those of the collaborative TT[1]s, it is striking that the mean score of the TT[1]s (17.6) is lower than the ITT mean score (18) (see Table 2). The TT[1] scores were expected to be on a par with mean scores of all ITTs by the MA students who worked individually combined. The lower than average score of TT[1] may be due to individuals becoming careless, knowing peers will address their typos or simple semantic errors later in the process. In 10 out of the 13 cases, product quality did improve throughout the collaborative process. In 2 cases, product quality was similar in the TT[1] and the CTT. Interestingly, one team produced a CTT (9/26) that scored lower than their TT[1] (10/26).

Initial observations of raw scores in the team assignment (Table 2) suggest that collaboration enhances TT quality. However, this hypothesis falters if Table 1 is considered: MA students working individually outshone peers, with their ITTs scoring .3 item higher than the CTT mean.

The next step was the elimination of items with unacceptable *p* values and d indices. Out of the 26 items, 18 items were eliminated: the items were deemed too difficult or too easy and did not allow for proper discrimination between (good and poor) translations. Tables 3 and 4 show the outcomes of the recalculation of PIE scores based on docimologically justified items. Table 3 shows that students achieved steady progress throughout the study programmes, with a mean score of 4 for ITTs produced by students in the first year of the bachelor's programme and mean scores of respectively 4.8 and 4.9 for

Table 3. Means of PIE scores ITTs (with docimologically justified items).

Individual Assignment	Mean justified items (/8)
BA1 (7)	4
BA3 (45)	4.8
MA (9)	4.9
ALL (61)	4.7

Table 4. Means of PIE scores ITTs (with docimologically justified items).

Team Assignment	Mean Justified Items (/8)
TT[1] (13)	4.7
CTT (13)	4.8

ITTs of last-year students in the bachelor and students enrolled in the master's programme. The mean PIE score of all ITTs combined was 4.7. On average, almost 59% of the total number of docimologically justified items were answered correctly.

A comparison of Tables 3 and 4 reinforces our initial findings: students working on the TT[1]s (4.7) were outflanked by individual students of the master's programme (see ITTs in Table 3). The master's students' individual work even outshone their fellow students' collaborative work (respectively 4.9 and 4.8). However, the differences between groups seem to be less marked here.

This pattern (of confirmed but subtle differences) is also apparent from the differences between TT[1]s and CTTs. In 8 out of the 13 cases, product quality had improved throughout the collaborative process. In 4 cases, quality was similar in the TT[1] and the CTT. One team's CTT score remained lower than TT1 upon recalculation based on justified items. These results suggest that collaborative effort adds minimal value, and that carelessness of individuals is apparent, when norm-reference is taken into account in quality assessment.

Translation processes

For this part of the project, we aimed to study translation quality evolution in our learner corpus by focussing on the textual differences between versions produced within the skills lab. The qualitative differences of the TT[1]s and CTTs – between versions as well as between groups – prompted us to critically examine both the textual material generated in the skills labs and certain process log files and QA forms. Our goal was to further our understanding of quality enhancement through collaborative translation. Since the mean (raw) score of the TT[1]s was lower than the ITT mean score and the product quality of CTTs appears to be superior to that of the TT[1]s, we were interested in exploring potential processual differences between teams. The aim was to find the key that led to successful collaboration in the teams that produced high-quality CTTs. With this aim in mind, we analysed process data of the two teams with the highest PIE scores for CTTs (Team 1 and Team 2; both 22/26) and contrasted these findings with patterns found in data of the team with the lowest score (Team 3; 9/26).

The translation processes of the TT1s of the CCTs in this study suggest numerous ways to produce quality translations. As shown in Table 5, the total time that the students spent on the TT1s differed considerably. Team 2 delivered their TT1 quite fast, pausing less often than Team 1. Team 1 worked for more than 2 hours on the first draft, pausing quite frequently, but the proportion of time spent pausing by Team 1 relative to the total process time is lower (43% of their total TT1 process time) than the proportion of time spent pausing by Team 2 (58% of process time). However, their average and median pause durations are fairly similar. The process time and pausing behaviour suggest that

Table 5. General process data of the TT[1]s.

TT1 Process Variables	T1	T2
Total Process Time	02:08:35	00:54:32
Total Pause Time	00:55:21	00:31:42
Total Number of Pauses (>2)	395	196
Average Pause Duration (ms)	8.408	9.704
Median Pause Duration (ms)	4.438	4.091

Team 2 is more efficient in producing the TT1, but Team 1 May have devoted more attention to the task (e.g. by reflecting on and researching translation problems more extensively).

For the subsequent analysis, the phasing of the translation process was examined. This involved distinguishing between orientation, translation, and revision phases (Englund Dimitrova 2005; Jakobsen 2002; Schrijver, Van Vaerenbergh, and Van Waes 2012). It is important to note that the description of the revision phase may be ambiguous, as it refers to self-revision conducted by the translator. In terms of organising the translation process, translators from Team 1 and Team 2 appear to take different approaches. This also becomes apparent in the process graphs (Figures 1 and 2), in which the green (product) line represents the number of characters contained in the translation and the blue (process) line the characters typed in the Word file. These two lines show that the translator of Team 2 started translating almost immediately, whereas the translation process of Team 1's translator was characterised by an orientation phase of roughly 40 minutes, during which the ST and parallel texts were consulted. Although the writing phases of both teams were quite linear, with relatively few revision loops during TT production, the frequency of source use could not have been more different. This provides interesting clues about differences in self-efficacy and the students' need and/or willingness to search for information in external sources. As shown by the orange lines below the graph, Team 1 frequently switched between Word and external sources, conducting internet searches for target language synonyms and gathering background information from hospital websites and online publications. By contrast, Team 2 exclusively consulted the ST, suggesting great confidence in their own bilingual and extralinguistic knowledge. Self-revision behaviour also seems to be highly distinct. In research

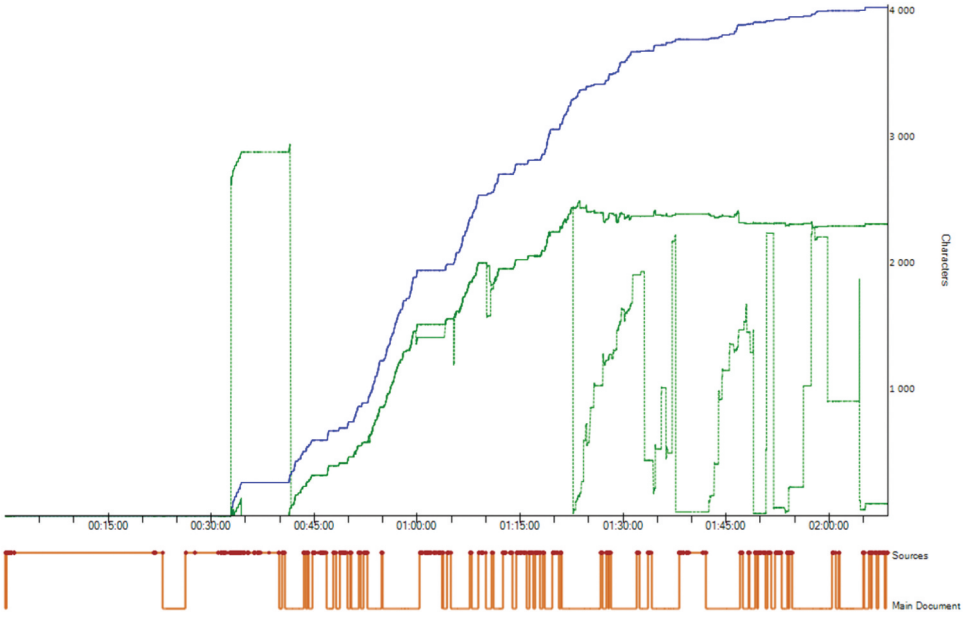

Figure 1. Process graph TT[1], Team 1.

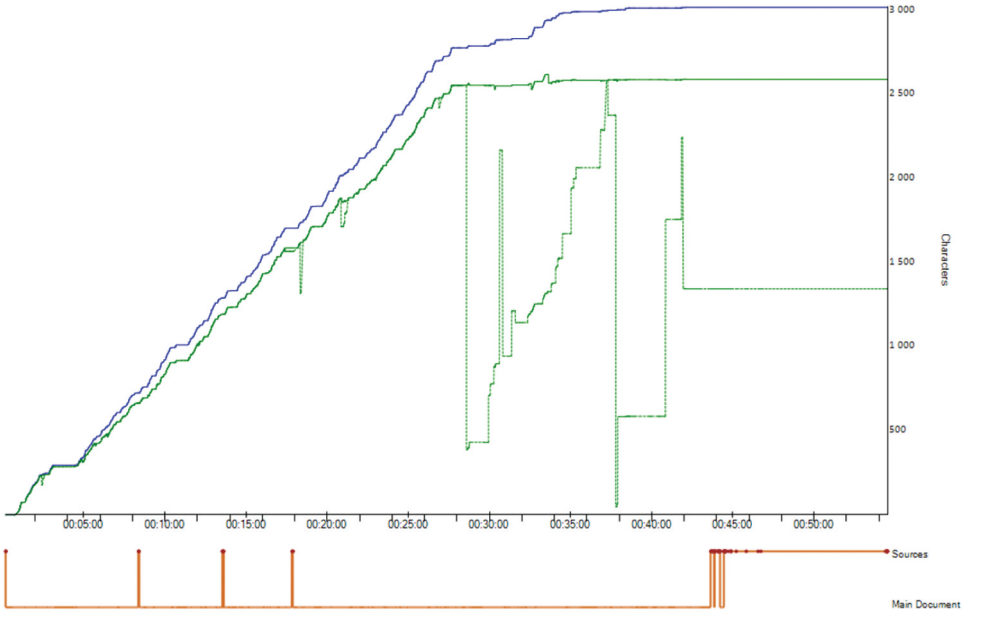

Figure 2. Process graph TT[1], Team 2.

(e.g. Dragsted and Carl 2013), a conventional distinction is made between online revision, end revision, and constant revision. Online revision involves self-correction during translation, while end revision occurs mainly at the final stage, after producing a draft TT. Constant revision combines elements of both online and end revision methods. Although Team 2 spent more time on end-revision, the revision or post-writing phase of Team 1 appears to be more intense. Following the dotted green line, which represents the cursor position movements, we can observe roughly six revision loops in Figure 1 (after 1h22); Team 2 revised their TT three times within 28 minutes after, finishing the initial draft.

Despite different work methods, both teams have equal confidence in their initial translation solutions for the preselected items (PIs). Team 1 revised only 4 PIs, and Team 2 revised 6 PIs in their TT[1]s. One of these concerned the same PI, 'those who were', which both translated rather generally at first with 'de mensen die' (the people that). They later revised their translations, using specific references: 'de patiënten die' [the patients who] and 'de participanten die' [the participants who]. No clear differences in pause patterns could be discerned regarding item solutions. The overall source usage difference became clear when analysing external sources before and after translation for each PI: Team 1 consulted sources for the first translation solution of 6 PIs, whereas Team 2 relied exclusively on internal resources.

Interestingly, the Team 3's TT[1] process, the team with the lowest number of correctly translated PIs, is a mixture of the previously discussed translation processes, with important distinctions. Like Team 2, Team 3 started drafting the TT almost immediately but their total process time (01:30:15) and external source usage resembled the process of Team 1. However, source use by Team 3 appeared to be predominantly

targeted at finding equivalents, as demonstrated by the frequent consultation of Linguee and Van Dale's English to Dutch dictionary. Team 3 consulted sources for 14 of the 24 PIs. The total time spent on finding equivalents also suggests that Team 3 did not reflect much on the information found, nor on alternative solutions. These observations align with existing research: non-expert translators tend to make limited and less critical use of sources (Krings 1986; Zheng 2014). This impression gains strength when we observe the total pausing time (33 minutes; 37% of total process time), which is short compared to the other two teams, and the large number of pauses (253; median pause duration of 4,031 ms). As shown in Figure 3, Team 3 spent nearly 49% of total process time on end-revision: after roughly 48 minutes, they performed 3 revision loops, carrying out relatively few revisions (e.g. spelling and idiom). Regarding the translation of the PIs, Team 3 appeared to be less confident than the other teams. They consulted external sources for considerably more PIs (14/24) and they also decided to self-revise more PIs (9) in their TT[1].

After the translators had produced the TT[1]s, the text was handed over to the designated revisors. The subsequent analysis focuses on the revisors' task. The QA forms provided insights into the revisors' actions and beliefs: attention below is primarily directed towards those forms. As can be inferred from the QA form, the revisor of Team 1 provided extensive feedback. Still, the revisor looks favourably upon the translation: they state that the meaning of the ST is conveyed correctly, and that the style makes for a pleasant read. Despite the positive reception of the TT[1], the revisor identifies no less than 14 'errors'. Only 1 solution was considered a major error (the mistranslation of 'new data'). Eight translation solutions were not considered incorrect by the revisor; changes were qualified as 'preferential'. The error report suggests that the reviser did not refrain from criticism, but that this was aimed at making an already acceptable text more fluent.

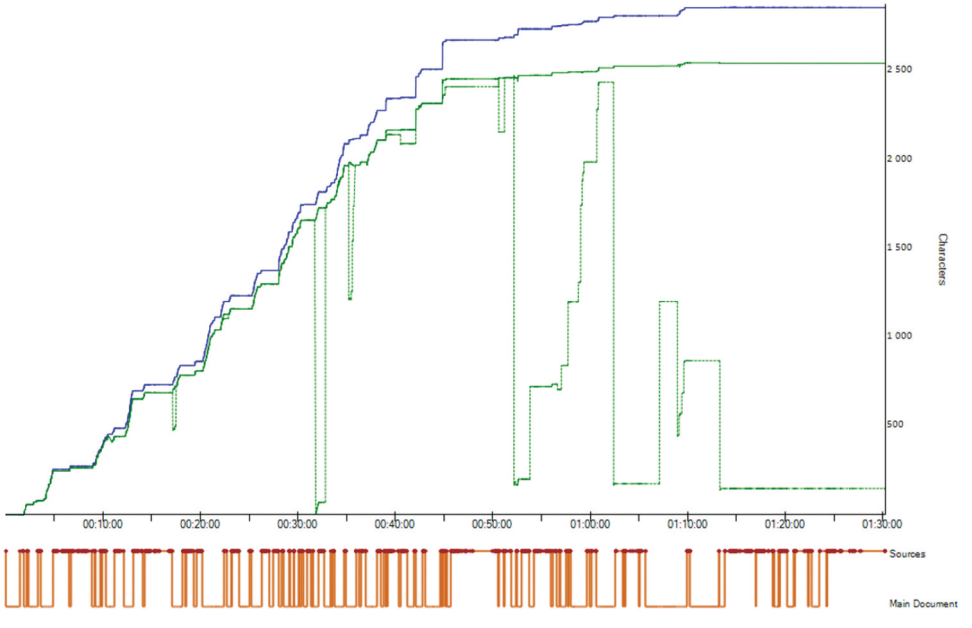

Figure 3. Process graph TT[1], Team 3.

The feedback quality is acceptable: some suggestions lead to the improvement of the TT (e.g. 'de eigen bijdrage' [their own contribution], 'overheid' [the government], others make little to no difference ('bewijslast' [burden of proof], 'dat wat hij financiële toxiciteit noemt' [what he calls financial toxicity]), and some suggestions are simply detrimental to text quality ('privé polis' [private insurance]).

As can be inferred from the QA form, the revisor of Team 2 put less effort into the revision task. Their general opinion was also relatively positive, but the report also suggests that the translator did not render the style adequately. They state that the text is too 'formal' for its intended purpose. At first glance, a small number of changes (6) seems to have been made to improve the TT quality in terms of style. However, upon closer inspection, the revisor of Team 2 appears to have made more changes to the text (9). The discrepancy between changes in the revised text and the number of errors in the report can be explained quite easily: whereas the revisor of Team 1 was precise and concise in their feedback, the revisor of Team 2 used larger units of revision in the QA form (entire sentences), thereby hindering error identification. No errors in TT[1] were considered severe; in 4 out of the 9 cases, the changes were qualified as 'preferential'. Looking at the 'correction' column in the QA form, we see that little effort was made to iron out stylistic creases, which is remarkable, considering the fact that the text was deemed 'too formal'. Text elements were removed (e.g. 'ook' [also], 'uiteindelijk' [finally]) with minimal effect. The only TT element that was correctly addressed, was 'en hen tijdig in contact brengen met ondersteuning' (as a translation for: 'and connect them with resources in a timely fashion'); as an alternative, 'en hen tijdig ondersteuning bieden' [and provide them with support in a timely fashion] was proposed. Unfortunately, the term 'resident' is translated as 'inwonend arts' [live-in doctor]: a mistranslation on the part of the revisor.

Product quality for Team 3's TT[1] was poor. Contrary to expectation, the revisor of TT[1] seemed satisfied with the first version; in their general impression, they stated that the translator 'did a solid job!' These kind words are supplemented with no more than 6 changes in the QA form. What is striking is that 4 out of the 6 implemented changes were considered preferential changes in the categories of semantics and style ('met [alle stadia van de ziekte]' [with all stages of the disease], 'behandeling' [treatment], 'medische [kosten]' [medical expenses], 'betalen' [pay]). The other changes address what the revisor qualified as minor errors ('heeft' [has], 'senior auteur' [senior author]. Remarkably, only two revisions help improve the text (improved congruence), but two other revisions further impair TT quality (incorrect collocation, poorer choice of term). Three minor changes were not registered in the form, but were observed in TT[2].

Before we continue with the CTT data analysis, we point out that Team 2's revision approach, as shown in the QA forms, differs from typical practices in higher-ranked texts. Figure 4 shows a downward trend in the number of suggestions and corrections; whereas high-ranked teams provided a thorough revision, the teams that performed poorly delivered slapdash revision work (few word-centred corrections).

The keystroke logging data of the proofreading stage largely show the same processes as for the TT[1]s: Team 1 worked considerably longer than Team 2 (44:00 versus 6:12 minutes). They also paused more frequently (117 versus 32 times) and on average for longer periods of time (12.442 ms versus 7.246 ms). However, the median pause duration (4.031 ms versus 5.484 ms) indicates that

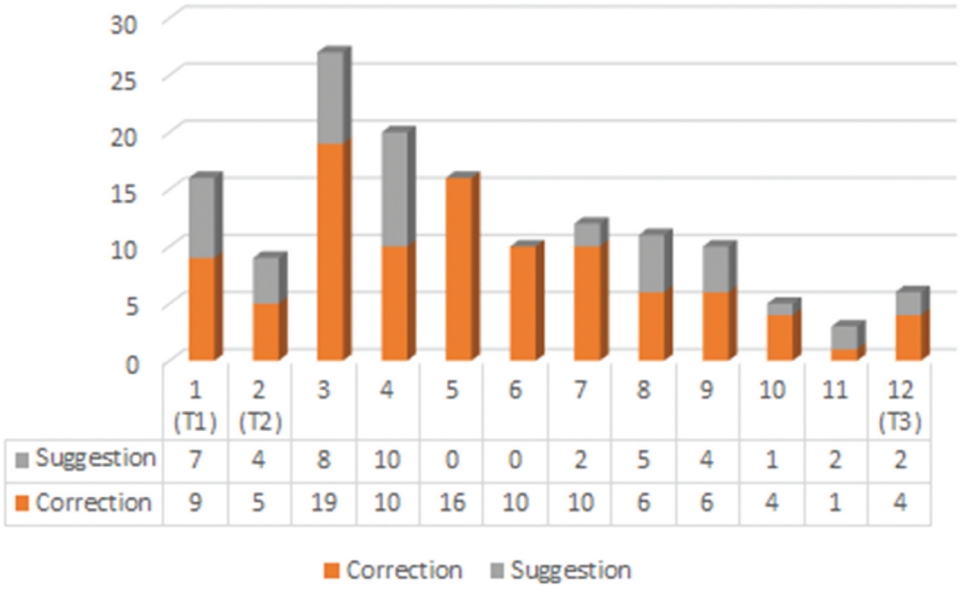

Figure 4. Number of suggestions and corrections in QA forms (ranked).

pause duration differed: throughout the CTT process of Team 1: a detailed pause analysis showed several pauses ranging from 10 seconds up to 50 seconds. This suggests that attention was paid to an array of aspects, posing different levels of cognitive load (e.g. ranging from spelling to interpretation issues). Team 1 went through the translation twice, carrying out 34 revisions, yet implementing only 6 of the revisor's suggestions. This detailed revision behaviour combined with the pausing behaviour and the use of external sources suggest that Team 1 took the proofreading process as an opportunity to 'revise' the translation once more. In contrast, Team 2's CTT process was more linear and shorter, without external source consultation. They went through the translation once, carrying out 10 revisions, nearly all of which were suggested by the revisor.

Interestingly, Team 3's proofreading process data resemble those of Team 2: short process time (10:05 minutes), few pauses (27 in total; but with a shorter median duration of 2.766 ms), a linear revision, no consultation of external sources, implementation of nearly all suggestions made by the revisor (plus two additional revisions). These combined findings suggest that the CTT processes were carried out far less critically by Teams 2 and 3 than by Team 1. However, since the baseline quality of Team 2's TT was far better than Team 3's TT, this lack of a critical approach was less detrimental to product quality.

Discussion

The evaluation analysis has shown that peer learning seems to benefit translator trainees: after evaluation with the full selection of PIs, the quality of team translations appeared to have improved during the collaborative translation process in 10 out of 13 cases; after evaluation with 8 docimologically justified items, the quality of 8 team translations

appeared enhanced as a result of collaboration. These results seem to support the hypothesis that collaborating translators are more likely to produce better TTs.

However, some findings were clearly at odds with this hypothesis. Having made a comparison between ITTs and TT[1]s, it was observed that students who worked on a collaborative task tended to be more careless: the mean scores of the ITTs, produced under similar circumstances as the TT[1]s, were better, and this was primarily due to 'simple' errors that could easily be avoided. We concluded that trainee translators who work in a team seem to rely heavily on fellow-translators, whom they entrust with the responsibility of improving the TT.

In a reconstruction of the creation of the versions in the team assignments, an attempt was made to tap into the social and organisational dynamics in collaborative processes. As in earlier research, process data revealed that various strategies can be employed with good result (see Verplaetse et al. 2018). Still, some elements attracted attention in the critical investigation of collaborative processes. Regardless of the working method, collaborative translation seems to benefit from high numbers of revision suggestions, which can be said to be indicative of higher engagement in negotiation of meaning. The data suggest that 'engaged' project managers-proofreaders did not merely rubber-stamp suggestions made by the revisors; the suggestions seemed to prompt them to think about translation problems at hand, examine problems thoroughly and initiate the renegotiation of meaning, ultimately encouraging them to work their way through the text to bring about changes throughout the text. Increased critical engagement in collaborative processes and processes of social translation was thus considered the true driver of quality. This finding is reinforced by the insights gained during the analysis of the group with the poorest results. Team 3 appeared to be easily satisfied, proofreading data attested to a less critical approach to suggestions for improvements, and superficial use was made of sources.

Conclusion

Starting from the notion of TT quality, an attempt has been made to assess whether collaborative processes can be considered drivers of translation quality. To this end, the quality of translated products of trainee teams were compared with those of individual peers. The results of the comparison seem to give rise to optimism: TT quality seems to improve in a skills lab setting. However, the textual comparison also provided extra food for thought: despite improved end product quality, the collaborative setting seemed to lead to carelessness and complacency. What is more, in many cases TT quality did not seem to improve dramatically with each new version: revisions were often scarce. Critical engagement in a collaborative setting seems conducive to higher quality output, and, as suggested in social constructivist theories, is an indicator of quality of learning (cf. INSTB et al. 2018; Kiraly 2000).

It should be emphasised that this study has limitations. Analysing quality is always risky, as it requires a sound operationalisation of the notion of quality. In this study, quality is conceived of as a textual characteristic: by adopting the PIE method, a product-based approach was favoured. A different take on quality, for instance, one that is user-based and takes heed of clients' and text users' appraisal (see Garvin 1984), might yield different results, since users can be said to play an equally important role in the

construction of meaning (through the act of reading the TT in its intended context). Additionally, the study was limited to a single ST. Future research would benefit from a variety of STs from similar (health economics) and different domains (such as law), as well as exploring other language pairs. Another crucial factor that requires attention is the limited size of sub-group samples in this study: small sample sizes impede the statistical determination of the significance of differences.

This brings us to the avenues for future research. Although the variables that were scrutinised in earlier process research did not correlate with product quality (see Verplaetse et al. 2018), confirmatory factor analysis might reveal a hidden relationship between observed process variables. Furthermore, a firmer handle can be gained on the improvement of product quality throughout the collaborative processes by annotating errors using analytical categories. Finally, future research should explore whether additional training sessions on interactive revision and quality assurance, beyond mere textual editing and QA procedures, could enhance collaborative translation. In the future, an additional test with a control group and a group that undergoes training might shed light on this matter. All these avenues have to be taken if we are to deepen our understanding of collaborative translation in a didactic context.

Notes

1. The students had access to free use of internet resources, incl. dictionaries, text corpora, open access MT systems.
2. Again, the students had access to free use of internet resources. In addition, team members were allowed to communicate with each other.
3. Prior to the skills lab module, students had to apply for at least one function. At that point, they had received training in translation, project management, revision and proofreading. Translation was addressed in the 'Vertaalateliers' (or translation workshops, 5 ECs) and project management, revision, proofreading were covered in the course unit 'Theorie en praktijk van het vaktalig vertalen' (Theory and Practice of Specialised Translation, 5 ECs).
4. Students had received instructions about the use of these quality assurance forms during an earlier course, where it was part of a full day's simulation assignment at an on-the-job translation internship address.
5. Both evaluators had over 5 years of experience in translation teaching and assessment.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Gys-Walt Van Egdom http://orcid.org/0000-0002-7521-6792

References

Adab, B. 2000. "Evaluating Translation Competence." In *Developing Translation Competence*, edited by C. Schäffner and B. Adab, 215–228. Amsterdam/Philadelphia: Benjamins.

Bowker, L. 2001. "Towards a methodology for a corpus-based approach to translation evaluation." *Meta Journal des traducteurs/Translators' journal* 46 (2): 345–364. https://doi.org/10.7202/002135ar.

Bowker, L. 2003. "Corpus-Based Applications for Translator Training: Exploring Possibilities." In *Corpus-Based Approaches to Contrastive Linguistics and Translation Studies*, edited by S. Granger, J. Lerot, and S. Petch-Tyson, 215–228. Amsterdam: Rodopi.

Bruffee, K. A. 1973. "Collaborative Learning: Some Practical Models." *College English* 34 (5): 634–643. https://doi.org/10.2307/375331.

Buysschaert, J., M. Fernández-Parra, K. Kerremans, M. Koponen, and G. Van Egdom. 2018. "Embracing digital disruption in translator training: technology immersion in simulated translation bureaus." *Revista Tradumàtica Tecnologies de la Traducció* 17:125–133.

INSTB, Buysschaert, J., M. Fernandez Parra, and G. W. van Egdom. 2017. "Professionalising the Curriculum and Increasing Employability Through Authentic Experiential Learning: The Cases of INSTB." *Current Trends in Translation Teaching and Learning E* 4:78–111. Retrieved from http://www.cttl.org/uploads/5/2/4/3/5243866/cttl_e_2017_3.pdf.

Colina, S. 2009. "Further Evidence for a Functionalist Approach to Translation Quality Evaluation." *Target: International Journal on Translation Studies* 21 (2): 235–264. https://doi.org/10.1075/target.21.2.02col.

Colman, A., W. Segers, and H. Verplaetse. 2021. "Refining the PIE Method (Preselected Items Evaluation) in Translator Training." *Current Trends in Translation Teaching and Learning E* 8:236–275. https://doi.org/10.51287/cttle20218.

Dillenbourg, P. 1999. "What Do You Mean by Collaborative Learning?" In *Collaborative Learning: Cognitive and Computational Approaches*, edited by P. Dillenbourg, 1–19. Oxford: Elsevier.

Dragsted, B., and M. Carl. 2013. "Towards a Classification of Translator Profiles Based on Eye-Tracking and Keylogging Data." *Journal of Writing Research* 5 (1): 133–158. https://doi.org/10.17239/jowr-2013.05.01.6.

EMT Board. 2017/2022. *European Master's in Translation: Competence Framework*. https://commission.europa.eu/document/download/b482a2c0-42df-4291-8bf8-923922ddc6e1_en?filename=emt_competence_fwk_2022_en.pdf.

Englund Dimitrova, B. 2005. *Expertise and Explicitation in the Translation Process*. Amsterdam/Philadelphia: John Benjamins.

Galán-Mañas, A., and A. Hurtado Albir. 2015. "Competence Assessment Procedures in Translator Training." *The Interpreter and Translator Trainer* 9 (1): 63–82. https://doi.org/10.1080/1750399X.2015.1010358.

Garant, M. 2009. "A Case for Holistic Translation Assessment." *AFinLA-e Soveltavan kielitieteen-tutkimuksia* 1:5–17.

Garvin, D. G. 1984. "What Does Product Quality Really Mean?" *Sloan Management Review* 26 (1): 25–43.

Göpferich, S. 2009. "Towards a Model of Translation Competence and Its Acquisition: The Longitudinal Study 'TransComp." In *Behind the Mind: Methods, Models and Results in Translation Process Research*, edited by S. Göpferich, A. L. Jakobsen, and I. M. Mees, 11–37. Copenhagen: Samfundslitteratur.

Gouadec, D. 2007. *Translation as a Profession*. Amsterdam: Benjamins.

Hönig, H., and P. Kussmaul. 1982. *Strategie der Übersetzung. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr.

Huertas Barros, E., and J. Vine. 2016. "Translator trainers' Perceptions of Assessment: An Empirical Study." In *Translation and Meaning, New Series*, edited by M. Thelen, G. W. van Egdom, D. Verbeeck, L. Bogucki, and B. Lewandowska-Tomaszczyk, 29–39. Vol. 41. Frankfurt am Main, Berlin, Bern, Brussels, New York, Oxford, Wien: Peter Lang.

Hurtado Albir, A., and P. Taylor. 2015. "The Acquisition of Translation Competence. Competences, Tasks, and Assessment in Translator Training." *Meta Journal des traducteurs/Translators' journal* 60 (2): 256–280. https://doi.org/10.7202/1032857ar.

ISO. 2015. *ISO 17100: Translation Services: Requirements for Translation Services*. Geneva: International Organization for Standardization.

Jakobsen, A. L. 2002. "Translation Drafting by Professional Translators and by Translation Students." In *Empirical Translation Studies: Process and Product*, edited by G. Hansen, 191–204. Copenhagen: Samfundslitteratur.

INSTB (Kerremans, K., M. Fernandez-Parra, K. Konttinen, R. Loock, S. Vandepitte, I. Van der Lek-Ciudin, and G. W. Van Egdom). 2018. *Assessing Interpersonal Skills in Translator Training: The Case of INSTB*. Accessed 07 14, 2021. http://www.instb.eu/wp-content/uploads/2018/09/didTRAD2018_INSTB_1.pptx .

Kiraly, D. 2000. *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*. Manchester: St.

Koby, G., P. Fields, D. Hauge, A. Lommel, and A. K. Melby. 2014. "Defining Translation Quality." *Tradumàtica tecnologies de la traducció* 12 (12): 413–420. https://doi.org/10.5565/rev/tradumatica.76.

Kockaert, H. J., and W. Segers. 2017. "Evaluation of Legal Translations: PIE Method (Preselected Items Evaluation)." *Journal of Specialised Translation* 27:148–163.

Kockaert, H., and W. Segers. 2014. "Evaluation de la traduction: la méthode PIE (Preselected Items Evaluation)." *Turjuman* 23 (2): 232–250.

Konttinen, K. 2022. *From Competence to Self-Efficacy: Measuring Student Progress in Translation Company Simulation*. Turku University. https://www.utupub.fi/bitstream/handle/10024/154672/AnnalesB584Konttinen.pdf?sequence=1&isAllowed=y.

Krings, H. P. 1986. "Translation Problems and Translation Strategies of Advanced German Learners of French." In *Interlingual and Intercultural Communication*, edited by J. House and S. Blum-Kulka, 263–275. Tubingen: Gunter Narr.

Leijten, M., and L. Van Waes. 2013. "Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes." *Written Communication* 30 (3): 358–392. https://doi.org/10.1177/0741088313491692.

Leijten, M., L. Van Waes, K. Schriver, and J. R. Hayes. 2014. "Writing in the Workplace: Constructing Documents Using Multiple Digital Sources." *Journal of Writing Research* 5 (3): 285–337. https://doi.org/10.17239/jowr-2014.05.03.3.

Mariana, V., T. Cox, and A. K. Melby. 2015. "The Multidimensional Quality Metrics (MQM) Framework: A New Framework for Translation Quality Assessment." *Journal of Specialised Translation* 23:137–161.

Matlock-Hetzel, S. 1997. Basic Concepts in Item and Test Analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, January, 1997. Retrieved from https://files.eric.ed.gov/fulltext/ED406441.pdf.

Moorkens, J., S. Castilho, F. Gaspari, and S. Doherty. 2018. "Introduction." In *Translation Quality Assessment: From Principles to Practice*, edited by J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty, 1–9. Berlin: Springer.

Nord, C. 1994. "It's Tea-Time in Wonderland: 'Culture-markers' in Fictional Texts." In *Intercultural Communication*, edited by H. Pürschel, 523–538. Frankfurt: Peter Lang.

Nord, C. 2002. *Lernziel: Professionelles Übersetzen: Spanish-Deutsch*. Wilhelmsfeld: Gottfried Egert.

Nord, C. 2011. *Funktionsgerechtigkeit und Loyalität: Theorie, Methode und Didaktik des funktionalen Übersetzens*. Berlin: Frank & Timme.

PACTE. 2003. "Building a Translation Competence Model." In *Triangulating Translation: Perspectives in Process Oriented Research*, edited by F. Alves, 43–66. Amsterdam: John Benjamins.

PACTE. 2020. "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.

Risku, H. 2016. "Situated Learning in Translation Research Training: Academic Research as a Reflection of Practice." *The Interpreter and Translator Trainer* 10 (1): 12–28. https://doi.org/10.1080/1750399X.2016.1154340.

Saldanha, G., and S. O'Brien. 2014. *Research Methodologies in Translation Studies*. London: Routledge.

Schrijver, I., L. Van Vaerenbergh, and L. Van Waes. 2012. "An Exploratory Study of Transediting in students' Translation Processes." *Hermes: Journal of Language and Communication in Business*, (49): 99–117. https://doi.org/10.7146/hjlcb.v25i49.97740.

Segers, W., and G. W. van Egdom. 2018. *De kwaliteit van vertalingen. Een terminologie van de vertaalevaluatie*. Kalmthout: Pelckmans.

Thelen, M. 2006. "The New European Standard on Translation Services and Translation Quality Assessment: Their Place and Role in a Professionally Oriented Course." *Qualifying Translators: The Role and Responsibilities of Higher Education*, Rennes, France. http://www.colloque.net/archives/2006/UCR206.htm .

Thelen, M. 2016. "Quality and Objectivity of Testing and Assessing in Translator Training: Dilemmas and Suggestions." In *Translation and Meaning, New Series*, edited by M. Thelen, G. W. van Egdom, D. Verbeeck, L. Bogucki, and B. Lewandowska-Tomaszczyk, 79–98. Vol. 1. Frankfurt am Main, Berlin, Bern, Brussels, New York, Oxford, Wien: Peter Lang.

Tijtgat, E., and W. Segers. 2019. "Wat is een goede vertaling? Vertaalevaluatie: methodes en technieken." In *In balans. Een inleiding tot vertaal- en tolkwetenschap*, edited by G. De Sutter and I. Delaere, 307–328. Antwerpen: Acco.

Van Egdom, G. W., and W. Segers. 2019. *Leren vertalen. Een terminologie van de vertaaldidactiek*. Kalmthout: Pelckmans.

Van Egdom, G. W., H. Verplaetse, I. Schrijver, H. Kockaert, W. Segers, J. Pauwels, H. Bloemen, and B. Wylin. 2018. "How to Put the Translation Test to the Test? On Preselected Items Evaluation and Perturbation." In *Quality Assurance and Assessment Practices in Translation and Interpreting*, edited by E. Huertas-Barros, S. Vandepitte, and E. Iglesias-Fernández, 26–56. Hershey: Narr.

Verplaetse, H., and A. Lambrechts. 2022. "Analytical Translation Quality Assessment: Insights for translators' Training." In *Research into Translation and Training in Arab Academic Institutions*, edited by S. M. Shiyab, 89–112. New York: Routledge.

Verplaetse, H., G. W. Van Egdom, I. Schrijver, W. Segers, H. Kockaert, and F. Van Santen. 2018. "Is Target Text Quality Indicative of Translation and Translator Competence? A Bi-National Study on Learning Progress in a Diversified Learner Translator Corpus." In *Using Corpora in Contrastive and Translation Studies Conference* (5th edition). Book of Abstracts, edited by S. Granger, M. A. Lefer, and L. A. de Souza, 186–187. Louvain-la-Neuve, UC Louvain: Penha Marion.

Waddington, C. 2004. "Should Student Translations Be Assessed Holistically or Through Error Analysis?" *Lebende Sprachen* 49 (1): 28–35. https://doi.org/10.1515/LES.2004.28.

Wiersma, W., and S. Jurs. 1990. *Educational Measurement and Testing*. Needham Heights, MA: Allyn and Bacon.

Zheng, B. 2014. "The Role of Consultation Sources Revisited: An Empirical Study of English–Chinese Translation." *Perspectives* 22 (1): 113–135. https://doi.org/10.1080/0907676X.2012.746380.

Appendices

Appendix 1 – P values and d indices per PI

PI's: 'Cancer patients with insurance still face huge costs'

PI	p-value AVG(ALL)	[p-value] AVG(TG)	[p-value] AVG(BG)	d-Index
Face	0.857143	0.9375	0.75	0.1875
out-of-pocket health care expenses	0.589286	0.8125	0.3125	0.5
despite having	0.642857	0.5625	0.4375	0.125
health coverage	0.928571	1	0.75	0.25
Copays	0.375	0.4375	0.375	0.0625
Deductibles	0.375	0.5625	0.1875	0.375
financial stress	0.678571	0.9375	0.375	0.5625
senior author	0.982143	1	1	0
adds to the growing evidence	0.339286	0.6875	0.125	0.5625
Intervene	0.553571	0.6875	0.4375	0.25
Providers	0.392857	0.375	0.3125	0.0625
health care-related costs	*0.607143*	*0.75*	*0.3125*	*0.4375*
By comparison.	*0.321429*	*0.6875*	*0.1875*	*0.5*
little or average financial stress	0.196429	0.3125	0.125	0.1875
Overall.	0.607143	0.6875	0.5625	0.125
lead author	0.982143	1	1	0
radiation oncology	*0.696429*	*0.8125*	*0.5*	*0.3125*

Appendix 2 – LISA QA Form

Linguistic Quality Assurance

Job title:

Reviewer:
Translation status: **Pass/Fail (for information only)**

Reviewer's general opinion of translation:

Revision report:

Segment #	Source Segment	Target Segment	Correction (suggestion)	Error type (→error report)	Repeated error (Y/N)
					(use tab to expand)

Error Report

Error category	Number of errors (minor	Number of errors (major)	Number of errors (critical)	Number of errors (preferential)	Number of error points	Max. Error points allowed	Result (pass/fail)
Mistranslation	(ex. 2)		(ex.1)	(ex. 3)	(ex. = 2×1 + 1×5 + 3 × 0 = 7)	1	(ex. fail)
Accuracy						2	
Terminology						2	
Grammar						1	
Semantics						1	
Spelling						1	
Punctuation						1	
Style						1	
Country (local convention)						2	
Consistency						2	
Total						**14**	