The handle http://hdl.handle.net/1887/40210 holds various files of this Leiden University dissertation

**Author:** Karoubi, Behrouz  
**Title:** Assessor-centered translation quality assessment : a theoretical model and a case study  
**Issue Date:** 2016-06-15
Chapter VI

Conclusion

In this final chapter I will provide a summary of the main points of the previous chapters of the thesis. The summary will be then followed by some final remarks on generalizability, potential practical applications, and limitations of the study.

6.1. Summary

This thesis was comprised of three distinct sections: 1) a critical historical overview of translation theories with particular attention to translation quality assessment, 2) a theoretical model for translation quality assessment, and 3) a case study.

The underlying core question in the first section, i.e. Chapter One, was whether the existing translation theories and assessment models that implicitly or explicitly dealt with translation quality were effective enough at explaining what translation quality really was and how it should be assessed. The best way to find the answer to this question was deemed to be a critical evaluation of a collation of previous translation theories and translation assessment models to identify their strengths and weaknesses as well as the areas of focus which had been largely overlooked. Chapter One thus was devoted to tracing the vicissitudes of the concept of quality in the history of translation theories. In the course of the critical analysis of various translation theories in the beginning of the chapter, attempts were made to show how the meaning of translation, the expectations about its quality, and accordingly the way it was assessed were
constantly changing alongside the socio-cultural and ideological developments in the Western world during different historical periods before the advent of modern linguistics. The second part of Chapter One dealt with the analysis of the major modern approaches to the study of translation in the linguistics era, in order to discover what is meant by translation quality in these approaches and how it is assessed. The critical review of these approaches was a prerequisite for designing a new methodology for translation quality assessment; because in that way the strong points of previous theories and methodologies could be incorporated into the new model, which was to be developed in the following chapters of the thesis, and the possible pitfalls and shortcomings could be avoided. This would also make it easier to show the value of the new methodology by demonstrating precisely how and where it was similar to or different from its previous counterparts. In the course of the critical survey of the existing translation theories and assessment models in the first chapter of this thesis the following major problems were identified:

1. They did not take enough account of the pivotal role of assessors in translation assessment as a complex process of decision making.

2. The theoretical foundations of many of these theories and assessment models were based on the contested concept of equivalence and they relied excessively on the source text as a point of reference for assessment.

3. Many of them limit the skopos of assessment, i.e. the purpose for which the translation being assessed is going to be used, merely to the purpose the source text is originally written for (intention of author) or the one the translated text is produced for (purpose of translator).

4. They often did not make a clear distinction between ‘translation quality assessment’ and ‘translation competence assessment’. Whereas the main purpose of translation quality assessment is to make a judgment about the quality of the translated text, in translation competence assessment the translated text serves merely as an assessment tool which helps the
5. The majority of them assessed the quality of translation against a set of fixed standards/criteria that were believed to be universally valid. These problems were seen as universal problems that motivated creating a new model to explain the concept of translation quality assessment better. The primary question in the second chapter of the thesis then became whether an alternative theory/model could explain the process of translation quality assessment and its relating concept more effectively.

In the second chapter of the thesis, attempts were made to address the above-mentioned problems and offer practical solutions for them. The chapter began with a discussion about the great amount of confusion that exists in the use of terminologies relating to the area of translation quality assessment and tried to disambiguate and (re)define key terms that are often taken for granted, such as translation, assessment, and quality. The chapter then developed a theoretical model to describe different stages of the process of translation quality assessment, i.e. collection, synthesis, and interpretation of data, with an emphasis on the significance of the role of assessors in the process. The theoretical model put forward in this chapter incorporated useful descriptive and analytical features of different approaches discussed in the first chapter and also encompassed different socio-cultural and ideological factors which might affect the process of evaluation. Toward the end of the chapter, the primary focus of attention was shifted onto the role of assessors and the importance of the assessment skopos in the process of translation quality assessment. In general, since theoretical models of assessment seemed to be, at best, just analytical tools for collecting data, the presence of knowledgeable and skillful assessors was deemed to be absolutely necessary to successfully apply them and interpret the data.
The focus of research in Chapters Three, Four, and Five was on a local problem in Iran. The underlying core problem in this section was the lack of knowledge about the evaluative behavior of the (academic) translation assessors in a specific local context. The primary goal of these chapters was to gain insight into the evaluative behaviors of the assessors during different steps of the process of translation quality assessment within the specific context of Iran. Emphasizing the necessity of understanding the evaluative behavior of the translation assessors within the socio-cultural and historical context in which they perform the task of assessment, Chapter Three presented a concise account of the history of translation in Iran, where the participant assessors in the case study in current research came from. This chapter addressed major developments of translation throughout the history of Iran from antiquity to the present time focusing on historical turning points, such as the Arab conquest of Iran, the Iranian Constitutional Revolution, and the Islamic Revolution of 1979, that have given rise to different translation movements in this country.

Chapter Four detailed a case study which attempted to investigate different aspects of the evaluative behavior of two groups of assessors during different stages of the process of translation quality assessment within the specific socio-cultural context of Iran, where the study was conducted. The main purpose of this chapter was to apply the theoretical model put forward in Chapter Two to collect information as to how the participant assessors evaluate the quality of the selected translation corpus of the study. The participants in the case study were comprised of two distinct groups: 1) a group of academic assessors consisting of 20 university professors/lecturers, who were native speakers of Persian and had the experience of teaching English into Persian translation courses, and 2) a group of ordinary assessors, as representatives of end-users of translations, consisting of 20 Iranian undergraduate/post-graduate students majoring in a variety of subjects in Iranian universities, who were also native speakers of Persian. The materials chosen for inclusion in the case study were three translated news articles in Persian and their English source texts, which were
extracted from the website of the Voice of America (VOA). The main apparatus used for data collection in the case study was a web-based questionnaire which was especially designed to acquire data about the evaluative behavior of the participants during different stages of the process of translation quality assessment.

In Chapter Five the overall evaluative behavior of the participant assessors in the case study during different steps of the process of the assessment, i.e. identification of the problematic areas, linguistic description of the problems, description of the frustrated expectations, determining the gravity of problems, suggesting solutions to the problems, and stating the results of the evaluation, were analyzed and discussed. The results of the analysis of data are briefly summarized step by step in the following.

**Identification of the Problematic Areas:**
1. The academic participants did **not** identify a considerably greater number of problematic areas on average in the translated material than their ordinary counterparts.
2. The most frequently identified problematic areas by the ordinary assessors were less frequently identified by the academic assessors.
3. As regards the degree of agreement on identification of the problematic areas by different assessors within the same group, the analysis of the data indicated that there was just a moderate degree of agreement between them.

**Linguistic Description of the Problems:**
1. While the academic assessors were heavily slanted toward identifying the problems related to the lexical and stylistic properties of translations, the ordinary assessors had a relatively balanced performance.
2. The academic assessors were particularly focused on the referential meaning of the linguistic items at the micro-level. This suggested that the academic assessors were excessively concerned with preserving the equivalence relation at the micro-level. The ordinary assessors, on the other
hand, paid special attention to the grammatical function and the order of the
textual material and punctuation in the translated texts.

3. In describing a substantial percentage of the problems they identified in the
translations, academic assessors preferred not to select an option among the
predefined linguistic categories provided in the questionnaire, and
consequently chose the option ‘Other’ to offer their own new categories.
After careful examination of all the instances where the option ‘Other’ was
used by the academic assessors, it appeared that they used this option
mainly in situations where, in their view, something was wrongly added to
or deleted from the translation.

4. It was observed that, in several cases, assessors in both groups provided
different linguistic descriptions for the very same items they identified in
the translations as being problematic.

**Description of the Frustrated Expectations:**

1. The majority of the problems identified by the academic group of assessors
resulted from the frustration of the assessors’ expectations about the relation
between the translations and their source texts.

2. The non-linguistic (ideological, socio-cultural, and political) factors were
largely neglected by the majority of the academic assessors in the current study,
whereas the ordinary group seemed to be more sensitive to these factors.

**Determining the Gravity of Problems:**

1. The individual assessors within both groups did not consistently assign the
same hierarchical gravity values to the problems that fell under identical
linguistic categories. The abundance of the examples of assessors in this
study who assigned different weight to the problems with the same
linguistic descriptions indicated that in determining the gravity of the
problems belonging to the same linguistic categories, assessors might have
also taken the specific context in which the problems occurred into
consideration.
2. No significant degree of agreement was observed between individual assessors in both groups with regard to determining the gravity of the problems. There were, however, a few cases where the ratings were concentrated around the mean gravity scores.

**Suggesting Solutions to the Problems**

1. In many cases, different assessors came up with different solutions to the same problems they identified in the translated texts.

2. Academic assessors frequently made reference to the translators of the texts in their proposed solutions.

**Stating the Results of the Evaluation**

1. There was a statistically significant difference in the mean ratings of the quality that the ordinary and academic assessors gave to each of the three translations in the study. The academic assessors showed more generosity in assessing the quality of the translations, because they consistently gave better quality ratings to the texts than the ordinary assessors.

2. A statistically significant degree of reliability was observed between the ordinary assessors in the *ordering or relative standing* of their ratings, as well as in the absolute value of the ratings. However, because of the wide CIs, we could not confidently say that the population values of the intra-class correlations were sufficient. In other words, one cannot confidently claim that the significant degree of agreement observed between particular assessors participating in the study will be the same if the experiment is repeated with different assessors.

3. Similarly to the ordinary group, there was a statistically significant degree of reliability between academic assessors in the *ordering or relative standing* of their ratings, as well as in the absolute value of the ratings. Here again, because of the wide CIs, we could not confidently say that the population values of the intra-class correlations were sufficient.
4. As regards the final decision, the mean ratings of the decisions suggested by the assessors in each group were very similar to each other and strongly resembled the quality ratings that were given to the translations earlier by the assessors.

5. Academic assessors seemed to be less likely to reject a translation. There were some differences between the translations (Translation 2 seemed to differ from Translation 1 and Translation 3). None of the translations seemed particularly more likely to be rejected than any of the other translations.

At the end of Chapter Five, the following suggestions were made for improving the evaluative behaviors of the academic assessors in Iran:

1. Sufficient inclusion of theory and practice of translation quality assessment in the curriculum translator training programs in Iran;

2. Promotion of the sense of accountability among the (future) assessors by reminding them of the role they play in the process of assessment;

3. Placing more emphasis on non-linguistic (ideological, political, socio-cultural …etc.) factors in the translator training programs at Iranian universities;

4. Making a clear distinction between the pedagogical assessment of the students’ translational competence in classrooms and the professional assessment of translation quality in the real world;

5. Placing greater emphasis on learning the target language norms and conventions in the curriculum of translator training programs at the Iranian universities; and

6. Avoiding excessively source-text oriented teaching methodologies that have been traditionally practiced at Iranian universities so far.
6.2. Final Remarks

When I first started my research on translation quality assessment, like many other researchers in the field, I was ensnared by the delusion of the possibility of developing a universal and comprehensive model which was applicable to all possible assessment scenarios. The quest for an ultimate panacea for all assessment problems, however, soon came to a dead end when I realized the complexity of translation quality assessment as a highly sophisticated decision-making process in which the assessor plays a central role. Then, in the course of the critical analysis of translation theories and assessment models in the first chapter of this thesis, I made a quite interesting observation. I noticed that many of these theories and models have a tendency to minimize or even ignore the role of the human agent in the process of assessment, probably because of the concerns about the potential negative impacts of the subjectivity factor involved. These models have generally reduced the process of assessment to benchmarking the quality of translations against a set of fixed and predefined criteria, so that the task of the assessor is relegated to that of a passive gauge reader who merely measures the translation with the yardstick of the assessment model. Translation quality assessment is, nevertheless, far more complicated than a simple act of measurement; it is a hermeneutic process that requires the assessor’s agency and judgmental skills. Assessment models are expected to assist assessors in the process of decision making concerning the quality of translations, rather than making decisions in lieu of them. As I have repeatedly emphasized throughout this thesis, the agency of the assessor should never be ignored, because an assessment model alone, no matter how well-designed, is not capable of interpreting the textual data in numerous different contextual situations and passing judgments on the quality of translation. Therefore, instead of trying to eliminate the subjectivity factor related to the human agency, a translation assessment model should ideally try to understand and control it.

What makes the methodology proposed in this thesis innovative and rather distinct from its previous counterparts, perhaps, is that it takes the assessors’
perspectives and the contingencies of the assessment exercise itself into consideration and incorporates them into a new model of translation quality assessment. The new dynamic model treats translation quality assessment as a process of decision-making and breaks it down into six different steps during which the assessors are allowed to make critical decisions, without forcing predetermined choices on them. It gives the assessors absolute freedom at the time of decision-making, but at the same time holds them responsible for their decisions, which are, nevertheless, expected to be in line with the overall skopos of assessment. The assessment skopos is in fact one of the most decisive factors in this dynamic process-oriented model\(^\text{14}\), serving as a guideline that directs all the decisions made at different levels during the process of assessment.

The new model of assessment was successfully put into application during an experiment in this thesis to analyze the evaluative behaviors of the participant assessors, leading to some interesting results. For instance, by application of the model in the specific local context of Iran, it was revealed that the academic assessors in this country tend to rely excessively on comparing the formal properties of translations with those of the original texts as a mean to measure the quality of translations. It was also discovered that, compared with ordinary readers, academic assessors in Iran pay less attention to socio-cultural and ideological factors. There were also some general findings that were not exclusively related to the local context in which the model was applied. For instance, the assignment of different hierarchical gravity values by individual assessors to problems that fell under identical linguistic categories showed that the assessment models which are based on subtraction of a fixed point from the quality index for each problem belonging to a certain linguistic category are not so reliable. Although the specific socio-historical and educational background of

\(^{14}\) The TQA model proposed in this thesis treats translation as a product as separate from the process through which it is produced or the person who has produced it. However, the model is referred to as a process-oriented model, since it focuses on the behavior of assessors during the “process” of assessment.
the assessors participating in the experiment and the unique context in which they performed their assessments limits the generalizability of its results, it does not mean than the experiment cannot be replicated in different contexts involving different languages and cultures.

There is a wide variety of people who may benefit from the new assessment model. Assessors will be more aware of their role and responsibilities, and will have a broader understanding of the decisions they have to make during different stages of the process of assessment. The model could serve as a basis for establishing a recursive mutual feedback loop between assessors and researchers: by using the model, assessors can provide researchers with helpful information about their evaluative behaviors. In return, researchers can share the results of their experiments with assessors and remind them of their common pitfalls and shortcomings, helping them to optimize their evaluative skills. Also, assessors can offer researchers constructive feedbacks as to how satisfied they feel with the assessment model, so that, in the light of their feedback, the model may be revisited and improved. Due to certain restrictions in this thesis, such as shortage of time and limited accessibility to the participants, however, the feedback loop was only partially completed, without conducting a post-interview with the participants – an issue that can be addressed in future studies. Translation companies may also benefit from the model, as it could be used as a dynamic tool to assess the job of assessors. The assessment model could help translation companies to make sure that the decisions made by a given assessor in his/her assessment of a given translation are indeed in line with their overall policies and objectives. Finally, translation teachers and students could potentially benefit from the assessment model. It can give them an insight into how assessors evaluate the quality of translations in the real world. If translation trainees are told how their translations will be assessed by the potential assessors in the real world, they could in turn produce translations that are successful and meet the expectations of the evaluators in different situations.
Nevertheless, the assessment model put forward in this thesis has some limitations. Since the model heavily relies on the ability of assessors to identify problematic areas in the translated texts, where their expectations are frustrated, it is not capable of capturing the positive evaluative behaviors of the assessors, i.e., it cannot identify those areas in a translation where the expectations of the assessors are surpassed rather than frustrated. The identification of strong points of a translated text, where assessors’ expectations are surpassed, is not so easy especially when it comes to assessment of a single translated version of a text, because in this case there is no point of comparison. However, in cases where multiple translations of the same text are available, it might be possible for the researchers to capture the positive evaluative behaviors of the assessors by discovering the areas wherein the assessors believe a certain translation is superior to others.

Using a questionnaire as a means for collecting data about the evaluative behavior of the participants in this research had restrictions which led to certain challenges. Composing a questionnaire exactly in accordance with the theoretical model of translation quality assessment proposed was a painstakingly difficult task. Although the application of a web-based questionnaire tool (Qualtrics) made the access to geographically dispersed participants in the study much easier, since it was not originally designed for developing a translation quality assessment model, it was not fully accommodating for the technical requirements of such a model, and I had to use a lot of creativity to make the best of its limited features. This problem affected the overall functionality of the model to some extent, and unwarily increased the time required for the participants to complete the questionnaire. The problem can be resolved in future studies through collaboration between translation researchers and computer specialists in developing a specialized software assessment application, something that was not possible due to limited financial resources available for my PhD project.
Another challenge that I faced was the decision about the number of participants I could recruit in my research. I was aware of the fact that the number of participants could affect the ability to generalize final findings, and that the greater number of participants could increase the reliability of the research. However, considering the fact that participation in the study was completely voluntarily and (again due to limited resources available) no financial compensation was offered to the participants, it was indeed difficult to persuade those who were potentially good candidates for participating in the research to cooperate, especially in the case of university teachers who had really hectic schedules. From about one hundred university teachers who were sent invitation letters (emails) to participate in the study, many refused to take part and many others quit halfway through the survey; consequently, it took me about a year to find twenty academic assessors who continued their cooperation to the end. However, at the end of the day, I was lucky enough to have 40 participants who generously devoted several hours of their precious time to completing the questionnaire in my study.

The processing and the analyzing of textual data were also among the most cumbersome challenges in this thesis. I had to transform a huge amount of raw data contained in the completed questionnaires into analyzable information and aggregate them into tabular form. The processing of data happened to be extremely time-consuming, since it had to be done manually. It took me almost a whole month of hard work to single-handedly categorize and combine responses in the questionnaires into 241 pages of information in tabular format (See Appendix A). Similarly, a significant amount of time and effort also went into analyzing the processed data, which again had to be done manually in parts. The processing and analyzing of data could have been much easier had the study been conducted as a research group project rather than an individual PhD thesis, so that the workload and responsibilities could have been shared and a greater amount of textual data processed and analyzed, increasing the reliability and generalizability of findings from the study.
So far, much time and effort has been spent by numerous researchers on developing flawless models to solve the problem of translation quality assessment, and yet the problem has remained largely unresolved. But maybe we are not asking the right question and missing the point entirely here. Maybe the time has come to invest more on training skillful assessors rather than on developing assessment models alone. I thereby end this thesis by recommending further research projects which focus on the agency of assessors, and raising their awareness about their important role in the process of translation quality assessment.