

Developing Guidelines for Evaluating Students' Translation at English Translation Program in Iranian Universities: Using Experts Views through Delphi Method

Saeed Taki¹ - Hossein Heidari Tabrizi²

Abstract

The present study aimed at investigating the Iranian translation experts' opinions and attitudes on the translation evaluation methods that are currently practiced in the context of Iranian universities through applying the Delphi procedure. In fact, it focused on answering the major question of translation evaluation: What criteria should be included in a systematic objective scoring rubrics for evaluating students' academic translations and how much weight should be given to each criterion? In so doing, Through the Delphi procedure, the opinions and attitudes of ten Iranian translation experts were collected, categorized, revised, and finalized by their responses and feedbacks on a two-round questionnaire-based Delphi procedure for the purpose of establishing some minimum areas of agreement on the student translation evaluation on final academic tests in the Iranian context. The research findings revealed that the Delphi panelists believed a shift should be made towards more direct, performance-based methods of testing and evaluation with essay-type tasks rather than recognition multiple-choice items. As part of the solution, to improve the reliability and validity of such tests, the present summative, product-oriented evaluation should be accompanied with some formative, process-oriented methods of evaluation. They also argued for a multidimensional scoring method in which the linguistic microstructures play a relatively minor role in comparison to more socio-pragmatic, functional macrostructures at discourse level.

Keywords: Evaluating, Translation, Translation Program in Iranian Delphi Method

¹ - Islamic Azad University, Shahreza Branch taki_saeed@yahoo.com

² - Islamic Azad University, Khorasgan Branch atbinus@yahoo.com

1. Introduction

Translator and interpreter education is now widely practiced around the world and Iran is no exception. During the last decade or so, the number of Iranian universities offering the "Academic Translation Program" as well as the number of the candidates entering such programs has been increasing. Whereas in mid 1360's/1980's only one university in Iran (namely, Allameh Tabatabai University in Tehran) offered this program at BA level, the number of Iranian universities offering just the "English Translation Program" at MA and BA level now exceeds to 130 accepting a total of more than 7500 students into the program.

Along with such a drastic increase in quantity, attempts have been made to improve the quality of translation programs too. Now an increasingly sophisticated body of research and knowledge is available on various aspects of translation training programs including pedagogical/educational programs, curriculum and materials development, teacher training, translation technology and translation evaluation, among which the often-neglected teacher evaluation of trainee translations has received the least attention. In other words, while translation evaluation is of central concern and significance in the context of translator training, it is, as observed by many a translation scholar (Arango-Keeth and Koby, 2003 or Bowker, 2000) to name just a few, one which is, despite being a common practice, "under-researched and under-discussed" (Hatim & Mason, 1997, p.197). In the debates on the subject of the assessment of translations in a Round-table discussion on translation in the New Millennium, McAlester (2003) even goes further: "this is an area in which Translation Studies has its worst failure" (p. 45). In sum, it can safely be concluded that in comparison "little is published on the ubiquitous activity of [translation] testing and evaluation" (Hatim & Mason, 1997, p. 197).

In an academic setting, evaluating translations is even much more daunting because a translator trainer has an obligation to help students improve their performance. Needless to say, every teacher of translation has an academic obligation to rank his/her students' work. In fact, translation teachers are said to play two major and simultaneous functions: they are both facilitators of learning and evaluators of what has been learnt. Thus, in training translators, judging the translation quality "should not be an end but a means" (Honig, 1998b, p.32). On the whole, translation evaluation is undoubtedly one of the most difficult tasks facing a translator trainer: the problem of evaluation and decision-making

Developing Guidelines for Evaluating Students' Translation

in translation. It is unlikely that there will ever be a ready-made formula that will transform this task into a simple one; however, attempts have been made to investigate this issue from different perspectives (for example, Williams, 1989, 2001, 2004; Waddington, 1999, 2000a, 2000b, 2001; House, 1981, 1997, 2001a, 2001b; Hatim & Mason, 1990, 1997; Sainz, 1994; Schaffner, 1998a, 1998b, 1998c; Schiaffino & Zearo, 2001, 2002, 2003, 2005; and Honig, 1998a, 1998b). Such attempts, to the best knowledge of the researchers, are rare in the academic Iranian context.

Teaching translation involves assessing the quality of the translations produced by students and giving a grade for the achievement of the intended goal; i.e., the instructional objectives. In fact, translation evaluation through quality assessment is an integral part of the career of every translation teacher. There are always mid-term and final tests as well as other more formative diagnostic assessments done for pedagogical purposes in the academic centers.

Having been teaching different courses of translation at a number of universities in Iran for more than ten years, the researchers themselves must confess that the reliability, validity and even in some cases the practicality of such tests as well as the way they are graded are under serious question. In most cases, translation students do not know on what criteria their work will be evaluated. Even much worse, some teachers and lecturers blithely substitute the authority of their position for any awareness of the complexity of the evaluative situations. The results are disastrous: students feel that the evaluation of their translations is done on the basis of arbitrary, subjective practices; they spend most of their energy adapting themselves to the personal non-objective criteria of their teachers and feel that it is a waste of time to gain insights into the nature of translation processes as provided by translation theories; consequently, they lack the self-awareness as well as the self-confidence they need to carry out translation tasks when they are on their own in the real — and sometimes confusing — world of translations.

It seemed to the researchers, at least based on common sense and experience, that translation teachers of Iranian universities are least informed and familiar, if at all, with the current translation evaluation approaches and methods in the field of translator training. This is in line with Honig (1998b), "Obviously, many teachers and lecturers are not aware of the fact that there is such a wide variety of evaluation scenarios and applied criteria (p.29)." Likewise, Newmark (2003, p.65) asserts that "... examination boards and examiners are not aware of the literature."

Proceedings of The Research in Language Science

In short, it could possibly be claimed that the dominant trend for evaluating translation quality in academic settings in Iran is far behind the modern ones practiced in accredited universities throughout the world. One piece of evidence can be the frequent negative feedbacks teachers are likely to receive from the students about the final tests of translation in every semester. Still another piece of supporting evidence is the countless anecdotes one hears in professional conferences about the deficiencies of translation tests. Again, Honig (1998b, p. 29) argues that "The least homogenous TQA criteria are assembled in university training course. The students feel that TQA is subjective and arbitrary, they try to adapt to the standards of teachers and they acquire neither self-awareness nor self-confidence."

2. Objective of the Study

The present study, therefore, aimed at investigating the Iranian translation experts' opinions and attitudes on the translation evaluation methods that are currently practiced in the context of Iranian universities through applying the Delphi procedure. In fact, it focused on answering the major question of translation evaluation: What criteria should be included in a systematic objective scoring rubrics for evaluating students' academic translations and how much weight should be given to each criterion?

3. Methods

3.1. Participants

To determine the prospective panelists for participating in the Delphi rounds of the present research study, first a listing of thirty male and female experts in the field of translation and translation studies including translation theorists, professional editors, translation researchers, competent writers, and translation-oriented applied linguists was prepared. The criterion for their inclusion was above all their well-known expertise on the issue of translation: they were selected since at first place they were recognized in the academic setting not as translation teachers but translation experts. Thus, efforts were made to select the respondents based on a rigorous scale of background education and academic performance.

In so doing, the researchers first reviewed the existing literature on translation evaluation and assessment including books, journals, articles, proceedings, electronic materials, homepages, web logs, and the like

Developing Guidelines for Evaluating Students' Translation

which resulted in a rather comprehensive list of influential key figures in the field continuously writing and working on assessing and evaluating translation. Next, the researchers tried to get access to their resumes (CVs) to make sure that they can safely be considered 'experts' on the subject matter in question. Moreover, those who could be experts in translation studies or language testing in general but not in translation testing in particular were excluded. Then, the researchers did their best to contact the people still on the list to ask them to honor the present study. In fact, one significant criterion in their final inclusion into the study was their willingness and sense of commitment to do so. In this way, the researchers guaranteed that the sample included just committed experts proper. As the last step, applying the snowball sampling method, the researchers used the prospective panelists identified as key informants to introduce others who, they thought, could qualify for inclusion in the Delphi panel. In this way, the researchers did their best to avoid leaving out any possible experts in the field.

Furthermore, on the recommendation of the first advisor professor of the present dissertation and approval of the dissertation committee, it was decided that at least ten non-Iranian translation experts should participate the Delphi rounds too so that it would be possible to compare and contrast their views with those of the Iranian experts. Yet, in practice, when the questionnaire of the first round of the Delphi procedure was sent to these non-Iranian scholars, they preferred to refer the researchers to their published materials on translation evaluation. Accordingly, such experts were excluded from the study.

However, to include in the questionnaire as much information as possible on the experts' attitudes towards translation evaluation in academic contexts, the researchers were compelled to extract the required information through critically reviewing the related literature. In addition to consulting the relevant articles or textbooks in this regard, the researchers decided to work rather on the reports of the round-tables, interviews, conference panels and the like where opinions are more likely to be expressed frankly and directly. The sources examined for this purpose were the scripts of a good number of experts' "Round-table Discussion on Translation in the New Millennium" held in 1999 in the University of Surrey published in Anderman and Rogers (2003, pp.13-67); Chesterman and Wagner's (2002, esp. pp. 80-107) compilation of experts' dialogues on translation theory and practice as well as Secara's (2005) state of the art survey, Williams (2004) and Waddington (2001).

Thus, in practice from among the thirty prospective panelists, ten Iranian translation experts actually participated in the rounds of the Delphi method by returning the round one questionnaire; in effect, making a 33 percent return rate for the first round. This is in line with Murry and Hammons' (1995) suggestion that the final panel should include a minimum of ten members. It should also be mentioned that the Delphi questionnaires were either distributed by e/mails or submitted in person.

3.2. Instruments

In this study, the Delphi procedure was used as the major research method to collect the required data. This research method is defined by Riazi (1999) as a method of data collection predominantly employed in qualitative research as a multi-phase ethnographic approach. The procedure uses questionnaires (as well as interviews) while ensuring that the anonymity of the participants so that nobody 'looses face'. In sum, the purpose of a Delphi study is to obtain the most reliable consensus of opinion of a group of experts by applying a series of intensive questionnaires interspersed with controlled opinion feedback.

As Doyle (1993) explains, the ultimate goal is to increase consensus among those already highly knowledgeable in one particular area, a key characteristic justifying its application in this very research study since "translation quality assessment and judgment of translations are also matters of communication, co-operation and consent" (Lauscher 2000, p.164). Likewise, Newmark (2003) argues that considering the increased number of Schools of Translation originated from the increase in types and quantities of translation throughout the world, "it is not helpful to continuously leave the subject of translation assessment to isolated individuals ... [W]hat is required are ... conferences ... for the purpose of establishing some minimum areas of agreement on the assessment of exams" (p.65) as well as "some kind of guidelines" (Mcalester, 2003, p. 46). The Delphi procedure best fits these goals. Moreover, the Delphi method is generally recommended where the objectives of the study are to elicit subjective opinions and attitudes from a group of highly knowledgeable people who are geographically separated and where precise statistical analysis is not applicable because of the nature of the study. In fact, the Delphi approach adds to the reliability of group decision by avoiding the 'bandwagon effect' and 'deference to authority' which are typical problems with face-to-face meetings (Murry & Hammons, 1995).

Developing Guidelines for Evaluating Students' Translation

The criterion for inclusion of the participants in the Delphi process is their relative expertise, not their representativeness of a target population. Delphi studies are processes that include the preparation, a structural survey in two or more rounds and some analyses and application (implementation) when the survey is finished. In fact, the Delphi method consists of two or more stages: The first round, an open-ended questionnaire, "amounts to an anonymous brainstorming on the part of the experts" (Murry and Hammons, 1995, p.424); in the second stage, the revised, now-structured questionnaires are given to the same experts who are asked to consider, rate, rank and comment on the responses developed during round one. Such a rating is typically done on a five-point Likert-scale type (1 for "Strongly disagree" while 5 indicating "Strongly agree"). In this situation, the researcher is actually "both a researcher and a moderator, acting as a go-between as ideas are shared, modified and 'debated'" (Pickard, 2007, p. 129).

As Murry and Hammons (1995, p.425) emphasize, "In higher education, the Delphi method has been used primarily for four purposes: (1) to develop goals and objectives, (2) to improve curriculum, (3) to assist in strategic planning, and (4) to develop criteria." The present study falls under the scope of the forth category; i.e. it is an attempt to determine the criteria, which the Iranian experts of translation recommend or believe translation university teachers use, for evaluating trainees translation. Moreover, the application of the Delphi procedure in this research revealed the often overlooked and unique aspects of this versatile powerful qualitative research methodology to Iranian investigators. The Delphi method offers several advantages:

- Group decision-making using anonymous controlled-feedback procedures are often more accurate than face-to-face discussions.
- Opinions using the Delphi method can be received from a group of experts who may be geographically separated from one another.
- Consensus reached by the group reflects reasoned opinions because the Delphi process forces group members to logically consider the problem under study and provide written responses.
- Group responses can be described statistically. (Murry and Hammons, 1995, p.426)

As for the optimal size of a Delphi panel, though there is no evident agreement found in the literature on, Murry and Hammons (1995) suggest that the final panel should include a minimum of ten members.

Proceedings of The Research in Language Science

Moreover, since within homogeneous groups few new ideas are generated when the panel size exceeds 30 (or even 25) well-chosen participants, the upper limit on the panel size is 30. In this very research, due to the limited number of experts in the field, the panel size was set on 10.

In practice, an open-ended questionnaire on translation evaluation theories and practices was first needed to conduct the initial round of the Delphi method. In fact, this format allowed the panelists to provide free responses using their own words to generate ideas. To prepare reliable valid essay-type questions of significant value, the researchers made a list of topics related to the purposes of the study. In so doing, again the extensive review of the related literature was very helpful. In fact, the researchers especially focused on the question raised in the discussion panels, round-tables and idea exchanges at different translation seminars, conferences and symposium held in the new millennium around the world. The starting point was the four major issues addressed in the one-day workshop held on March 22, 2003 at Aston University:

- What is the relationship between assessment strategies and theoretical frameworks for teaching translation?
- Which criteria are used for assessing and scoring translations in a university program?
- Do these vary according to the purposes of the assessment?
- How do these criteria compare to translation quality assurance in the real world?

After comparing the above-mentioned questions with those posed in ATA annual conferences, in Antwerp's 2005 workshop for the Flemish translation day, in Secara (2005), in Williams (2004), in Anderman and Rogers (2003), in Chesterman and Wagner (2002), or in Waddington (2001), the researchers came up with a question checklist of ten items. However, it is suggested that the essay-type items of the Delphi's first-round questionnaire should be to-the-point, precise, and limited in number to guarantee its practicality (Murry & Hammons, 1995; Doyle, 1993). Hence, the researchers limited the questions to three major essay-type open-ended unstructured ones of general nature produced in English as follow:

- Which major criteria do you recommend for scoring and evaluating student translations on final tests in a university program? Why do you apply such criteria but not others?

Developing Guidelines for Evaluating Students' Translation

- What test type(s) or test item format(s) do you prefer as final translation tests to evaluate student translations?
- What do you think about how to establish the quality of such tests in terms of selecting materials, deciding on the test length, adjusting the item difficulty, writing instruction, allocating the amount of time needed, and allowing dictionaries?

The questionnaire also asked the respondents to edit or reword any statement they felt needed revision and to provide additional comments if they wish so. As the next step, two experienced teachers of translation were invited to pilot-test the initial questionnaire. After they completed the questionnaire, critiqued its format, content, and clarity, and made their suggestions for improvements, the finalized round one questionnaire was formed incorporating a number of changes in diction and style. The questionnaire was then sent to fifty translation experts. Moreover, to encourage the prospective panelists to respond to the questionnaire, a personalized formally-written cover letter accompanied the questionnaire as an invitation. Introducing the researchers as well as establishing their own credibility and status in the field, the letter explained the problem at issue, provided the participant with necessary but brief information about the research study in general and the Delphi rounds in particular, explained the time it required, and promised the panelist to be provided with a summary of the research findings at the end of the study. Returning the round one questionnaire indicated their agreement to be included in the Delphi process.

After two months or so, at last out of twenty ones sent, ten questionnaires completed by members of 'Group A' were returned, of course, with some follow-up contacts the researchers made with a couple of the respondents who failed to return the questionnaire by the specified return date. As each response was received, each respondent was assigned a number for identification as responses were compiled. To increase objectivity, this number was used to identification on responses, so anonymity was ensured. In addition, the compilation of the responses included as much of the original wordings as possible. As the number of reactions to the questionnaire items increased, concepts shared by more than one panelist were simply counted while unique ones were noted. The concepts most often mentioned by individuals were collapsed into categories so that the number of items respondents needed to evaluate in the second round was reduced. By intentional processing, the common themes were identified and conceptualized from many different

Proceedings of The Research in Language Science

perspectives, forming phrases for the round two questionnaire. Considering the respondents' time and the practicality issues, however, not all of the phrased responses were included in the final version of the second round questionnaire. Actually, only those phrases which were significant opinions frequently mentioned by the panelists were included in a precise, concise and comprehensive way in the second questionnaire.

In sum, based on these analyses of the statistical results and panelists' responses and comments, the round two questionnaire was developed. In preparing this second questionnaire, the researchers followed the principles proposed by Dornyei (2003) as well as the guidelines recommended by Cohen, Manion, and Morrison (2000) regarding the format, content, organization, sequencing, attractiveness, and comprehensibility of the instrument. To encourage the panelists to respond to the questionnaire, again a personalized formally-written cover letter was prepared to accompany the questionnaire as an invitation. The letter was developed in order to inform the Delphi panelists about the purpose of this second questionnaire, thanking them for reacting to the round one questionnaire. The respondents, ensured about the confidentiality and anonymity of the data gathered by this instrument, were invited to honor once again the present study with their meticulous reactions. As for the item wording of the questionnaire itself, the researchers did their best to aim for short yet precise meaningful items written in simple and natural English language avoiding ambiguous or loaded words, negative constructions, leading statements, and double-barreled items.

Thus, in practice, the questionnaire started with an attractive title followed by informative and well-pitched instructions printed in italics. Apart from its introductory section on personal information (i. e., age, sex and degree), the questionnaire was actually divided into three major parts. The first section encompassed 20 topic items on the value or weight the respondent believed should be assigned to the criteria for evaluating academic translation. To these, the respondent reacted on a six-point Likert-type response options, given just next to the questions (and not below), from 1 (zero weight) to 6 (full weight). The second part consisted of 30 statement items on the reliability, validity and feasibility of final tests on academic translation. Again, the response options, given just next to the questions (and not below), were in the format of a six-point Likert-type scale from 1 (Strongly Agree) to 6 (Strongly Disagree). The last part was just one rank order item on which test formats are the

Developing Guidelines for Evaluating Students' Translation

most appropriate ones for evaluating academic translation. As a rule of thumb in basic courtesy, the respondents were thanked once again for their cooperation at the very end of the questionnaire too. Thus, the questionnaire was organized in three one-sided A4 pages..

Finally, the then-structured questionnaire was sent only to the same sample of participants who had returned the first questionnaire. They were kindly asked to comment on the responses proposed in the first round by rating them on a six-point Likert scale as to agreement-disagreement, following Dornyei's (2003) recommendation. Although the researchers had a very hard time to write some follow-up letters, make some phone calls, and even arrange to visit the participants in person, all participants returned their questionnaires. This round took approximately four months to complete.

Moreover, as stated before, the ultimate goal of the Delphi study is to reach consensus having allowed the panel maximum opportunity to contribute their thoughts and ideas." As such, "at least two rounds' is an absolute minimum; three or more is more common" (Pickard, 2007, p. 129). Accordingly, the researchers was actually ready to conduct still one or even two more rounds of questionnaires if needed, but since "the Delphi procedure stops after either consensus or stability of responses has been achieved" (Murry & Hammons, 1995, p.429), there was no need to do so as explained below.

4. Research Findings

Since the number of the panelists was limited to ten experts, based on the assumption(s) required it was not justified to run a non-parametric Chi-square test or any other statistical procedure. Thus, the data analysis in this part includes just the percentages observed. To make it easy-to-follow, the very sequence applied in the questionnaire is observed here in reporting the findings and results of the analysis. As it was stated before, the Delphi round-two questionnaire was actually divided into three major parts. To refresh the reader's mind, the questionnaire encompassed 20 topic items followed by 30 statement ones and a single rank item at the end.

4.2.1.1. The Topic Items

The first part of the questionnaire consisted of 20 topic items on the value or weight the panelists believed should be assigned to the different

criteria for evaluating academic translation. Table 4.1 summarizes the responses elicited on each of these items by the Delphi panelist:

Table 4.1: Descriptive statistics of the topic items of the Delphi questionnaire

PHRASES	Zero 0%	Least 20%	Less 40%	Much 60%	Heavy 80%	Full 100%
1. Grammatical Points				4	5	1
2. Lexical Equivalence				2	3	5
3. Register				6	3	1
4. Genre				6	3	1
5. Cohesion				2	3	5
6. Coherence				1	4	6
7. Beauty	3	4	3			
8. Accuracy/Loyalty				4	2	4
9. Fluency/Meaningfulness				2	4	4
10. Naturalness			2	3	4	1
11. Equivalence Effect				3	3	4
12. Mechanics of Writing		1	1	3	3	2
13. Culture			2	3	3	2
14. Style	2	3	3	2		
15. Purpose/Function				3	3	4
16. Target Language Norms			2	2	3	3
17. Item-Test Format				3	3	4
18. Text Type				3	4	3
19. Instructor's own Translation	3	4	3			
20. Error Types			2	3	2	3

The data presented in this table immediately reveals that there is a general consensus among the Iranian translation experts as the Delphi panel of the present study that the model for evaluating academic translation must be a sophisticated multidimensional one. To them, such a model must include in itself a variety of criteria at both macro- and micro-level. In other words, in their opinion there are a great number of factors involved in a 'good' translation all of which must be observed in an adequate model for evaluating academic translation. In practice, the

Developing Guidelines for Evaluating Students' Translation

panelists unanimously marked only the right side ('much/heavy/full weight') of the Likert scale for 12 factors displaying the high importance they believe must be allocated to these criteria in a translation evaluation scale. Among these, coherence, cohesion and lexical equivalence are regarded by the panelists as the most important ones which should receive full weight. For the criteria presented in items 10, 12, 13 and 20, majority of the experts believed that they should also be given great weight: most panelists (80%) have marked the left side and only two members believed that these criteria are of less value in a translation evaluation model. On the contrary, they marked only the left side ('less/least/zero weight') for just two factors; namely 'beauty' and 'instructor's own translation', showing that they believe these two factors must be assigned very low value, if any at all, in translation evaluation models. The same can almost be said about 'style' another criterion included in the questionnaire as item 14. In fact, just two members of the Delphi panel believed that this criterion is of much value in a translation evaluation model while most panelists (80%) marking the left side of the Likert scale.

4.2.1.2. The Statement Items

The second part of the instrument consisted of 30 statement items regarding the reliability, validity and practicality of final tests on academic translation. Table 4.2 presents the way the Delphi panelists responded to each item:

Table 4.2: Descriptive statistics of the statement items of the Delphi questionnaire

STATEMENTS	Strongly Agree	Agree	Partially Agree	Partially Disagree	Disagree	Strongly Disagree
21. The instructor's own translation must be used as the criterion for scoring student translations in the final tests.				7	2	1
22. To me, academic translation can be assessed objectively.	5	3	2			
23. Students should be informed of the evaluation criteria before test.	5	3	2			
24. Students should be informed of the evaluation criteria during the tests through explicit instructions.	4	3	3			
25. I believe test items of a translation test should be contextualized.	3	4	3			
26. I believe Limited-Response items are appropriate for translation tests.				2	3	5
27. I believe Free-Response items are appropriate for translation tests.	8	1	1			
28. To me, Multiple-Choice items are appropriate for translation tests.	3	4	3			
29. To me, Recognition items are appropriate for translation tests.	2	3	5			
30. To me, Reaction items are appropriate for translation tests.	2	4	4			
31. To me, Completion items are appropriate for translation tests.				4	4	2
32. To me, Essay-type items are appropriate for translation tests.	6	3	1			
33. To me, Cloze Tests are appropriate for translation tests.				4	3	3
34. To me, True-False items are appropriate for translation tests.					1	9

Developing Guidelines for Evaluating Students' Translation

Table 4.2: Continued

35. The test should contain some items on translation theory.	2	3	3	2		
36. The required test time should be allotted based on the length of the translation tasks.	3	3	2	2		
37. The required test time should be allotted based on the time that instructors themselves spend to translate the test tasks.	2	3	3	2		
38. "Unseen Texts" are appropriate for evaluating testees' translation ability.	3	4	3			
39. "Seen Texts" actually evaluate testees' memory rather than their translation ability.	3	4	3			
40. "Seen Texts" are appropriate for evaluating testees' translation ability.				4	4	2
41. "Readability Index" of a translation task in a test is an appropriate criterion for adjusting its difficulty level.			1	4	3	2
42. The topic of a translation task in a test is an appropriate criterion for adjusting its difficulty level.	2	3	5			
43. The length of a translation task in a test is an appropriate criterion for adjusting its difficulty level.	1	3	4	2		
44. "Analytic Scoring" is the appropriate method for scoring testees' translations.	5	3	2			
45. "Holistic Scoring" is the appropriate method for scoring testees' translations.	2	4	4			
46. The "penalty" system is an appropriate way of scoring testees' translations.				3	3	4
47. Information resources such as dictionaries or glossaries must be allowed in translation tests.	6	2	2			
48. "Performance-based" tests are more appropriate for translation evaluation than "paper-and-pencil" tests.	5	4	1			
49. Translation must be evaluated through indirect NOT direct testing.				2	5	3
50. "Production" tests are better indicators of student academic translation ability than "recognition" tests.	4	4	2			

Proceedings of The Research in Language Science

Again, as the table shows, in majority of the cases, the respondents marked only one of the either sides of the Likert scale. In facts, almost a general consensus exists among the panelist members on all items of the questionnaire. In other words, they either show their agreement with an item unanimously or disagree with it by consensus: that is the case for 25 out of 30 items (over 83%). Besides, in the cases where there was not a complete consensus among the respondents, still the majority of the responses were gathered on one side of the scale.

As such, the Delphi panelists showed their relatively full agreement over 18 statements (60%) with items 27, 21, 32 and 47 at the top respectively. As for items, 35, 36, 37 and 43, 80% of the panelists marked the right side of the scale while just two members partially disagreed with these statements. On the other hand, the respondents unanimously disagreed with 7 statements (over 23%) as the most disagreement found with items 34, 26 and 46 respectively. Again, as for item 41, just one of the panelist members partially agreed with the statement while the other nine members conveyed their disagreement.

Accordingly, the panelists showed their strong belief that academic translation can be assessed objectively. They believed that test items of a translation test should be contextualized while most argued for some items on knowledge of translation theory as well. They were strongly of the opinion that students' translation ability can best be tested and evaluated by response-free items especially essay-type format while other test methods such as multiple-choice questions, reaction as well as recognition items were considered suitable for this purpose too. On the contrary, they rejected limited-response items such as completion items, cloze procedure and especially true/false items as inappropriate methods for evaluating academic translation. To them, "seen texts" are not suitable for evaluating student translation since they actually evaluate testees' memory rather than their translation ability; they preferred "unseen texts" instead.

The Delphi panelists held the belief that the topic and the length of a translation task in a test can be appropriate criteria for adjusting its difficulty level whereas the "Readability Index" of a translation task is not. Moreover, according to the panelists, the amount of time required should be allotted based both on the length of the translation tasks and on the time that instructors themselves spend to translate the test tasks. They believed that information resources such as dictionaries or glossaries must be allowed in translation tests. While both "analytic" scoring and

Developing Guidelines for Evaluating Students' Translation

“holistic” scoring were regarded as appropriate methods for scoring testees’ translations, “analytic” scoring was more strongly supported by the panelists.

In addition, they supported the idea that students should be informed of the evaluation criteria before the tests and during the tests through explicit instructions. They also considered the “penalty” system as an inappropriate way of scoring testees’ translations. Besides, the instructor’s own translation must not be used as the criterion for scoring student translations in the final tests. Finally, the panelists emphasized the point that translation ability can be evaluated more appropriately through performance-based direct testing procedures with production tasks rather than by classical paper-and-pencil tests with multiple-choice recognition items.

4.2.1.3. The Rank Item

Finally, there was just one rank order item asking the Delphi panelist which test formats, they believe, are the most appropriate ones for evaluating translation in academic contexts. The responses elicited are presented in Table 4.3:

Table 4.3: Descriptive statistics of the ranking item of the Delphi questionnaire

Format Rank	Essay Type	Classical Cloze	Completion	Multiple Choice	MC Cloze	Reaction	Recognition	T/F	Total
One	7	--	--	2	--	1	--	--	10
Two	5	--	--	3	--	2	--	--	10
Three	--	--	--	4	--	3	3	--	10
Total	12	--	--	9	--	6	3	--	30

As the table shows, , seven respondents ranked the essay-type format as their top priority; that is to say, almost the majority of the Delphi panelists (70%) believed that the most appropriate method for testing and evaluating student translations in academic contexts is the essay-type format. This format has also received the highest frequency as the second priority having been marked by five members of the panel. In total, this test format was chosen 12 times, the most frequently selected one. Multiple-choice items, reaction items and recognition ones stand in the 2nd, 3rd and 4th ranks respectively. Interestingly enough, the panelists unanimously believed that classical cloze procedure, multiple-choice cloze items, completion items and true/false ones are by no means suitable for translation evaluation.

5. Discussion & Conclusion

As the Delphi procedure revealed, translation experts believe that the length, the topic, the diction and the linguistic (structural) complexity of the texts can be useful in determining their difficulty level. They reject the adequacy of readability formulas in this regard, however. It was also recommended by the Delphi panelists that the amount of time required should be allocated according to the length of the translation tasks and the time the teachers themselves spend to translate the test texts.

In addition, the Delphi experts, like other translation experts (e.g. Newmark, 1988, p. 221), argue that to guarantee the authenticity of translation job & to avoid artificiality, in all translation tests students must be allowed to use dictionaries during their test since they can always consult human/non-human resources & references especially a dictionary during translating a text in their normal career as a translator. The Delphi panelists proposed, student performance on translation tasks of final tests may be scored and evaluated better through production performance-based tests with essay-type tasks rather than recognition multiple-choice items.

Moreover, the reactions the Delphi panelists made to the second round questionnaire showed that consistent with fashion of the time, Iranian translation experts believe that the focus of attention should shift from mere linguistic, text-oriented factors to more socio-pragmatic macrostructures as far as possible criteria for translation rating schemes are concerned. In other words, as for the test instructions and directions, the findings of the Delphi procedure indicated that the Iranian translation experts, like those in the field of language testing, believe that the testees must be informed of how to perform the translation tasks and how their performance is to be evaluated and scored. They recommended that to do so, explicit written test instructions be developed for the tests. They believe that the scoring method should include in itself elements relating to more macrostructures of the texts as well. They argue for a multidimensional scoring method in which the linguistic microstructures play a relatively minor role in comparison to more socio-pragmatic, functional macrostructures at discourse level.

The Delphi procedure presented a bell-jar summary of common-core factors that, the panelists believe, must be used in evaluating a translation in academic environments. Thus, the potential translation evaluation scheme for the Iranian context may consist of a list of criteria composed of a manageable number of items selected based on the ideas the Iranian

Developing Guidelines for Evaluating Students' Translation

translation experts, as the Delphi panelists of the present research, suggested as well as those proposed by other scholars in the literature to date. It is more appropriate to present them as general criteria so that they can be applicable to any translation tests regardless of their subject matter. Besides this very checklist, to guarantee its flexibility, the scheme should provide enough room for the teacher-evaluator to determine the significance of each criterion before the application of the instrument. In so doing, a column can be added to the left side of the list where the teacher-evaluator can put a double checkmark for "highly significant" criteria, a single checkmark for "fairly significant" ones, a single cross or double cross for "not significant" or "totally in significant" or even a question mark when s/he is "not sure". Moreover, another column can be put to the right side of the list where the teacher-evaluator puts the ratings based on a Likert-like scale from 1 to 5 to arrive at a final decision.

References

- Anderman, G. M. & Rogers, M. (Eds.). (2003). *Translation today: Trends and perspectives*. Clevedon: Multilingual Matters Limited.
- Arango-Keeth, F. & Koby, G. S. (2003). Assessing assessment: Translator training evaluation and the needs of industry quality assessment. In B. J. Baer. (Ed.), *Beyond the ivory tower: Rethinking translation pedagogy* (pp.117-134). Amsterdam/ Philadelphia: John Benjamins Publishing Company.
- Bowker, L. (2000). A corpus-based approach to evaluating student translations. *The Translator*, 6 (2), 183– 210.
- Chesterman, A. & Wagner, E. (2002). *Can theory help translators: A dialogue between the ivory tower and the wordface*. Manchester: St. Jerome Publishing.
- Cohen, L. Manion, L. & Morrison, K. (2000). *Research methods in education* (5th ed.). London: Routledge.
- Dornyei, Z. (2003). *Questionnaires in second language research: Construction, administration, & processing*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Doyle, C. (1993). The Delphi method as a qualitative assessment tool for development of outcome measures for information literacy. *SLMA*, 11, 132-144.
- Durieux, C. (1998). Translation Quality Assessment. *Seventh Seminar on Translation Theory and Applications, United Nations Office at Vienna*.
- Hatim, B. & Mason, I. (1990). *Discourse and the translator*. London & New York: Longman Ltd.

Proceedings of The Research in Language Science

- Hatim, B. & Mason, I. (1997). *The Translator as communicator*. London: Routledge.
- Honig, H. (1998a). Complexity, contrastive linguistics and translator training: Comments and responses. In C. Schaffner (Ed.), *Translation and quality* (pp.83-89). Clevedon: Multilingual Matters Limited.
- Honig, H. (1998b). Positions, power and practice: Functionalist approaches and translation quality assessment. In C. Schaffner (Ed.), *Translation and quality* (pp.6-34). Clevedon: Multilingual Matters Limited.
- House, J. (1981). *A Model for translation quality assessment* (2nd ed.). Tübingen: Gunter Narr.
- House, J. (1997). *Translation quality assessment: A model revisited*. Tübingen: Gunter Narr.
- House, J. (2001a). How do we know when a translation is good? In E. Steiner and C. Yallop (Eds.), *Exploring translation and multilingual text production: Beyond content* (pp. 127-160). Berlin: Mouton de Gruyter.
- House, J. (2001b). Translation quality assessment: Linguistic description versus social evaluation. *Meta*, 46(2), 243-257.
- Lauscher, S. (2000). Translation quality assessment: Where can theory and practice meet? *The Translator*, 6(2), 149-168.
- McAlester, G. (2003). Comments in the 'Round-table discussion on translation in the New Millennium'. In G. M. Anderman, & M. Rogers, (Eds.). *Translation today: Trends and perspectives* (pp. 13-51). Clevedon: Multilingual Matters Limited.
- Murry, J. W. & Hammons, J. O. (1995, summer). Delphi: A versatile methodology for conducting qualitative research. *The Review of Higher Education*, 18(4), 423-436.
- Newmark, P. (1988). *A textbook of translation*. New York: Prentice-Hall.
- Newmark, P. (2003). No global communication without translation. In G. M. Anderman & M. Rogers (Eds.), *Translation today: Trends and perspectives* (pp.55-67). Clevedon: Multilingual Matters Limited.
- Pickard, A. J. (2007). *Research methods in information*. London: Facet Publishing.
- Riazi, A.M. (1999). *A dictionary of research methods: Quantitative & qualitative*. Tehran: Rahnama Publication.
- Sainz, M. (1994). Student-centred correction of translations. In C. Dollerup and A. Lindegaard (Eds.), *Teaching translation and interpreting 2: Insights, aims, visions; [Selection of] Papers from the Second Language International Conference* (pp. 133– 141). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Schaffner, C. (1998a). Introduction: From 'good' to 'functionally appropriate': Assessing translation quality. In C. Schaffner (Ed.), *Translation and quality* (pp.1-5). Clevedon: Multilingual Matters Limited.

Developing Guidelines for Evaluating Students' Translation

- Schaffner, C. (1998b). Qualifications for professional translators: Translation in language teaching versus teaching translation. In K. Malmjaer (Ed.), *Translation and language teaching, language teaching and translation* (pp. 117– 133). Manchester: St. Jerome Publishing.
- Schaffner, C. (Ed.). (1998c). *Translation and quality*. London: Routledge.
- Schiaffino, R., & Zearo, F. (2001). *The Measurement of quality in translation*. [Presentation]. *42nd Annual Conference of the American Translators Association, Los Angeles, CA*.
- Schiaffino, R., & Zearo, F. (2002). The measurement of quality in translation using statistical methods. [Presentation]. *43rd Annual Conference of the American Translators Association, Atlanta*.
- Schiaffino, R., & Zearo, F. (2003). Measuring quality in translation: The translation quality index and other methods. [Presentation]. *44th Annual Conference of the American Translators Association, Phoenix*.
- Schiaffino, R., & Zearo, F. (2005). Translation quality measurement in practice. [Presentation]. *46th Annual Conference of the American Translators Association, Seattle*.
- Secară, A. (2005). Translation evaluation - A state of the art survey. *Paper presented in eCoLoRe / MeLLANGE Workshop*. Retrieved March, 25, 2007 from <http://www.leeds.ac.uk/cts/research/publications/leeds-cts-2005-03-secara.pdf>.
- Waddington, C. (1999). *Estudio comparativo de diferentes métodos de evaluación de traducción general (Inglés-Español)*. Madrid: Publicaciones de la Universidad Pontificia Comillas.
- Waddington, C. (2000a). *Measuring the effect of errors on translation quality*. Paper presented at the Saarbrücker Symposium on Translation and Interpretation: Models in Quality Assessment, held at the Universität des Saarlandes 9th–11th March 2000.
- Waddington, C. (2000b). Should student translations be assessed holistically or through error analysis? *Hermes, Journal of Linguistics*, 26.
- Waddington, C. (2001). Different methods of evaluating student translations: The question of validity. *Meta*, 46(2), 311-325.
- Williams, M. (1989). Creating credibility out of chaos: The assessment of translation quality. *TTR*, 2(2), 13-33.
- Williams, M. (2001). The application of argumentation theory to translation quality assessment. *Meta*, 46(2), 326-344.
- Williams, M. (2004). *Translation quality assessment: An argumentation-centered approach*. Ottawa: University of Ottawa Press.