Quality Assessment of Simultaneous Interpreting: Teaching and Learning Perspective to English and Arabic Renditions

Safa'a Ahmed
Interpreting and Translation, Faculty of Languages, MSA University, Egypt
Corresponding author: Safa'a Ahmed (e-mail: ssaleh@msa.eun.eg).

Article Info

Received: February 11, 2020
Accepted: August 09, 2020
Published: August 18, 2020

Keywords

Assessment
Rubrics
Simultaneous Interpreting

ABSTRACT Despite the unprecedented interest in interpreting teaching and training, there exists a gap in the knowledge of quality assessment of simultaneous interpreting (SI) in the academia. This study aims to investigate the assessment of the quality of SI from a teaching and learning perspective, in a new way especially of looking at holistic and subjective judgements. It raises three questions. How to make assessments? How to design a rubric? How to use assessments and rubrics to help teachers and learners get quality education? This multidisciplinary study derives its theoretical tenets from interpreting studies and interactive pedagogical assessment, employing case study and questionnaire methods. The data was collected from actual assessments of bi-directional English/Arabic renditions, given to final-year university learners of SI unit. Proper assessment and clear rubrics 'with some holistic and subjective' characteristics can help enhance teachers' assessment and learners' creative performance.

I. INTRODUCTION

Despite the unprecedented interest in interpretation teaching and training, there exists a gap in the knowledge of assessing the quality of simultaneous interpreting (SI) in academia that makes assessing a performance a challenge. Campbell and Hale (2003) argue that serious concerns about SI assessment validity and reliability should be clarified (see also Hatim and Mason, 1997; Sawyer, 2004). Assessment in education is defined as "an action, usually in measurable terms, to determine the importance, size, or value of a skill or subject knowledge" (Wu, 2010, p.30). The terms 'test' and 'assessment' are used here interchangeably.

Interpreting is one of the most complicated human activities, especially on the cognitive and communicative levels (Liu, 2001, p.86). Therefore, examiners, scholars and trainers become worried about inconsistent, holistic, and subjective judgements (Wu, 2010, p.7). Furthermore, SI in industry vanishes once performed, and "what remains is the impressions received by the audience" (Riccardi, 2002, p.116). Recordings are sometimes used for evidence in education and industry (Ahmed, 2016, pp.185-186). SI assessment in this context is an under-researched challenging area.

Hence, SI researchers suggest multidisciplinary approaches based on disciplines like language testing and educational assessment to better address the issue (Hatim and Mason, 1997, pp.165-166; Sawyer, 2004,p.93; Pöchhacker, 2004,p.187). Campbell and Hale explain the literature is still "in its infancy" and it can benefit significantly from educational evaluation (2003, p.221).Valid and reliable assessment is inseparable from proper education. Sawyer presumes that "High quality education is based upon sound assessment" since participants should have 'evidence' for being assessed objectively according to Intended Learning Outcomes (2004, pp.5-7).

Therefore, this study aims to investigate SI assessment from a teaching and learning perspective, in a new way especially of looking at holistic and subjective judgements. It raises three questions about how to make assessments, design a rubric, and use them to help teachers and learners get quality education. It is a qualitative and quantitative multidisciplinary study delving into interpreting studies and interactive pedagogical assessment, through case study and questionnaire. The data is collected from actual assessments of bi-directional renditions of speeches from English (Language B) into Arabic (Language A) and vice versa, given to final-year university learners registered in SI unit. Its significance lies primarily in multidisciplinary, teacher/learner perspective and the new way of looking at quality and creativity. The study is divided into introduction,

literature review, theory, methods, findings and discussion, and conclusion and implications.

II. LITERATURE REVIEW

Most studies on SI assessment depend on surveys and interviews whose validity is questioned or tackle SI products as texts i.e. 'talk-as-text' (Wadensjö, 1998, p.79).

A. EXPERIENTIAL AND IMPRESSIONISTIC RESEARCH

Studies on SI quality started in industry rather than academia in the 1980s. They tackle users' expectations. Bühler (1986; in Wu, 2010) examined experts' expectations regarding fluency, accent, cohesion, grammar, completeness, style, etc. But experts' opinions may not reflect the other users'. Kurz (2001) extended Bühler's framework to include other end-users and concluded that interpreters have higher expectations. These studies are 'experiential and impressionistic' (Sawyer, 2004,p.20) and exclude other factors (Kalina, 2005).

B. MORE SYSTEMATIC RESEARCH

Since the 1990s researchers have approached SI quality more systematically and from different perspectives, like linguistic analysis, the interpreter's role and mediation, audience and speaker's evaluation (Kurz, 2001; Pöchhacker, 2001). Kopczyński investigates conference interpreting from linguistic and pragmatic perspectives (1994). Wadensjö's (1998, pp.50-52) explores the quality of interpreting by explaining where the interpreter's loyalty should go. Kalina (2005)'s psychological approach utilizes a datasheet listing 77 quality items. Townsley (2007) maintains that SI service providers do not concur about the quality of interpreting. Milcu (2012) deems SI assessment tools as subjective and often based on the classical approach 'the interpretation sounds good'. She proposes a vague and complicated scale of error analysis, errors, holistic method, and competence.

C. INTEREST IN SI AND ASSESSMENT IN EDUCATION

Lately, an interest in SI assessment for education has emerged. Pöchhacker (1994, 2001) differentiates between quality assurance professionally and educationally. His model is based on communication and interaction between the various actors (2004). Similarly, Riccardi (2002) distinguishes between two assessment criteria: macrocriteria for professional interpreters (e.g. equivalence, accuracy, appropriateness) and microcriteria for learners (e.g. register, omissions, alterations), but the validity of her research tools needs further testing (p.125). Sawyer (2004) ascertains that quality education requires sound curriculum and assessment. His scientific and humanistic approaches consider curriculum as a process and interaction. He admits the inability of his model to deal with all aspects of SI assessment. Carroll attempted a scale for measuring machine translation intelligibility and informativeness (1966; in Tiselius, 2005). Tiselius adapted Carroll's method to suit SI, but her generalisation may raise doubts. A large-scale survey concludes that professional experience determines the assessment in higher education and that test design is usually subjective and intuitive (Wu, 2010,p.10). Angelelli (2009) and Jacobson (2009) stress the need for a more comprehensive model. Results of evaluating Australian Sign Language refer to assessment uncertainty and subjectivity (Wang et.al, 2015). Nadir studies the 'holistic and impressionistic' features persistent to SI assessment (2017, p.1). A bird's-eye view shows 'a very uneven picture' (Pöchhacker, 2001, p.411). Little has been done to fully understand the issue.

III. THEORY

A. WU'S CONCEPTUAL MODEL OF SI TEST CONSTRUCTS

Wu, who noticed the 'holistic and subjective' or impressionistic judgement of most SI examiners, assumes "understanding how the examiners exercise their judgement can help balancing out the ephemeral nature of SI in assessment" (2010,p.7). Investigating 30 examiners' assessments, Wu thinks six factors affect performances: presentation and delivery, fidelity and completeness, audience point of view, interpreting skills and strategies, foundation abilities of interpreting and examiner behavior; the first five are assessment criteria.

Presentation and delivery refers to learners' vocal presentation and language usag, with three dimensions acoustic, word/phrase usage and flow of information (Wu, 2010, pp.161-181). Fidelity and completeness encompasses faithfulness and message completeness; this implies content accuracy, speaker intention and contextual consistency. Audience point of view means the interpreter should deliver the message faithfully from the audience perspective instead of the examiners'. Interpreting skills and strategies denotes resourcefulness (skills and strategies, background, preparation and anticipation of the topic) and multitasking. Wu adopts Gile's Efforts Model (1995) of SI: listening and analysing the original, memorizing, and producing the target. Foundation abilities for interpreting are related to personality and aptitude (like staying calm under stress) and comprehension. Of the five criteria, fidelity and completeness weighs 56%, and presentation and delivery 30%. Wu criticizes the rest of criteria, which are difficult to operate in exams, and concludes that examiners are subjective and find it hard to evaluate audience opinion. He admits his inability to "identify a prevailing pattern of the examiners' use of the assessment criteria in relation to their judgements" (pp.208-209). However, his model is useful.

B. BROWN'S INTERACTIVE APPROACH

Since the 1990s, 'assessment' has been energized with some freedom and responsibility in testing. Brown argues that we should trust our subjectivity and move away from the 'tyranny' of traditional objectivity as "our challenge was to

test impersonal, creative, communicative, interactive skills, and in doing so, to place some trust in our subjectivity, our intuition" (2007, p.461). It should target "more subjective evaluation, more individualism, and more interaction in the process of offering feedback better possibilities for intrinsic motivation, and ultimately greater validity" (p.462). An assessment becomes an interaction between teachers and learners. There are five assessment principles (Brown, 2007, pp.446-451). A good assessment should be: practical, financially, time-wise, and administratively; reliable, consistent and dependable (this applies to the test, its administration, test-taker and scoring); valid, measuring what it intends to measure; authentic, real world-inspired; and providing a washback of learners' strengths and weaknesses.

IV. METHODS

From the problem statement and aim of the study, I was able to raise three questions

How to make assessments? How to design a rubric? How to use assessments and rubrics to help teachers and learners get quality education? The study approached the issue from a multidisciplinary perspective based on Wu's categorization of SI assessment criteria with some amendments and Brown's interactive theory, through case study and questionnaire methods.

A. PARTICIPANTS

Participants are two teachers (also called examiners or markers) and learners. One teacher has a long academic and practical experience, referred to as Teacher 1 and the other with a close experience but in translation, Teacher 2. Training blind-marking sessions were organized before and during the semester to guarantee the consistency and reliability of results. Learners (L) are 100 undergraduates, the whole population of final year university students of languages, studying SI (Fall semester 2019). The learners were divided into 5 groups, 4 taught and examined by Teacher 1 and 1 by Teacher 2. The questionnaire population consists of 60 random students who happened to be available at the time of the questionnaire.

B. DATA COLLECTION AND PROCEDUR

The data was collected from 3 formal assessments, a rubric and a questionnaire. Each assessment followed Brown (2007)'s 7 steps to make a good assessment: a) assess toward clear, unambiguous objectives; b) draw up assessment specifications from objectives; c) draft your assessment; d) revise your assessment; e) final-edit; f) utilise your feedback; and g) provide ample washback. Assessments, extracted from real speeches were given in Weeks 5 (Assessment 1/A), 7 (Mid-term Exam), 11 (Assessment 1/B) and 14 (Final-term Exam). Performances were recorded on CDs for authentication and easy retrieval. Examiners, First Markers (FM) in this case, heard performances, compared them to originals and gave each learner a mark according to a rubric designed in the light of the Intended Learning Objectives of the unit, Then 20% of total performances was second-marked by a Second Marker (SM), and selected randomly from various levels of performances. This means the FM's group(s) was second-marked by the SM while the SM's was second-marked by FM. The examiners' marks were presented in tables and graphs for analysis and discussion from the teacher's perspective.

The questionnaire was conducted in Week 12 to address the issue from the learners' perspective. It asked them to listen to assessment 1/B, given in Week 11, and use the rubric to mark themselves. Teacher 1 clarified the overall goal and objective of this process and gave instructions clearly. Their marks were compared to Teacher 1's and the findings were presented in tables and graphs for analysis and discussion. They were also asked to 'rate the easiness of using this rubric' on a scale from very easy to very difficult, to choose from totally agree to totally disagree 'how far the rubric can help them improve their performance', and to answer an open-ended question about their suggestions for a more useful rubric.

C. MAKING ASSESSMENTS

The Intended Learning Outcomes for this unit are set clearly, so that learners are able to demonstrate the basic knowledge necessary for understanding SI nature and acquire professional skills and competencies gradually under various challenges e.g. stress, time, speed-rate, language competency, etc. The Egyptian Ministry of Higher Education identifies three assessments: mid- and final-term exams and a coursework whose weights are 20%, 30% and 50% respectively. Assessments reflect gradual time stress (ranging from 5-10 minutes), difficulty of speeches and speed rate (wpm). Each of the three assessments includes bi-directional renditions. Brown (2007)'s five assessment specifications were duly taken into consideration

D. DESIGNING THE RUBRIC

As mentioned above, Wu (2010) demonstrates 5 criteria of SI assessment. In fact, 'SI skills and strategies' and 'foundation abilities of interpreting' can stand as subcategories of 'presentation and delivery' and 'audience point of view' respectively. SI skills and strategies are used naturally in interpreting. In addition, the interpreter's personality and aptitude can be sensed from the audience evaluation. Hence, to design a simplified rubric adapted from Wu's study, three descriptors arise here: fidelity and completeness, presentation and delivery and audience point of view (see Shape 1).

Wu's results assign the following weight for his criteria: 56% for fidelity and completeness, 30% presentation and delivery, and 14% for the rest (2.4% audience point of view, 6.1% interpreting skills and strategies and 5.4% personality and aptitude). This distribution gives a minimal percentage to audience point of view (in my terms 2.4% plus 5.4%). To overcome the dissatisfaction with SI services, teachers should put more emphasis on audience evaluation, though difficult to measure. Based on this rationale then, marks are going to

be redistributed here: 50% for fidelity and completeness, 30% presentation and delivery, and 20% audience point of view. These are not clear-cut criteria however (check the appendix). Thus this rubric combines between holistic (evaluating intuitively each criterion as a whole) and analytic (with a more detailed rating scale) features.

The researcher was aware of the study limitations. There is no guarantee that examiners and learners use the rubric only in assessment. Examiners' second-marking and learners-FM's blind-marking were manipulated. This may result in SM getting affected by FM's mark, so the researcher held marking sessions and adopted SF-SM turn-taking to maintain reliability and consistency. The size and type of data could have been larger. Results need more testing.

V. FINDINGS AND DISCUSSION

The results of this study are given below. Only differences between the two FM and SM marks are reported in Tables I to Table VII.

Table I (see also Figure 1) shows the differences between the marks given by Teacher 1 as a First Marker (FM) and Teacher II as a second marker (SM) of 20%, i.e. 15 out of 75. The mark is calculated from 100. 9 marks are similar, so not reported. Differences (6) range from 1%-2%, signifying high consistency between both.

In Table II the differences between marks are presented, but this time Teacher 2 is FM and Teacher 1 SM of 10 renditions. 3 marks are similar and differences range from 1%-4%, reflecting high consistency too.

The results of the second part of assignment 1 (1/B) is demonstrated in Table III showing the differences (7 out of 15) between those given by Teacher 1 as FM and Teacher 2 as SM. Differences are minor, 2%.

While Table IV reveals high consistency between the marks given by Teacher 2 as FM and Teacher 1 as SM regarding the second part of assignment 1, 1/B. Differences (6 out of 10) are minus or plus 3%-6%.

Moving to Final-term Exam, Table V shows the differences between the marks given for a batch of renditions marked by Teacher 1 as FM and Teacher 2 SM of 20%. Differences (4 out of 15) range from plus or minus 1%-6%.

For the batch marked by Teacher 2 as FM and Teacher 1 as SM, differences range from minus or plus 4%-6% as shown in Table VI within consistent results.

There are no differences (0 out of 15) between the marks given by Teacher1 as FM and Teacher 2 as SM in mid-term exam; nothing is reported in a table. Meanwhile, Table VII shows 2-7% differences (3 out of 10) between the marks given by Teacher 1 as SM and Teacher 2 as FM.

The previous tables indicate that out of 80 second-marked performances, 44 (55%) show no differences between FM and SM marks, while 36 (45%) show differences of 1%-7%. The data reflects high consistency. Table I to Table VII deal with the teachers' perspective, whereas Table VIII (see also Figure 2) tackles the learners' perspective, where they mark assignment 1/B, using the same rubric for the three criteria.

TABLE I
ASSESSMENT 1/A MARKS GIVEN BY TEACHER 1 (FM) AND TEACHER 2

SI Learner	FM	SM
L1	45	47
L2	49	50
L3	69	70
L4	64	66
L5	66	68
L6	52	54

TABLE II
ASSESSMENT 1/A MARKS GIVEN BY TEACHER 2 (FM) AND TEACHER 1

SI Learner	FM	SM
L1	87	83
L2	86	87
L3	83	79
L4	59	55
L5	91	87
L6	78	74
L7	77	74

TABLE III
ASSESSMENT 1/B MARKS GIVEN BY TEACHER 1 (FM) AND TEACHER 2

SI Learner	FM	SM
L1	63	65
L2	62	64
L3	61	63
L4	60	62
L5	60	62
L6	70	72
L7	70	72

TABLE IV
ASSESSMENT 1/B MARKS GIVEN BY TEACHER 2 (FM) AND TEACHER 1

SI Learner	FM	SM
L1	89	85
L2	83	78
L3	80	77
L4	86	81
L5	82	76
L6	78	73

TABLE V
FINAL EXAM MARKS GIVEN BY TEACHER 1(FM) AND TEACHER 2

SI Learner	FM	SM
L1	44	50
L2	70	67
L3	70	69
L4	45	50

TABLE VI
FINAL EXAM MARKS GIVEN BY TEACHER 2 (FM) AND TEACHER 1

SI Learner	FM	SM
L1	70	64
L2	74	70
L3	70	65

TABLE VII
MID-TERM EXAM MARKS GIVEN BY TEACHER 2 (FM) AND TEACHER 1

SI Learner	FM	SM
L1	58	65
L2	60	67
L3	40	42

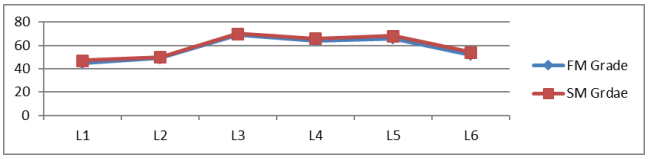

FIGURE 1. Difference between Teacher 1(FM) and Teacher 2(SM), Assessment 1/A.

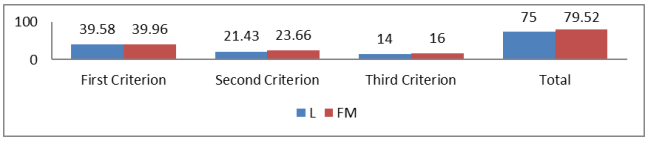

FIGURE 2. Learners' vs. FM's Marks.

Results show differences in each criterion as 67% of learners tended to have lower expectations than the teacher's, 28% had higher expectations and 5% gave similar marks. On average, learners' mark was 39.58 vs. FM's 39.96 for the first criterion (i.e. 0.7% lower expectations), 21.43 vs. 23.66 for the second (i.e. 7.32% lower expectations), and 14 vs. 16 for the third (i.e. 10% lower expectations).

As per the questionnaire, learners responded to a question about how easy the rubric was as follows: 30% selected easy, 56% neutral, 7% very easy, 7% difficult and null very difficult. The majority, 93%, found it easy and accessible. Responses to a question about how far learners think it can help improve their performances varied from 27% totally agree, 46% agree, 22% neutral, 3% totally disagree and 2% disagree. An open ended-question asked students to write down their suggestions for improving the rubric. 10 learners sought more details, 6 using it more often, 7 specific marks, and 19 no modifications. The others wrote irrelevant, miscellaneous or no comments.

To answer research questions, then, it is noteworthy to mention that before giving learners the assessments; they had had some formative assessments to experiment different solutions. "Learners must have the freedom in the classroom to experiment, to try out their own hypotheses without feeling that their overall competence is being "judged" in terms of those trials and errors" (Brown,2007,p.445). Experimentation and constructive feedback helps get creative solutions. Solving problems enhances performance; experimentation and feedback can lead to creativity. For example, some learners develop grammatical techniques to summarize the original accurately and save time to catch up with a speedy speaker. Regarding the question, how to design a rubric, the study suggested a mixture of holistic (evaluating intuitively a performance as a whole) and analytic (with marks that are more detailed judgements (see the appendix). Holistic assessment provides an opportunity to evaluate creativity and individualistic characteristic, whereas analytic assessment maintains results objectivity and reliability. The third research question was raised about how to help teachers and learners get quality education. Tables I to Table VII indicate that 55% of the second-marked performances show complete consistency between the FM's and SM's marks. 36 performances (45%) show differences ranging from 1%-7% and high consistency. Teachers are recommended to use similar assessments and rubric. As for the learners, they generally tend to underestimate their performances, maybe due to market inexperience and lack of self-confidence. The 0.7% difference for fidelity and completeness may indicate criterion easiness or clarity. A 7.32% difference in delivery and presentation may reveal vagueness or lack of language competence, yet the percentage does not affect consistency much. The biggest difference, 10%, in evaluating the audience opinion may be attributed to inexperience. Indeed, learners' highly amazing ability to mark their performances, together with teachers' feedback, enhances performances as they become well-aware of weaknesses and strengths

VI. CONCLUSION AND IMPLICATIONS

This study has attempted to explore the assessment of the quality of simultaneous interpreting from a teaching and learning perspective in a new way. A qualitative and quantitative multidisciplinary approach, delving into interpreting studies and interactive pedagogical assessment, has employed case study and questionnaire methods.

The data shows high consistency between the First Marker's and Second Marker's marks. The same applies to the learners' vs. the FM's. The interesting results are related to the learners' marks. Their surprisingly amazing ability to mark using the rubric enriches their performances. Teachers' washback is also valuable in informing them of their weaknesses and strengths. They can think creatively of what to do next to improve performances. Assessments should allow learners to experiment and have washback; a matter if utilized properly can help them be creative in solving SI problems. A rubric that reflects holistic and analytic features is inevitable for creativity, objectivity and reliability. Many SI researchers express their fears about holistic and subjective assessment of SI. They are right if there is no clear objectives, no proper assessment or rubrics. The study concludes that proper assessment and clear rubric can help both teachers and learners achieve the ILOs and get quality education. SI quality assessment as such becomes a vital part of the learning process. Therefore, teachers are encouraged to use similar assessments and rubric. The implications are significant for SI teaching and testing academically and professionally. Indeed, further research is needed to test the rubric criteria and rating scale in similar (academia) and different (like trainings and SI services) contexts.

REFERENCES

Ahmed, Safa'a (2016) Quality Assessment in Community Interpreting: A Case-study of Talk-shows Renditions from English into Arabic. *Occasional Papers in the Development of English Language Education*, CDELT, 61 (Jan.), 403-456.

Angelelli, Claudia V. (2009) 'Using a Rubric to Assess Translation Ability: Defining the Construct', in Angelelli, Claudia V. and Jacobson, Holly

TABLE VIII
LEARNERS' VS. FM'S MARKS WITH THEIR HIGHER (HE) OR LOWER EXPECTATIONS (LE)

Criterion Learner	L: 1st Criterion	FM:1st Criterion	L: 2nd Criterion	FM:2nd Criterion	L: 3rd Criterion	FM:3rd Criterion	L: Total	FM: Total	HE	LE
L1	35	36.5	26	23.5	12	15	73	75		2
L2	45	37.5	25	20.5	18	15	88	73	15	
L3	40	39	20	24	15	16	75	77		2
L4	40	37	20	20	12	15	72	72	-	-
L5	25	45	15	27	10	18	50	90	40	
L6	50	41	26	25	17	16	93	82	11	
L7	45	41	20	25	10	17	75	83		8
L8	30	44	20	27	10	19	60	90		30
L9	45	41	20	23	15	16	80	80	-	-
L10	45	44	22	26	14	17	81	87		6
L11	48	48	27	28	19	19	94	95		1
L12	40	36	22	20	15	14	77	70	7	
L13	45	44	23	26	15	17	83	87		4
L14	49	48	28.5	28	19	19	96.5	95	1.5	
L15	38	40	19	24	13	16	70	80		10
L16	37	42	17	23	12	16	66	81		15
L17	40	38	15	23	14	16	69	77		8
L18	40	43	20	26	12	17	72	86		14
L19	37	39	15	22	10	16	62	77		15
L20	40	35.5	20	20	15	14.5	75	70	5	
L21	40	41	15	25	10	16	65	82		20
L22	44	42	27	26	17	17	88	85	3	
L23	40	46	20	26	15	18	75	90		15
L24	30	35	20	23	15	15	65	73		8
L25	35	36	15	22	12	15	62	73		11
L26	35	39	20	23.5	15	15.5	70	78		8
L27	40	36.5	20	23	15	15.5	75	75	-	-
L28	30	40	20	22	10	15	60	77		17
L29	40	45	23	26	17	17	80	88		8
L30	43	46	25	26	15	18	83	90		7
L31	40	41	25	24	14	16	79	81		2
L32	46	42	25	25	15	16	86	84	2	
L33	42	46	26	28	14	18	82	92		10
L34	45	35	22	20	15	15	82	70	12	
L35	45	44	26	26	18	18	89	88	2	
L36	40	46	20	27	15	18	75	91		16
L37	45	45	25	27	15	17	85	89		4
L38	40	41	20	25	10	16	70	82		12
L39	25	39	13	23.5	9	16.5	47	79		32
L40	40	25	20	15	15	10	75	50	25	
L41	35	35	20	22	14	15	69	72		3
L42	40	47	25	29	17	19	82	93		11
L43	30	35	20	20	10	15	60	70		10
L44	40	36	23	22	12	15	75	73	2	
L45	40	42	20	24	15	15	75	83		8
L46	40	35	22	21	13	14	75	70	5	
L47	30	38.5	15	23.5	12	16	57	78		21
L48	47	49	25	29	18	20	90	98		8
L49	40	32	25	19	16	12	81	63		18
L50	41	40	22.5	24	16	16	78.5	80		1.5
L51	40	40	20	24	10	16	70	80		10
L52	40	45	25	25	15	18	80	88		8
L53	38	42	22	24	15	17	75	83		8
L54	45	41	27	24	20	18	92	83	9	
L55	40	24	25	16	15	10	80	50	30	
L56	40	32	20	18	10	12	70	62	8	
L57	40	38	25	20	15	14	80	72	8	
L58	40	38	15	23	15	16	70	74		4
L59	35	35	20	20	12	15	67	70		3
L60	35	43	22	26	12	16	69	85		16
Total	2375	2397.5	1286	1417.5	840	960	4500	4771		
Average	39.58/ 50	39.96/50	21.43/ 30	23.66/ 30	14/20	16/ 20	75/ 100	79.52/100	7 28.2%	40 66.8%
Percent	79.17%	79.92%	71.43%	78.75%	70%	80%	75%	79.52%		

E. (eds.) *Testing and Assessment in Translation and Interpreting Studies*. John Benjamins. pp.13-47.

Brown, H. Douglas (2007) *Teaching by Principles: An Attractive Approach*. 3rd edn. New York: Pearson Education.

Campbell, S. and Hale, S. (2003) Translation and Interpreting Assessment in the Context of Educational Measurement, in Anderman, G. and Rogers, M. (eds.) *Translation Today: Trends and Perspectives*. Clevedon, UK: Multilingua Matters Ltd, pp. 205-224.

Carroll, J.B. (1966) An Experiment in Evaluating the Quality of Translations. *Mechanical Translation*, 9, 55-66.

Gile, D. (1995) *Basic Concepts and Models for Interpreter and translator Training*. John Benjamins.

Hatim, B. and Mason, I. (1997) *The Translator as Communicator*. Routledge.

Jacobson, Holly (2009) Moving beyond words in assessing mediated interaction: Measuring interactional competence in healthcare settings, in Angelelli, Claudia V. and Jacobson, Holly E. (eds.) *Testing and Assessment in Translation and Interpreting Studies*. John Benjamins. pp. 49-72.

Kalina, S. (2005) Quality Assurance for Interpreting Processes. *Meta*, 50(2), 768-784.

Kopczyński, A. (1994) Quality in Conference Interpreting: Some Pragmatic Problems, in S. Lambert and Moser-Mercer, B. (eds.) *Bridging the Gap - Empirical Research in Simultaneous Interpretation*. John Benjamins, pp.87-99.

Kurz, I. (2001) Conference Interpreting: Quality in the Ears of the User. *Meta*, XLVI, 394-409.

Liu, M. (2001) *Expertise in Simultaneous Interpreting: A Working Memory Analysis* [unpublished PhD dissertation]. Graduate School of the University of Texas at Austin. Austin, Texas: University of Texas.

Milcu, Marilena (2012) Evaluation and Self-evaluation in Simultaneous Translation: Assessment Methods, *Procedica-Social and Behavioural Sciences* 46, 4253-4257.

Nadir, Abdelkader (2017) *Simultaneous Quality Assessment at the Higher Arab Institute of Translation from Teachers' and Students' Perspectives* [MA Thesis]. Algeria: Faculty of Foreign Languages, University of Algeria.

Pöchhacker, F. (1994) Quality Assurance in Simultaneous Interpreting. In C. Dollerup and A. Lindegaard (eds.) *Teaching Translation and Interpreting 2*. John Benjamins, pp.233-242.

--(2001) Quality Assessment in Conference and Community Interpreting. *Meta*, XLVI (2), 411-425.

--(2004) *Introducing Interpreting Studies*. Routledge.

Riccardi, A. (2002). Evaluation in Interpreting: Macrocriteria and Microcriteria, in

Hung, E. (ed.) *Teaching Translation and Interpreting 4*. John Benjamins, pp.115-126.

Sawyer, David B. (2004) *Fundamental Aspects of Interpreter Education*. John Benjamins.

Tiselius, Elisabeth. (2005***)*** *"The Work Shows the Workman": An Investigation of Two Methods for Measuring Interpreting Performance and Studying Expertise in Interpreting*. Stockholm: University of Stockholm.