

University of Algiers 2  
Faculty of Foreign Languages  
Department of English/Anglophone Studies

# **Simultaneous Interpreting Quality Assessment at the Higher Arab Institute of Translation from Teachers' and Students' Perspectives**

**A Dissertation submitted in partial fulfillment of the requirements for the  
Magister Degree in English Linguistics and Didactics**

**Submitted by**

**Mr. Abdelkader NADIR**

**Supervised by**

**Prof. Faïza BENSEMMANE**

**January 2017**

## **Board of Examiners**

- **Chairman :** Prof. Zoulikha BENSAFI, University of Algiers 2
- **Supervisor :** Prof. Faïza BENSEMMANE, University of Algiers 2
- **Examiner :** Dr. Yasmine BOUKHEDIMI, University of Algiers 2

## **Declaration**

**I hereby declare that the substance of this dissertation is entirely the result of my investigation and that reference or acknowledgment is made, whenever necessary, to the work of other researchers.**

**I am duly informed that any person practicing plagiarism will be subject to disciplinary sanctions issued by university authorities under the rules and regulations in force.**

**Date: January 2017**

**Signed by: Mr. Abdelkader NADIR**

# Dedication

To my parents and grandparents

To my brothers and sisters

## **Acknowledgements**

I would like to thank Prof. Bensemmane for supervising this modest work. But, it is hard to find the words and expressions to express the intended sense, because their overuse weakened their meanings. Dear Professor, words are powerless to thank you for your invaluable feedback and the time you devoted to improve this dissertation.

I would like to extend heartfelt thanks to the Board of Examiners for their time to read and assess my dissertation.

I would like to extend special words of thanks to all my professors at the University of Algiers 2.

I would like to express my gratitude to the professors and students of the Higher Arab Institute of Translation for accepting to take part in this study.

## **List of Acronyms, Abbreviations and Shortenings**

**CI:** Consecutive Interpreting

**ComI:** Community Interpreting

**IQ:** Interpreting Quality

**IQA:** Interpreting Quality Assessment

**HAIT:** The Higher Arab Institute of Translation (in French: ISAT)

**QA:** Quality Assessment

**QE:** Quality Evaluation

**S:** Semester

**SI:** Simultaneous Interpreting

**SIQ:** Simultaneous Interpreting Quality

**SIQA:** Simultaneous Interpreting Quality Assessment

**SL:** Source Language

**ST:** Source Text

**STU:** Student

**TEA:** Teacher

**TL:** Target Language

**TQA:** Translation Quality Assessment

**TT:** Target Text

## **List of Diagrams and Figures**

**Diagram 1.** Quality standards for the product and service of interpreting.

**Diagram 2.** Conceptual spectrum of interpreting.

**Diagram 3:** Taxonomy of Assessment Constructs according to the Teachers' Perspective

**Diagram 4:** Taxonomy of Assessment Criteria according to the Teachers' Perspective

**Diagram 5:** Suggested Hierarchical Taxonomy of Teachers' Assessment Constructs

**Diagram 6:** A Criterion-based Assessment Rubric for Proficiency Testing.

**Figure 1.** Dimensions of interpreters' output quality

**Figure 2.** Factors relevant to interpreting quality

## **List of Tables**

### **Chapter Four:**

Table 1. Curriculum and Assessment of SI at the (HAIT) ISAT – (2005-2007)

Table 2. Curriculum and Assessment of SI at the (HAIT) ISAT – (2008-2010)

Table 3. Curriculum and Assessment of SI at the (HAIT) ISAT – (2012-2015)

Table 4. Hourly Volume of Credits - HAIT

### **Chapter Five:**

Table 5: An example of thematic, open and axial coding

Table 6. Number of Collected Questionnaires per Class

Table 7. Gender per Level of Study

Table 8. Age of the Students per Level of Study

Table 9. Nationality of the Students per Level of Study

Table 10. Degrees of the Students enrolled at the HAIT (ISAT)

Table 11. Students' Mastery of Languages – HAIT (ISAT)

Table 12. Number of Students who took the Admission Test & the IAT (HAIT)

### **Appendices:**

Table 13. Data Coding of the Pilot Study Questionnaire (for Teachers)

Table 14. Data Coding of First-year Students' Responses (Semester 1)

Table 15. Data Coding of First-year Students' Responses (Semester 2)

Table 16. Data Coding of Second-year Interpreter Students' Responses (Semester 3 + Semester 4)

Table 17. Data Coding of Graduate Students' Responses

## **Abstract**

This dissertation is an exploratory study of the process of assessment practices of simultaneous interpreting quality (SIQ) at the Higher Arab Institute of Translation (HAIT) from teachers' and students' perspectives. It delves into the perceptions, views and opinions of the teachers and students alike regarding the procedures of quality assessment of simultaneous interpreting along the process of training, i.e. from the Admission Test to graduation. It aims at generating hypotheses on quality assessment issues for further research in this field in order to improve the teaching of simultaneous interpreting in this Institute and elsewhere. This research looks into the scoring of the Admission Test, the Interpreter Aptitude Test and the formal Exams of Simultaneous Interpreting against the principles of educational assessment, namely validity, reliability, authenticity, practicality in addition to fairness, usefulness and meaningfulness. It is based on a post-positivist paradigm and a mixed-methods approach, and it is mainly qualitative as it does not just deal with the statistical data, i.e. the numbers, but with the subjective experience of the respondents. It relies on the methods of case study, grounded theory and a questionnaire-based survey. As far as the results are concerned, they have revealed, on the one hand, the lack of teachers' awareness about how to construct and design valid, reliable and authentic educational tests and exams. On the other hand, the results revealed a mismatch between the perspectives of the teachers, who are professionals, and those of the students regarding the skills and criteria of assessment in almost all the tests and exams.

**Key words:** Quality Assessment, Simultaneous Interpreting Quality Assessment, the Higher Arab Institute of Translation, Teachers' Perspective, Students' Perspective

## عنوان الرسالة: تقييم جودة الترجمة الفورية في المعهد العالي العربي للترجمة من منظور الأساتذة والطلبة

### ملخص باللغة العربية

هذه الرسالة هي دراسة استكشافية لمسار ممارسات تقييم جودة الترجمة الفورية في المعهد العالي العربي للترجمة من منظور إدراكات الأساتذة والطلبة. وهي تبحث في مدارك ووجهات نظر وآراء الأساتذة والطلبة على حد سواء بشأن كيفية وإجراءات تقييم جودة الترجمة الفورية طيلة مسار التكوين أي ابتداء من اختبار القبول إلى التخرج. وتهدف هذه الدراسة لإنتاج وبلورة افتراضات حول قضايا تقييم الجودة لتعميق البحوث في هذا المجال من أجل تحسين تدريس الترجمة الفورية في هذا المعهد وفي غيره من المعاهد. ويفحص هذا البحث مدى استجابة اختبار القبول واختبار استعداد الترجمان وكذا الامتحانات الرسمية للترجمة الفورية لمبادئ التقييم التربوي مثل الصدقية والثبات والأصالة وسهولة التطبيق إضافة للعدالة والإفادة والدلالة. وترتكز هذه الدراسة على برادام ما بعد الوضعية ومقاربة متعددة الطرائق إذ إنها عموما دراسة نوعية لأنها لا تقتصر على فحص المعطيات الإحصائية بل تتطرق للخبرات الذاتية للمبحوثين. وتعتمد هذه الدراسة على ثلاثة طرائق هي دراسة الحالة والنظرية التجديرية والمسح الاستبائي. وأما النتائج فقد بينت من جهة نقصا في وعي الأساتذة بخصوص بناء وتصميم اختبارات وامتحانات تربوية ذات صدقية وموثوقية وأصالة. ومن جهة ثانية، بينت النتائج عدم تطابق بين منظور الأساتذة المنحدرين من فضاءات مهنية ومنظور الطلبة بشأن المهارات ومحكات التقييم في كل الاختبارات والامتحانات تقريبا. وبالتالي فإننا نوصي بإيلاء قضية التقييم اهتماما أكبر ومزيدا من الدراسة المعمقة الإمبريقية منها والتجريبية إضافة لتكوين الأساتذة وتدريبهم عمليا. علاوة على ذلك، يجب الأخذ بعين الاعتبار أهمية وعي الطلبة بقضايا التقييم في مسار التعلم.

**الكلمات المفتاحية:** تقييم الجودة، تقييم جودة الترجمة الشفوية، المعهد العالي العربي للترجمة، منظور الأساتذة،

منظور الطلبة

## Table of Contents

<i>Board of Examiners</i>	ii
<i>Declaration</i>	iii
<i>Acknowledgments</i>	iv
<i>Dedication</i>	v
<i>List of Acronym, Abbreviations and Shortenings</i>	vi
<i>List of Diagrams and Figures</i>	vii
<i>List of Tables</i>	viii
<i>English Abstract</i>	ix
<i>Arabic Abstract</i>	x
<i>Table of Contents</i>	xi
<b>General Introduction</b>	<b>1</b>
<b>Chapter One: Assessment Concepts, Terms and Issues</b>	
1.1. Definition of Basic Concepts and Terms in Assessment	8
1.1.1. Measurement, Assessment and Quality Assessment	8
1.1.2. Towards a Movement of Educational Assessment	9
1.1.3. Principles of Assessment	14
1.1.3.A. Validity	14
1.1.3.B. Reliability	15
1.1.3.C. Practicality	16
1.1.3.D. Washback or Backwash	16
1.1.3.E. Authenticity	17
1.2. Construction and Design of Quality Assessment Tools	17
1.3. Main Assessment Issues	20
1.3.1. Accountability	20
1.3.2. Fairness, Ethics and Equity	21
1.3.3. Educational Outcomes	21
1.4. Interpreting Assessment and Related Fields	22
Summary & Conclusion	24
<b>Chapter Two: A Critical Review of Works on Simultaneous Interpreting Quality Assessment (SIQA)</b>	
2.1. Complexity of Interpreting Quality Assessment (IQA)	26
2.1.1. Purpose and Theoretical Frameworks of IQA	28
2.1.2. Methodological Considerations in IQA	33
2.2. A taxonomy of SIQA Models:	35
2.2.1. Academe-based SIQA Models	36

2.2.2. Industry-based SIQA Models	40
Summary & Conclusion	42

### **Chapter Three: Criteria, Norms & Standards of Simultaneous Interpreting Quality Assessment (SIQA)**

3.1. Definitions of SIQA	43
3.2. Educational Criteria for SIQA of Trainee Interpreters	45
3.2.1. Aptitude Criteria	46
3.2.2. Ability Criteria	59
3.2.3. Achievement Criteria	61
3.3. Industry-based criteria, norms & standards for SIQA	62
3.3.1. Self & Peers Criteria	62
3.3.2. Certification and Accreditation Criteria, Norms & Standards	66
3.3.3. Professional Criteria, Norms & Standards: Employers & Clients	70
3.3.4. End-users Criteria	70
Summary & Conclusion	73

### **Chapter Four: Methodology, Design & Procedure**

4.1. Context of the Study: The Higher Arab Institute of Translation (HAIT)	74
4.1.1. The programmes: Degree Tracks & Language Combinations	75
4.1.2. Interpreting Modes at the HAIT	76
4.1.3. Assessment of SI at the HAIT	77
4.2. Research Paradigm, Approach and Methods	82
4.3. Research Instruments	85
4.4. Questionnaires' Design	86
4.4.1. Design of the Pilot Study (The Questionnaire to Teachers)	87
4.4.2. Design of the Follow-Up Study Questionnaires	88
4.4.3. Population of the Study	89
4.5. Administration of the Questionnaires	90
4.5.1. Administration of the Pilot Study Questionnaire	90
4.5.2. Administration of the Follow-Up Study Questionnaires	90
Summary & Conclusion	91

### **Chapter Five: Presentation of the Results**

5.1. Qualitative Analysis of the Pilot Study Results	92
5.1.1. Coding of Teachers' Responses	92
5.1.2. Analysis of Teachers' Responses	95
5.1.2.A. Results of the Hierarchical Axial Coding	96

5.1.2.B.	Results of the Non-Hierarchical Selective Coding	108
5.2.	Quantitative Analysis of the Follow-Up Study Results	114
5.3.	Qualitative Analysis of the Follow-Up Study Results	120
5.3.1.	Coding of Students' Responses	120
5.3.2.	Analysis of Students' Responses	121
5.3.2.A.	Results of the Hierarchical Axial Coding	121
5.3.2.B.	Results of the Non-Hierarchical Selective Coding	128
	Summary & Conclusion	131

## **Chapter Six: Interpretation and Discussion of the Results**

6.1.	Interpretation of the Results	132
6.1.1.	Validity	132
6.1.2.	Reliability	135
6.1.3.	Scoring and Grading	136
6.1.4.	Practicality	137
6.1.5.	Feedback and Washback	137
6.1.6.	Fairness, Usefulness and Meaningfulness	138
6.1.7.	Authenticity	138
6.1.8.	Tests and Exams' Specifications	139
6.2.	Synthesis and Discussion of the Results	145
6.2.1.	Theme 1: The Admission Test	145
6.2.2.	Theme 2: The Interpreter Aptitude Test	147
6.2.3.	Theme 3: Exams	148
6.2.4.	Theme 4: The Whole Procedure of SI Assessment	149
6.2.5.	Theme 5: Emerging Hypotheses	151
6.3.	Limitations of the Study	152
	Summary and Conclusion	154

## **General Conclusion & Recommendations** 155

### ***Bibliography*** 159

### ***Appendices*** 178

## **General Introduction**

This Magister dissertation is intended to investigate thoroughly Simultaneous Interpreting Quality Assessment (SIQA) as practiced at the Higher Arab Institute of Translation (HAIT)<sup>1</sup> from the perspective of Educational Assessment (EA).

### **a- Rationale for the Study:**

The choice of this topic came after reflecting introspectively about my own training at this institution, i.e the HAIT, where I obtained a Magister Degree in Interpreting. Besides, after attending Magister Seminars in English Linguistics and Didactics at the University of Algiers 2, I acquired valuable knowledge in Methodology, Linguistics, Classroom Research and Assessment which triggered my interest and reflective thinking about my learning experience of translation and interpreting at the HAIT, increased my understanding of the topic and raised my awareness about the interfering and complex issues involved in this research.

Through informal discussions I had with my classmates during that training, I realised that teachers' assessment at the HAIT is holistic and impressionistic. Although the HAIT is equipped with a SANAKO lab, not all the teachers use it, neither to present the teaching or testing prompts nor to record the performance of the trainees. No rubrics or checklists are used for scoring and no explicit assessment criteria are provided to the trainees before sitting for exams. The students have their own understanding and perceptions about what a 'good interpretation' is and perform on that basis. Moreover, they seem to see no difference between the first and second year in terms of progress in skills acquisition and assessment. There seems to be a mismatch between teachers and trainees on defining good interpreting that this study will attempt to identify. Students' views, perceptions and thoughts on quality assessment will be compared with the teachers' views, perceptions and thoughts.

---

<sup>1</sup> The acronym that is used in the documents of this Institute, i.e. The Higher Arab Institute of Translation, is the French one, ISAT. It stands for "Institut Supérieur Arabe de Traduction". The HAIT has changed its premises several times; therefore, for updates, we refer the reader to its URL which is: <http://isat-al.org>

## **b- The Research Question:**

In this study, we will try to answer the main research question that follows: Is the current assessment procedure of Simultaneous Interpreting (SI) based on valid, reliable, authentic and practical tests and exams? This research question has been broken down into the following secondary or sub-questions:

1. Is the Interview an adequate and sufficient test as a tool for admission?
2. Is the Interpreter Aptitude Test (IAT) valid, reliable and practical as a tool to assess the trainees' aptitudes?
3. Do the exams target the competencies needed for future SI at the 2nd year degree level?
4. Are the tests and exams built on a well-defined rationale?

A holistic approach will be used to examine the different tools and techniques for assessing interpreting quality along the process of learning, from the Admission Test to graduation. Yet, the overall performance or the end-of-training achievement performance of the trainees in simultaneous interpreting is targeted. The aim is to examine the whole testing and examining procedure and not only a specific test or exam. We will examine quality assessment of the would-be interpreters' aptitude, ability and achievement performance at the HAIT in this linear way. Specially, the following assessment principles will be explored in relation to the assessment tools in force at the HAIT: validity, reliability, authenticity, practicality.

## **c- Significance of the Study:**

This study is significant insofar as it is, to our knowledge, the first attempt to examine in-depth interpreting quality assessment practices at the HAIT and to focus on the whole assessment procedure of SI from an educational perspective. This study examines the construction and design of the tests and exams at the HAIT and triangulates teachers' and students' perspective regarding assessment practices at the HAIT. This is quite innovative and will hopefully bring insights into the issue of quality assessment in Algeria and elsewhere. This study is qualitative and Grounded Theory has been used to examine the assessment procedure as a whole in order to generate hypotheses about the tests and exams at the HAIT.

#### **d- Methodological Considerations:**

It is worth noting that delving into the theme of Quality Assessment (QA) and IQA in particular is not an easy task, because of a number of methodological considerations and a lot of lines of investigation that must be taken into account. It is worth pointing to the controversial findings of research raised by many scholars due to the complexity of the phenomenon of interpreting, the newness of the fields of translation and interpreting studies and their unestablished traditions of investigation in addition to deficiencies in basic research design (e.g. Shlesinger, 1995; Hatim & Mason, 1997: 164; Gile, 1998; Setton, 1999; Martínez Melis & Hurtado Albir, 2001; Pöchhacker, 2001, 2004; Sawyer, 2004; Angelleli & Jacobson, 2009).

Firstly, terminology in the field of quality assessment is difficult to define. For example, Hatim and Mason (1997) consider that there are many sub-fields in the area of evaluation in translation: Translation Quality Assessment (e.g. House, 1981), Translation Criticism and Translation Quality Control (TQC) on the one hand and Performance Assessment (e.g. Nord, 1991) on the other. This research does not deal with these fields, because it is focused on SIQA only. Nonetheless, it does not neglect them totally.

In this research, we are interested in the assessment of trainee interpreters and not industry-based professional practitioners whose assessment is done in professional contexts. Apart from Quality Control which refers to Translation Project Management Quality Controls, elsewhere called Quality Assurance or considered as part of it (Gouadec, 1982), we consider that all the other activities of assessment mentioned above are interrelated and inform each other and are both insightful and useful in interpreter assessment, because they are focused on the translation or interpreting activity per se. We will exclude Translation Quality Controls, because it refers to a very different and complex process which involves many other agents. For example, in written translation, TQC includes the reviewers, the proofreaders, the editors and the project managers or the translation agency in general. In interpreting, the process might include peers and supervisors. Nevertheless, in the real-life job, most of the

time, the interpreter has no second chance to ensure quality. Overall, in Language Assessment, scholars distinguish between three levels which are, from the broadest to the narrowest, evaluation, assessment and testing. According to Brown and Abeywickrama (2010: 348), evaluation refers to *“making decisions and/or value judgement based on the results of tests, other assessments, and/or teachers’ reports”*. However, practically speaking, the distinction between evaluation and assessment is not always clear-cut and the terms are often used interchangeably. In this study, we will consider that IQA refers to any kind of evaluation, assessment or testing of the interpreter students’ performances during the training or at its end. However, it should be pointed out here that performance assessment is also different from proficiency assessment, because the first one must be related to a given curriculum or course content; whereas, the second *“is not limited to any one course, curriculum, or single skill in the language; rather, it tests overall global ability”* (Brown & Abeywickrama, 2010: 351). So, performance assessment will be used, because it relates to the HAIT curriculum like any educational institution.

Secondly, eventhough there is interference between translation and interpreting, IQA is different from translation quality assessment (TQA) (Sawyer, 2004; Angelleli & Jacobson, 2009). As explained by Angelelli (2012: 1),

*“There are similarities and differences between testing for translation and testing for interpreting, just as there are similarities and differences between testing for writing and testing for oral production in any language. They differ in the construct they measure (e.g. translation versus interpreting skills, producing written language versus producing oral language respectively) and, at the same time, they are part of a larger whole (e.g. translational/interpreting competence, communicative competence, respectively).”*

She rightly concludes that:

*“T&I testing varies in terms of approaches and constructs. In the last twenty years, testing has evolved significantly and it is gaining more interest among T&I scholars. Undoubtedly, methodology and rigor are essential components of test development. Both academe and industry will continue to benefit enormously from making accurate and sound decisions*

*on translation/interpreting ability and quality, processes and products, based on meaningful testing.”*

But, TQA is beyond the scope of the present study.

Thirdly, due to the relative nature of quality, the term Quality is very complex, multifaceted and multi-dimensional. As we will expound in the second chapter, especially in interpreting, there are many criteria, norms and standards which determine the nature of quality and no consensus on what it refers to. Due to the fact that there cannot be one common definition shared among all the scholars, the professionals and the end-users, quality is deemed to be in the eye of the beholder and generalization is very difficult, as pointed out by many scholars (Sager, 1983; Williams, 1989; Waddington, 2003, etc). According to Grbić (2008), quality is a social construct and its relativity is evident because quality depends both on the individual who is assessing it and on the given context or specific situation where it is assessed. She warns that even the benchmarks used to gauge quality are relative in the way that quality as an attribute or notion pertaining to an object is relative.

Fourthly, quality can be evaluated in the fields of translation and interpreting from various perspectives. As pointed out by Pöchhacker (2001), in *Interpreting Quality Assessment*, a lot of parts have stakes, the researcher, the client, the end-user and the practitioner (i.e. the interpreter himself or his colleagues). Sometimes, the client is not the direct employer, for instance the client can be a university but the employer is the translation agency; therefore, they may have different views on quality. As far as this research is concerned, we are interested in assessment in the context of interpreting pedagogy at post-graduate level. Hence, our research is both process-oriented and product-oriented.

Fifthly, trainees' evaluation in educational settings can be approached with different objectives. We may have diagnostic, formative and summative forms of assessment. Each type has different objectives and different instruments. We will deal with these forms in the next chapter. Yet, it is worth pointing out here that we are referring to novice student interpreters' assessment and not to the assessment of the

professional educational interpreters who might become trainees for the sake of skills building and professional development. Sawyer (2004: 5) confirmed that “*high quality education is based on sound assessment*” and that IQA is a challenge. He explained the integrative role of assessment in the implementation and success of the curriculum in interpreting by relating research in four main areas: curriculum design and implementation, building expertise in interpreting, assessment practices, and the pedagogical aspects of language transfer competence. By taking examples from the three programmes of the Graduate School of Translation and Interpreting (GSTI) at the Monterey Institute for International Studies, he tried to consider the extent to which constructs from curriculum, expertise, and assessment theory are related to interpreting pedagogy and the role of translation instruction in building interpretation competence. Additionally, he presented an enhanced Y-track curriculum as a proposed model for future discussion where he included formative, ipsative and summative forms of assessment. However, he failed to deal fully with SIQA. He acknowledged the limitations of his study by stating that:

*“Developing an overarching understanding of the role of assessment in learning and instruction necessarily involves approaching multiple areas, not all of which can be explored thoroughly [here]. They include, for example, (1) the role of assessment and testing in all stages of the curriculum, (2) the role of various types of assessment in learning and specific events of instruction, (3) the relationship between assessment for educational purposes and the evaluation of quality in professional practice, and (4) the similarities and differences in assessment for the teaching of interpreting and the teaching of translation.”*  
(p.31-32)

Our study takes research up from Sawyer’s (2004) conclusion and will hopefully contribute to research on interpreter education with focus on SI assessment practices at postgraduate level.

Finally, there are different modes of interpretation (sight translation/interpreting, community interpreting/public service interpreting, consecutive interpreting and simultaneous interpreting, sign language interpretation, etc). Each type requires different types of skills and sub-skills. We will focus only on simultaneous interpreting and its related skills and sub-skills. Unlike Sawyer (2004) who explored the

assessment of ‘conference interpreting’ (as practiced at the GSTI) by examining the assessment of two modes of interpreting at the same time in addition to written translation, we will focus on one mode only to narrow down the scope of this research and provide a deeper analysis with empirical data to reflect the whole process of assessment, the assessment procedures and the exit profile of the graduates.

**e- Structure of the Study:**

This dissertation is divided into six chapters. Before explaining the methodology, we will review the literature. In the first chapter, we will shed light on the concepts, terms and issues of assessment. In the second chapter, we will deal with a critical review of works on SIQA and show how it relates to Language Assessment (LA). In the third chapter, we will explore the criteria, norms and standards of SIQA, both in academic and professional contexts. In the fourth chapter, we will explain and justify our methodology, design and procedure. First, we will introduce the HAIT and its SI curriculum. Then, we will deal with the research methodological choice of paradigm, approach and methods. We will also describe the design and administration of the research instrument. In the fifth chapter, we will present and analyze the results of the Pilot Study and the Follow-up study, both quantitatively and qualitatively. In the sixth chapter, we will interpret and discuss the results. Further, the limitations of this research will be pointed out. Finally, we will conclude and give recommendations to improve the assessment procedure of SI at the HAIT and teaching, learning and assessment of SI in general.

# **Chapter One :**

## **Assessment Concepts, Terms and Issues**

In this chapter, we will introduce the main concepts and terms used in the field of assessment. After that, we will deal with the main assessment principles and steps for designing high-quality assessment tools. Then, we will tackle the long standing and current assessment issues such as accountability, ethics and fairness. Moreover, we will shed light on the difference between IQA and some relative fields such as Language Assessment (LA). Finally, we will conclude the chapter by a short summary of the most important points that should be retained and the most salient conclusions.

## **1.1. Definition of Basic Concepts and Terms in Assessment**

In this section, we will deal with the main concepts and terms such as measurement, quality assessment, standardized assessment, educational assessment, classroom assessment, referencing, benchmarking. Besides, we will deal with the principles in the fields of quality assessment and educational assessment, namely validity, reliability, authenticity, practicality, etc.

### **1.1.1 Measurement, Assessment and Quality Assessment**

Concerns about measurement and assessment of aptitude, educational attainment and achievement have very ancient roots: China (210 B.C.E), Europe (late 12<sup>th</sup> C.), Cambridge in the UK (mid-1700), and USA (1840s). Moreover, the measurement of intelligence and aptitude goes back to the 1800s and 1920s. Furthermore, the integration of Bloom's Taxonomy of Educational Objectives which prioritized cognitive skills in examinations goes back to the 1950s - 1970s (Wood, 1993; Black, 1998; Odendahl, 2011), and the assessment movement in higher education dates back to the 1980s (Ewell, 2002). However, assessment concepts and issues such as quality, especially in connection with validity and reliability, are still controversial (Broadfoot & Black, 2004; McAlpine, 2002). Basing this review on exploring terminology and clarifying the concept of quality assessment is necessary for this study. We will try to draw a picture of the main trends in assessment with distinctive colours to draw attention to their importance, worth and implications for interpreting, without delving into all the intricacies of the heated debates.

Initially, the notion of QA was used in the areas of industry, manufacturing and project management which use practical assessment procedures to get rid of any defects that might be found either in the process or the final product. To confirm this view, Clifford (2005) states that the literature in Translation Quality Assessment (TQA) and Interpreting Quality Assessment (IQA) conduct assessment in the same way as in the manufacturing industry, i.e. the main concern is about products' defects through a set of procedures of examination. In addition, we may add that the assessors focus on the quality of the process through a thorough analysis of its traceability. Thus, we notice that quality in these fields is guaranteed both along the process and through the output. There are now even specialized bodies which award specific stamps like ISO (International Standards Organisation) to prove the compliance of the companies with the international norms and standards of production. So, the term quality is now related to anything; from speech quality, voice quality to healthcare quality, public service quality, etc.

### **1.1.2. Towards a Movement of Educational Assessment**

The use of the concept of Educational Measurement (EM) by authors like Ebel and Frisbie (1991) and Ayres (1918, as cited by Odendahl, 2011) is sometimes confusing as they use it interchangeably with assessment. It is worth pointing out that Educational measurement refers to the quantification of educational results by applying a set of rules (a score scale) (Bachman, 1990; Brookhart, 2004; Odendahl, 2011); whereas, Educational Assessment refers, *grosso modo*, to three types of assessment: Assessment of learning, assessment for learning, and assessment as learning (Chappuis & Stiggins, 2002; MacMath et al, 2009). Hence, the use of multiple forms of assessment in addition to traditional tests within the view of enhancing learning (i.e. assessment for learning by adjusting the teaching learning process). EA might target the process (if formative) or the product (when summative). The term Educational Measurement is replaced by Educational Assessment when the scope is widened to enabling the teachers not only to infer the abilities and proficiencies of students (i.e. assessment of learning) but also to decide on the actions

to be taken in order to consolidate their weaknesses during the training before they sit for the final examination. Feedback (Wiggins, 1993) and self-assessment (Black, 1998) are also highly encouraged, because they foster and promote learning. In fact, self-assessment is deemed as a tool of learning, and assessment is then viewed as learning per se.

Educational measurement and educational assessment are subsequent alternatives to Psychometrics, because the recent studies on what Banta and Associates (2002) call ‘scholarly assessment’ have shown a shift in paradigm from psychometrics to a broader educational paradigm, and from a testing and examination culture to an assessment culture (e.g. Angelo & Cross, 1993; Gipps, 1994; Wiggins, 1993, 1998; Black, 1998; Torrance and Pryor, 1998; Janesick, 2001; Siggins, 2001; Valencia, 2002; Volante, 2006). Wiggins (1993) points out that the whole relationship between the tester and student must be changed, not only the form of tests, because the focus must be on performance against standards, not ranking students against each other. In addition to promoting the culture of assessment, these works and others (e.g. McAlpine, 2002; Chappuis & Stiggins, 2004; MacMath et al., 2009) tapped into the purposes of assessment, the principles of assessment and the steps and techniques of construction and design of reliable assessment tools, including the teacher-made tests and exams. This shift is due to the criticism leveled at the traditional classroom tests and the large-scale standardized tests in addition to the introduction of new forms of assessment with diversified purposes. Scholars advocating educational assessment examined a considerable number of assessment purposes: Educational assessment supports a continuous monitoring of teaching and learning; it provides information about the learners’ progress, teachers’ and schools’ effectiveness and accountability; it acts as a selection and certifying device; and it drives the improvement of the curriculum and teaching. As for the forms of assessment, they range from teacher assessment, standard tasks and coursework to records of achievement, oral and written tests and examinations in addition to the high-stakes standardized tests or the standards-based tests. Moreover, the discussions over the utility of these forms raised many technical issues, namely norm-referenced assessment, cohort-referenced

assessment, criterion-referenced assessment, standards-referenced assessment, domain-referenced assessment, field-referenced assessment (Hughes, 1989; Black, 1998), construct-referenced assessment (William, 1998) and ipsative-referenced or ipsotive (McAlpine, 2002); formative assessment and summative assessment; authentic assessment and performance assessment; informal assessment and formal assessment; process assessment and product assessment. All these types are useful for our study which deals with the procedures of the whole process of assessment.

According to Gipps (1994), Psychometrics is still one of the forms of assessment in schools alongside the other forms which may be clustered into three areas. These are (i) classroom-based assessment, (ii) large-scale assessment and (iii) cognition assessment. Psychometrics is the science which stemmed from the work on intelligence and intelligence testing and date back to the 1800s and early 1900s. Also worth mentioning are the works of Linquist (1951), Cronbach (1951), Thorndike (1971) Linn (1989), Tyler, and Brennan (2006) (as cited in Odendahl, 2011: 1-20; Brennan, accessed 2015). Building on the works of Glaser (1963) and Wood (1986), Gipps (1994) who attempted to develop a theory of educational assessment, believes that the assessment model developed within Psychometrics is interested in measuring fixed attributes of the individual; therefore, it is seen as model of limitation. This model is norm-referenced and has two other disadvantages. The first is linked to the assumption of '*universality*', i.e. the extrapolation of performance in a given ability. Technically, psychometric tests are not aced in terms of construct-related validity and consequential validity, because they do not consider all the aspects of the skills. Thus, the user of the score would find it difficult to know which aspect of the skill has been assessed. The second assumption is '*unidimensionality*' which stipulates that items in a test should measure a single underlying attribute even though some attributes are multi-dimensional.

Further, to paraphrase Wood (1986), as quoted by Gipps (1994), educational measurement differs from psychometrics in many respects:

1. It deals with the individual's own achievement, no comparison with others.

2. It focuses on competence, not on intelligence;
3. It does not consider educational settings only;
4. It looks for best rather than for typical performances;
5. It is restricted by the rules and regulations of standardized testing;
6. It embodies a constructive outlook on assessment by helping the individual to learn.

One type of educational assessment is cognitive assessment or the measurement of capacity in teaching and learning (Gipps, 1994; Osuala & Onwuagboke, 2014). According to Mondofacto (2006), it is “*the process of systematically gathering test scores and related data in order to make judgments about an individual's ability to perform various mental activities involved in the processing, acquisition, retention, conceptualisation, and organisation of sensory, perceptual, verbal, spatial, and psychomotor information* (as cited in Osuala & Onwuagboke, 2014: 178). Generally speaking, in classroom settings, there are three types of tests used in assessing student’s cognitive domain. These include:

- a) *General intelligence test: a test given to determine the individual’s general intellectual level at any given time. It measures the intelligent quotient (IQ) of an individual.*
- b) *Aptitude test/mental ability test: a test designed to measure or assess student’s potential or capability to succeed in a training programme leading to a specific profession. [...].*
- c) *Achievement test: a test designed to determine how much pupils have learnt from what skills they have been exposed to in a system.” (Osuala & Onwuagboke, 2014: 178)*

From 1900 onward, standardized testing was introduced in the USA both for college admissions and achievement purposes and it became customary (Odendahl, 2011). Currently, there are many calls (e.g. Messick, 1992; Gipps, 1994; Briggs, 1998; Stiggins, 2002; Valencia, 2002; Volante, 2006) to bridge the gap between standardized assessment and classroom-based assessment by developing a ‘hotch-potch’ system of mixed-methods of assessments to serve the purposes of classroom assessment and those of large-scale assessments of achievement, because the latter were construed on highly standardized procedures that had limiting, damaging and negative effects on teaching and the curricula as a whole in the USA, and relatively in the UK, in addition to other disadvantages. These drawbacks were addressed by Archbald and Newman

(1988), Wiggins (1993), Valencia (2002) and Volante (2006b). Their main arguments are that standardized testing:

- establishes a distance between the tester and the testee (Wiggins, 1993)
- supports unscrutinized policy-decisions that are based on measures with important psychometric limitations (Kane, 2002; Popham, 2001),
- acts as a form of educational and social control that undermines educators' professional autonomy (Canadian Teachers Federation, 1999)
- increases educators' feelings of lower self-efficacy in case of obtaining bad results that are due to inequitable resources (American Educational Research Association, 2000; Darling-Hammond, 2004),
- provides unrealistic results which widen the gap between minorities and white students by ignoring key factors known to affect student performance such as socio-economic status, language of origin, and the students' physical and/or emotional health (McNeil, 2000; Valencia & Villarreal, 2003),
- undermines the quality of education by ignoring the importance of divergent thinking, creativity, and intellectual work in general (Anderson & Postl, 2001; Glaser, Linn, & Bohrnstedt, 1997; Hess & Brigham, 2000; Wideen et al., 1997),
- narrows and distorts the curriculum by encouraging 'teaching to the test' techniques which take valuable time away from non-tested subjects, particularly in high-stakes tests (Kohn, 2002; Smith & Fey, 2000; Volante, 2004),
- gives little consideration to improving classroom practice (Stiggins, 2002; Falk, 1998), and
- inflicts damages on children's self-concept and leads to student disengagement (Meaghan & Casas, 2001; Miller & Tovey, 1996).

Angelo and Cross (1993: 4) see that classroom Assessment is "*an approach designed to help teachers find out what students are learning in the classroom and how well they are learning it. This approach is learner-centered, teacher-directed, mutually beneficial, formative, context-specific, ongoing, and firmly rooted in good practice*". Thus, classroom-based assessment is linked mainly with the terms 'formative assessment' and 'authentic assessment' but it is sometimes applied even in a 'summative assessment' such as teacher-made final exams. Summative assessment is of course applied in all the other types of assessment. It could be formative if it is combined with useful feedback, rather than only with numerical scores. Performance-based assessment is used interchangeably with authentic assessment (Janesick: 2001), especially in the USA. This type of assessment is based on performance tasks to show what a student can really do. Performance-based assessments are criterion-referenced, because they measure the amount of learning which happens during a given taught

content, course or curriculum, then the teaching would be adjusted to meet its objectives and goals. By contrast, large-scale tests and examinations are norm-referenced assessment and always summative without any feedback.

In the following section, we will deal with the fundamental principles of educational assessment. These principles play a crucial role in formal assessments such as final exams, but scholars advocate respecting them even in performance-based assessment.

### **1.1.3. Principles of Assessment**

There are five principles underlying assessment: validity, reliability, practicality, washback and authenticity. We will define these principles and discuss the extent to which they apply to quality assessment.

**1.1.3.A. Validity:** Validation is a process undertaken to scrutinize the assessment procedure for relevance, accuracy, consistency, usefulness, accessibility, and fairness (Odendahl, 2011). Validity has at least five facets (Hughes, 1989). It is related to content, criterion, construct, consequence and examinee judgment. In theory, validity refers to *“the extent to which an assessment measures what it purports to measure”* (Gipps, 1994: 2). Yet, as explained by many authors (Brown & Abeywickrama, 2010; Genesee & Upshur, 2001; Odendahl, 2011), this extent has been extended by recent research to include appropriateness, meaningfulness and usefulness of assessment purpose and the inferences made from assessment results.

As far as validity facets are concerned, content-related Validity, for instance, is defined by looking at *“the extent to which a test actually samples the subject matter about which conclusions are to be drawn”* (Brown & Abeywickrama, 2010: 354). Other researchers divide content validity into two types, relevance validity and coverage validity. The second type means that the test must sample the whole content of the unit or the course. The second facet which is criterion-related validity refers to *“the extent to which the linguistic criteria of the test (e.g., specified classroom*

*objectives) are measured and implied predetermined levels of performance are actually reached”* (Harmer, 1983: 354). In this regard, two types must be highlighted: Predictive validity of a given test refers to the extent to which its results are useful in gauging future performance; whereas, concurrent validity is concerned about how the test results correlate with another test which is assessing the same skill. That is to say, Concurrent validity refers to *“the extent to which results of a test are supported by other relatively recent performance beyond the test itself”* (Brown & Abeywickrama, 2010: 353). In fact, these two types are often combined to give evidence of criterion validity, because they both relate to predicting performance on some criterion either at the same time or in the future (Gipps, 1994; Brown & Abeywickrama, 2010). According to Brown and Abeywickrama, (2010: 354), construct validity refers to *“any theory, hypothesis, or model that attempts to explain observed phenomena in one’s universe of perceptions”*. Consequential validity refers to *“a test’s impact, including such considerations as its accuracy in measuring intended criteria, its effect on the preparation of test-takers, and the (intended and unintended) social consequences of a test’s interpretation and use”* (Brown & Abeywickrama, 2010: 353). The final aspect is face validity. Brown and Abeywickrama (2010) state that the face validity of any assessment is measured by the test-taker’s views on its fairness, relevance, and usefulness for improving learning. We will see how these principles relate to the interpreter trainees’ perceptions and views in chapter five.

**1.1.3.B. Reliability :** Scholars agree that reliability refers to the extent to which a test yields consistent and dependable results. Like validity, reliability is not a unitary concept (Brennan, 2001; McAlpine, 2002) but it has many types: inter-rater reliability, intra-rater reliability, student-related reliability, test reliability, test administration reliability, items’ internal consistency reliability, test-retest reliability, parallel forms reliability (Brown & Abeywickrama, 2010, Odendahl, 2011).

Indeed, both the examinee and the rater(s) might be a source of error. The learner’s ‘observed’ score might deviate from his/her ‘true’ score and lead a decreased student-related reliability due to the learner’s fatigue, anxiety, or any physical or

psychological factor. Also, if the exam is scored by one examiner, rater reliability or intra-rater reliability is calculated by looking at the consistency of his/her scores across all tests; whereas, Inter-rater reliability refers to the condition in which two or more scorers yield consistent scores for the same test. In addition, as far as the intrinsic features of the test are concerned, internal consistency reliability refers to the degree of relationship between the items of the test. And, test reliability refers to consistency among its different facets (e.g., instructions, item types, organization) in each test administration. But, test administration reliability refers to consistencies in the conditions in which the test is administered. Moreover, test-retest reliability is increased by the replication of the same test on different occasions to check the consistency of examinees' scores. Instead, another method would be parallel forms reliability in which multiple versions of the tests are administered (Brown & Abeywickrama, 2010; Odendahl, 2011).

The question which is worth investigation is whether a test is reliable for different levels of study? Another concern should be highlighted here is that not all the teachers master mathematics and statistics and it would be difficult for them to calculate the reliability coefficients. Even if they can, educational assessment is not interested in coefficients such as Cronbach's Coefficient Alpha and the Coefficients of Stability and equivalence (Odendahl, 2011). As a result, the reliability of their classroom exams has to be ensured by other means.

**1.1.3.C. Practicality** refers to *“the extent to which resources and time available to design, develop, and administer a test are manageable and feasible”* (Brown & Abeywickrama, 2010: 346). Fulcher (2010) points out that these resources are not always available for a well-planned programme of test development or for writing test progress and evaluation reports.

**1.1.3.D. Washback or Backwash** are terms which refer to the effect of testing or assessment on classroom teaching and learning. They can be positive or negative both on the learners and on the teachers, because they can be beneficial or harmful.

For example, if the content of the test is at variance with the course objectives, it is likely to be harmful (Hughes, 1989; Brown & Abeywickrama, 2010).

**1.1.3.E. Authenticity:** Archbald and Newman (1988) are believed to be the first to use this term in teaching. In the USA, the use of authentic assessment expressed the will to move away from the standardized, multiple-choice type tests towards an approach where the assessment task reflects the desired performance either in the classroom or a genuine context. Performance-based assessment aims at enhancing learning by modeling the real-life activities that the learners are encouraged to engage with such as written communication skills and problem-solving activities (Gipps, 1994: 11). Authentic assessments are “*assessments that emulate the performance that would be required of the student in real-life situations*” (SCASS/AEAC Glossary of Assessment Terms). In language assessment, authenticity refers to “*the degree of correspondence of the characteristics of a given language test task to the features of a target language task.*” (Brown & Abeywickrama, 2010: 346).

The next section is devoted to exploring all the steps which must be taken into account in the construction and design of the assessment tools.

## **1.2. Construction and Design of Quality Assessment Tools**

Practically speaking, assessment includes traditional tests and all the other alternative forms which are used to gauge students’ learning such as homework assignments, portfolios of accomplishments and running records, conferences and debates, journals and diaries, questionnaires and interviews, classroom projects, informal reading inventories, writing samples (Genesse & Upshur, 2001; Valencia, 2002) and many more. All these forms of assessment aim at achieving high quality outcomes. They must be constructed and designed taking into account, mainly, eight steps (determination of purpose, objectives, specifications, content and format, constructs, tasks, prompts and items, scoring). These are briefly presented below:

i. Determination of the purpose: Newton (2007) draws attention to the three different meanings underlying the term ‘assessment purpose’: the judgment level meaning, the decision level meaning and the impact level meaning. So, educational assessment can be used for summative purposes (accountability; certification) or for formative purposes (learning) (Gipps, 1994; Black, 1998; Black et al., 2005). In practice, assessors are frequently encouraged to shift their focus from accountability purposes to the purpose of enhancing learning (McAlpine, 2002; Valencia, 2002) through high quality assessment. Stiggins (2001) stresses that high quality is determined by the extent to which the assessment answers clearly the following five questions: What? Why? How? How much? How accurate it is?

The first two questions are relevant to the determination of the purpose. The first question is ‘What’ is the learning target of the assessment? The key to this question would be a clear and specific chunk of content which define both what to teach and what the students need to learn. The second question is ‘Why’ and the answer must be a clear and focused purpose for the assessment (Kusimo et al., accessed 2012). Therefore, it is worth noting that assessment may be used to diagnose, classify, select, place, predict, monitor, or change (Odendahl, 2011). A diagnostic test refers to a test that is designed to diagnose specified aspects of language. A placement test is a test meant to place a student into a particular level or section of a [language] curriculum or school. An achievement test refers to a test which is used to ascertain whether overall course objectives regarding the acquisition of the appropriate knowledge and skills have been met by the end of a given period of instruction (Hughes, 1989: 9-19; Brown & Abeywickrama, 2010: 346-351). All these assessments would be beneficial in assessing the adequacy of instructional plans and practices and help in adjusting them (Genesse & Upshur, 2001; Valencia: 2002).

ii. Determination of the objectives: the operational instructional objectives must be clear to the students and reflected in the assessment (Genesse & Upshur, 2001).

iii. Determination of the specifications (Blueprints Table): According to Brown and Abeywickrama (2010: 352), the term ‘specifications of a test or an exam’ refers to

the “*planned objectives, features, methods, and structure of a test*”. Wood (1993) pointed out that according to the purpose the Blueprints Table must include all the details such as the construct, questions, items, scoring scale, etc. Kusimo et al. (2012) stress that the assessors must answer the question: How appropriate is the match among the targets, the purpose and the methods used for the assessment. Specifications for process assessment are more difficult to be determined, because the attributes are not relevant to a well-known final product.

iv. Determination of the Test Content and Format: Odendahl (2011) reminds that the format of the questions in the tests has always been a source for the thorny issue of subjectivity in scoring since the introduction of questions which measure higher-order analytic skills such as reasoning and thinking in Chinese examinations by the Sung Dynasty (960-1279).

v. Description of the constructs: A construct refers to “*the specific definition of an ability, often not directly measurable (e.g., fluency) but which can be inferred from observation*” (Brown & Abeywickrama, 2010: 347). Fulcher (2010: 96) adds that “*constructs are the abilities of the learner that we believe underlie their test performance, but which we cannot directly observe*”. It is used in Psychometrics to label underlying skills or attributes. It is an explanatory device, because it is a theoretical construction about the nature of human behaviour (Gipps, 1994). Ward et al. (1996: 2) warn that “*A basic consideration in construct validity is to define the construct (or domain), and this is exactly where the debate in achievement testing has been for the past decade. There is no longer an agreement as to what should be covered on academic achievement tests, if there ever was.*”.

Description of the criteria and standards: Meeting criteria improves learning, because the students should be active participants in their own learning. Clearly, they perform better when they know what goals they are working towards, when they have the opportunity to examine models of excellence, and when they understand how their own performance compares to a set of established criteria (McTighe, 1996) and standards. Scholars such as Wiggins (1996), Guskey (2006), Stiggins (2007), Jung and Guskey (2007), Jung (2009) point out that a high-quality grading and reporting system

must have two basic characteristics: First, clearly articulated standards for student learning so that the meaning of a grade change from being a single overall assessment of learning (How did this student perform?) to a description of the student's performance on an explicit set of skills (How well did the student master the ability(ies)?). Second, standards must be based on three distinct types of learning criteria relating to product, process, and progress: First, product criteria describe students' knowledge and abilities at a specific level of achievement or proficiency as demonstrated by final examinations; final reports, projects, exhibits, or portfolios; or any other final assessment of learning. Second, process criteria describe students' behaviors (efforts; participation; classroom, homework and daily work habits, etc) in reaching a given level of achievement or proficiency. Third, progress criteria relate to students' improvement, gain from their learning experiences and educational growth. These criteria focus on comparing students' performances in time.

vi. Description of the tasks, prompts and the items. Quality tasks should be authentic, i.e. they should include real-world performances. Therefore, the prompts and items should be carefully selected to correspond both to students' level in terms of knowledge and skill and to the required standards in the industry. For instance, Fulcher (2010) points out descriptions of tasks, prompts and items depend on whether the test is traditional or online where the structure must be different.

vii. Determination of scoring and the grading methods: Scoring and grading methods should be clear to the audience. These are determined in tandem with assessment referencing, i.e. norm-referenced, cohort-referenced, domain-referenced, criterion-referenced, standards-referenced, or ipsative-referenced assessment, etc (McAlpine, 2002). Gipps (1994) stresses for instance that giving a single statistic score is misleading. It cannot replace the descriptive score which is encouraged by educational assessment, because the numerical score diminishes the value of the detailed picture that is required about the students' performance and achievement.

We will see in chapters 5 and 6 whether the steps of quality assessment described above have been accounted for in the practice of assessment of simultaneous

interpreting at the HAIT. Nevertheless, we will move now on to shed light on the main long-standing issues in assessment (Accountability, fairness, ethics, equity, educational outcomes).

### **1.3. Main Assessment Issues**

The main outstanding and controversial issues in assessment are related to accountability, fairness, ethics, equity and education outcomes.

**1.3.1. Accountability:** Government officials used testing to hold teachers and schools' principals accountable for the scores, because the learners' results reflect the effectiveness of the schools (Black, 1998; Odendahl, 2011). Thus, according to Gipps (1994), assessment specialists must strike a balance between teaching and assessment by designing accountability assessment tools which will provide good quality information about examinees' performance without distorting good teaching practice. Therefore, they would ensure fruitful learning which fosters higher-order skills, because teaching for the test would focus only on the basic skills. Hence, the call for mixed-methods to make classroom assessment more standards-based and the standardized tests more performance-based.

**1.3.2. Fairness, ethics and equity:** These issues are interrelated. For instance, Gipps and Stobart (2004, 2009) who addressed the key issues related to fairness in assessment considered equal educational opportunities as one of them, in addition to bias and validity. They stated that assessors must provide equal opportunities for the learners, be clear about what they assess, avoid the sources of bias and increase the validity of the procedure. Indeed, equity is focused on respecting the learners (Wiggins, 1993) and on providing equal opportunities by considering the students' personal experience and all the sociocultural factors which might be a source of difference between them such as gender, culture, language in addition to the other aspects of students' lives and identities such as ethnicity, age, social class, socioeconomic status, family history, physical attributes, learning styles, religion, nationality, etc (Gipps, 1994; Miller, 2006; Kusimo et al., 2012). Besides, the

International Association for Educational Assessment (2012) emphasizes the role of ethics by stressing all the rights of all the candidates, because it deems that the assessment ‘belongs’ to them, not to its administrators.

**1.3.3. Educational outcome:** Gipps (1994) points out that the learning goals must be attainable and the exemplars and the standards must be clearly specified, because students’ motivation is not enough to reach high-quality achievements. In addition, if the teachers use holistic criteria, the exemplars and standards will inform them about the nature of the multi-dimensional and complex skills being assessed. Moreover, these exemplars and standards will improve their consistency of scoring.

According to Gipps (1994), we must articulate the model of learning on which we are to base new developments in assessment. Psychometrics was based on a theory of intelligence, while multiple choice standardized tests were based on a behaviourist model of learning. Yet, Gipps does not describe the model upon which educational assessment is based. Does it suffice to say it “*must be based on our best understanding of theories of learning*”(Gipps, 1994: 4)?

#### **1.4. Interpreting Assessment and related fields**

Ostensibly, as mentioned above, the movement of educational assessment had an impact on Classroom Language Assessment (CLA) in the 1990s due to the negative results of Large-scale Language Assessment (LLA) which was based then on high-stakes testing only. LA means the assessment of the four language skills that the learner must have developed in order to prove his mastery of a language, i.e. listening, speaking, reading and writing, in addition to the intercultural communicative skills and the strategic skills. This area is well explored, so literature is abundant. For example, one may mention the seminal works of Harmer (1984), Bachman (1990), Hughes (1996), Genesee and Upshur (1996), Davies (2013), Brown and Abeywickrama (2010), Fulcher (2010) in addition to the tremendous number of articles which examined the implementation of educational assessment in the language classroom.

Yet, it is worth noting here that LA is different from IQA in many respects. What are the similarities and differences between them? Assessment of interpreting transcends the assessment of the speaking skill which is known in spoken language assessment. It is true that both of them assess oral production; nevertheless, Language Assessment is the assessment of oral language but Interpreting Quality Assessment is the assessment of Interpreting as process and as product. Speaking and interpreting do not involve the same processes and outputs. It is true that they both involve ‘Speech production’ but the skills and sub-skills which are required in this production are different as we will see below. For example, in her analysis of 11 interviews with professional interpreters whose experience exceeds 27 years, Albl-Mikasa (2014) identified four sub-competences for the ‘intuitive competence’ which is considered as a crucial competence in interpreting. These are: 1) Social and emotional competence, 2) Empathy and instinct, 3) Identification, 4) Touch of genius.

Scholars such as Sawyer (2004) tried to distinguish the profession of interpreting from other activities which involve the use of oral language by considering the nature of each of them. For example, he stated that in Europe the distinction between interpretation and other multilingual activities is reflected in the pedagogy of interpretation, which has separated itself from the field of foreign language teaching (Arjona 1984, as cited in Sawyer, 2004). He emphasized the fact that either in translator and interpreter education or in the profession, language is a means to an end, because language competence is a foundation upon which language transfer skills are built. But, in language teaching and literature programs, language remains the goal of study (Snell-Hornby, 1998; Hönig, 1995).

What is involved in the process of the spoken word production or the production of speech and discourse? What are the models of Spoken Language Assessment? Are they different from the models of IQA described earlier? Hughes (1989) and Bachman (1990) point out that the appropriate development and use of language tests for the purposes for which they are intended requires a clear definition of the abilities one wishes to measure and the means by which we observe and measure them. Thus, one

must be aware that the abilities required for speaking are different from the ones required for interpreting. In other words, testing communicative competence in relation to language proficiency is different from testing the communicative competence related to interpreting competence. That is to say, testing Spoken L1 or L2 is different from testing interpreting from L1 to L2 or vice versa.

The difference between LA and IQA can be illustrated further by highlighting it from the perspectives of Text Analysis, Discourse Analysis and Conversation Analysis, which is not always obvious. Pöchhacker (2004: 58) stated that “*the conceptual distinction between ‘text’ and ‘discourse’ is anything but clear, and is sometimes a matter of geolinguistic tradition and intellectual preference*”. Indeed, the use of the terms ‘text’ and ‘discourse’ are very confusing both in the translation and interpreting literature and in the Linguistics literature, that is why, as stated by Roy (2000), it is difficult to know what is meant by discourse analysis. Roy (2000) points out that the linguist Harris (1952) was the first to use the term ‘discourse’ and that discourse analysis, which sprouted in Linguistics from several branches, often refers to the analysis of ‘*language beyond the sentence*’. When Roy (2000) regarded Interpreting as a discourse process and said that the analysis examines how an interpreter manages the discourse process between two participants who do not speak the same language, she was referring to the community mode of interpreting, because in the simultaneous mode, the listener does not speak, he only listens. The discourse is mono-directional. Thus, from her perspective, discourse is seen as language as uttered by people who are engaged in a social interaction to accomplish a goal. Nevertheless, this kind of interaction is considered as a ‘dialogue’ or a ‘conversation’. Hence, this interaction might be the subject of study of Conversation Analysis and not Discourse Analysis.

In a nutshell, from a Discourse Analysis viewpoint, Language Assessment is concerned with how arguments are communicated; whereas, IQA looks into how equivalent speech is produced.

## **Summary & Conclusion**

Psychometrics, Educational Measurement, Educational Assessment refer to different sub-systems of the whole system of assessment, which is part of educational research. Besides, each assessment model reflects a learning model which was built on a separate theory of Learning. These models might be mixed, but their use must be justified by a clear rationale because they are underpinned by different philosophical and educational foundations.

All the forms of assessment, both traditional tests and alternative forms of educational assessment, aim to achieve high quality results. That is why researchers explored the principles of assessment that would ascertain valid, reliable, practical and authentic assessments of the real attainment and achievement of the learners. This issues triggered controversies as a result of the applications of the various forms of assessment in different socio-cultural contexts for different purposes.

The construction and design of teacher-made tests (e.g. mid-term tests, end-of-course exams) must follow a long validation procedure which touches upon the determination of: 1) Assessment Purpose; 2) Objectives; 3) Specifications; 4) Content and format; 5) Constructs; 6) Criteria and Standards; 7) Tasks, Prompts and Items; 8) Scoring and grading. In addition, the designer must consider the long-standing issues in the field of assessment which are related to accountability, comparability of the results, fairness, ethics, equity, educational outcomes.

Interpreting Quality Assessment is different from Language Assessment in many respects, and the literature on language assessment is not sufficient in IQA. A specific type of assessment is needed to account for the specificities of the field of Simultaneous Interpreting. In the next chapter, we will review the literature on IQA to gain insights and be able to advance research in this area.

**Chapter Two:**  
**A Critical Review of Works On**  
**Simultaneous Interpreting**  
**Quality Assessment**  
**(SIQA)**

This second chapter is a review of the most relevant literature to the issues of Interpreting Quality Assessment (IQA), in particular Simultaneous Interpreting Quality Assessment (SIQA). In the first section, we deal with the complexity of IQA to pave the way for a better understanding of what is at stake in SIQA. We will look into the definitions of IQA, then delve into its purpose and its theoretical frameworks. After that, we will examine the methodological considerations in IQA. In the second section, we will tackle the suggested taxonomy of SIQA models.

### **2.1. Complexity of Interpreting Quality Assessment (IQA)**

The concern with quality assessment in interpreting, dating back either to the late 1980's or according to Clifford (2005) to the 1990s, is more recent than the concern with TQA (Kopczynski, 1994; Shlesinger, 1997; Pöchhacker, 2004). Broadly speaking, quality in interpreting is still a moot point, because the field of interpreting encompasses many modes the input, process and output of which vary tremendously as we will see below. This brief overview of literature is meant to shed light on Quality Assessment of the different modes of Interpreting to pave the way for a better understanding and a more narrowed review of literature of QA in Simultaneous Interpreting below and in chapter three.

Generally speaking, Interpreting Quality Assessment refers to the assessment of the quality of the interpreters' performance through specific criteria, norms and standards relevant to the context from different perspectives. Interpreting Quality (IQ) is defined in the Encyclopedia of Applied Linguistics by Barranco-Droege et al (2012) as:

*“a multifaceted and complex concept resulting from a variety of communication situations and interpreting types, including conference, community, court, and signed language interpreting. It also comprises the views, motivations, and goals of the interpreters and all other stakeholders-speakers, event organizers, public institutions, trainers, and so on”.*(Online)

Thus, as pointed out by Collados Ais and Gile (2002), there is not one absolute quality but a set of qualities. Each assessor defines the ideal IQ from a specific standpoint via his own list of criteria, norms and professional standards. Consequently, the number of lists is not limited. For example, Collados Ais and Gile (2002) define IQ as the sum of the characteristics of many components of the interpreter's discourse such as information fidelity, linguistic accuracy, prosody and the quality of the voice at their optimum levels, because in reality it would be difficult to assemble some characteristics at their maximum levels. For instance, according to them, maximum fidelity means a waste of time and excellent style can be achieved at the expense of comprehension for some non-native speakers. That is why, quality which is a dependent variable can be measured through the success of communication between the interlocutors. Kalina (1998, 2005) defines IQ as appropriate strategic processing and its deterioration means that strategic processing is difficult, has broken down and does not lead to an adequate target text.

For Sawyer (2004), IQ means the development of expertise which can be achieved through addressing at least three challenges by the interpreter educators in order to enable the learner to develop expertise more rapidly and efficiently to the skill levels required at the top of the language industry. These challenges are linked to (1) Curriculum design for the new programmes and for updating of the existing ones. (2) The appropriate, meaningful, and useful assessment of students' performances at all stages of the curriculum. (3) The optimization of the academic environments and instructional settings through efficient syllabus design and lesson planning, the use of cutting-edge teaching methodology, innovative technology, proven classroom management techniques, and other pedagogically sound practices. Additionally, Sawyer (2004) believes that research on quality informs assessment and helps in the development of constructs, because in the absence of clear assessment criteria there are many concerns about the inadequacy of assessment practices which are based on 'unsystematic hit-and-miss' methods of performance assessment.

### 2.1.1. Purpose and Theoretical Frameworks of IQA

It is worth noting that IQA in higher education settings is under-researched, because the literature which considers the implementation of the latest findings in testing and assessment in IQA with reference to an educational environment is rare. The bulk of literature refers to theoretical speculations or to the professional realities of the job. Sawyer (2004) seems to be the only scholar who wrote a book in which he dealt with the integrative role of assessment combined with curriculum design and implementation in one educational institution, the Graduate School of Translation and Interpreting (GSTI) of the Monterey Institute for International Studies in the USA. Indeed, the number of book chapters, articles and presentations, either in books, peer-reviewed journals or in conference proceedings (e.g. Almuñécar, 2001, 2011; EMCI, Graz 2006, 2008) urging for more research, in particular empirical research, in these areas is considerable.

All the scholars interested in the field of assessment in interpreting stress the fact that *'quality'* is a central concept in translation and interpreting studies alike (Gile, 1995, 2009; Pöchhacker, 2001, 2002, 2004, 2013; Clifford, 2005; Kalina, 2002, 2005; Grbić, 2008; Angelelli & Jacobson, 2009; Zwischenberger, 2010; Napier 2012; Kellett Bidoli, 2013; Becerra et al., 2013; Barranco-Droege et al., 2013; Tsagari & Van Deemter, 2013; Schwieter & Ferriera, 2014; Zwischenberger & Behr, 2015), but there is neither consensus on what it refers to (Product, process, interaction,...) nor on its criteria, norms and standards. Indeed, broadly speaking, the purpose of IQA is to investigate, theorize and model the interpreting activity taking into account its multiple modes and their various forms in addition to the various and complex dimensions of quality.

Pöchhacker (2001, 2002, 2004), Collados Ais and Gile (2002), Kellett Bidoli (2013) have dealt with the paradigms, approaches, theories and the methodological strategies and techniques, for both experimental and empirical studies, that have been used so far in IQA. They have sketched out the progression of the scholars' reasoning,

the findings of these studies and their flaws. It is also worth noting that some of these scholars use the term conference interpreting which does not refer solely to simultaneous interpreting but includes consecutive interpreting as well. Others have used just the generic term interpreting which means that they refer to the whole 'activity'. Also, there are many definitions for Simultaneous Interpreting. Gerver et al. (1989) define it as the immediate transformation of spoken textual information; whereas, its transformation after some delay is termed Consecutive Interpreting.

Pöchhacker (2004) reviewed the paradigms of Interpreting studies and showed that there was a shift from pure speculation to more interdisciplinary research. In fact, his contribution is an attempt to lay insightful foundations for future research in IQ, because he explained that the studies under these paradigms do not target explicitly the investigation of quality. Then, he drew a table summarizing these paradigms according to their memes, influential theory or model, leading members, paradigm cases, interdisciplinary orientation, methodological strategy – fieldwork, survey, experiment - and, finally, the interpreting mode. For instance, to shed light on them in short here, one might cite the Paris School Paradigm, which was built on the Interpretive theory of Translation (IT) (Seleskovitch, 1975; Lederer, 1981), the Neuro-Linguistic Paradigm (NL) (Fabbro et al., 1990 and Fabbro & Gran, 1994), the Cognitive Processing Paradigm (CP) (Gerver, 1976; Moser-Mercer, 1978; Gile, 1985, 1999), the target-text translation-theoretical approach, the target-oriented text production, which is referred to as the (TT) paradigm (Shlesinger, 1989; Pöchhacker, 1994), and finally, the dialogic discourse-based interaction paradigm or (DI) paradigm (Wadensjö, 1998; Roy, 2000).

It is worth noting that most interpreting scholars who dealt with IQA discuss the criteria of all the modes or some of them wholesale (For example, Pöchhacker 2001, 2002) did not draw a line between the different forms of interpreting when highlighting the different quality-oriented methodological approaches used by researchers to investigate this issue both experimentally and empirically. So, he cited,

with an in-depth analysis of some studies, survey research, experimentation, corpus-based observation and case study as applied in Conference and Community Interpreting. Additionally, Pöchhacker (2004) gave a broader account of all the psycholinguistic and the socio-cultural approaches used in the discipline of IS. That is why we believe that they are irrelevant in a review which is focused on the literature that is intended to be connected explicitly with the study of IQ. Nonetheless, if we take for granted here that all the studies inform in a way or another research on IQA, one might consider not the approaches but the '*memes*' introduced by Pöchhacker (2004: 51-61). In general, the term meme refers to ideas, practices, creations, inventions, metaphors and 'ways of seeing' which drove the evolution of thinking about a given phenomenon. Drawing on Chesterman's (1997) memes in Translation Studies, Pöchhacker introduced five '*memes*' which are connected within a matrix of four conceptual dimensions: language, cognition, interaction and culture. These four concepts are separate poles but can combine to form dimensions such as language-culture, language-cognition and cognition-culture. The five '*memes*' which are processing information, verbal transfer, making sense, text/discourse production and mediation represent the key ideas that geared theoretical inquiry into interpreting as activity, interaction and communication. These '*memes*' reflect scholars' priorities in research. For instance, the '*meme*' making sense which conceptualizes the interpreter's task as grasping the intended meaning and expressing it to the listener in a different language was taken for granted by the researchers who investigated the '*memes*' of text/discourse production and mediation. We will see the impact of these '*memes*' on teachers later when we discuss the findings of our empirical study.

Collados Ais and Gile (2002) reviewed the evolution of empirical studies on interpreting quality which aimed at measuring the value and relative impact of each of the different criteria. The first studies showed that there is consensus on the importance of some criteria among the interpreting community but did not determine their relative weight on quality for the different groups of users. Besides, they explained that empirical studies can be divided into two big '*families*': Studies on

expectations (e.g. Büler's 1986 and Kurz's 1989 surveys), and those on evaluation (e.g. Gile's 1990 study). The first category focuses on the hypothetical expectations of the different users; whereas, the second deals with the *a posteriori* real evaluative reactions of all the concerned parties, namely the interpreters' auditors, the events' managers, the interpreters, the technicians and so on. Their reactions reflect their own experience but also their *a priori* conceptualization and mental representations of quality. So, Collados Ais and Gile reviewed the questionnaire-based works of Meak (1990), Marrone (1993), Vuorikoski (1993), Mack and Cattaruzza (1995), Kopczynski (1994), Kurz and Pöchhacker (1995), and some unpublished works such as Andres (2000) and Doerflinger (2001). They also reviewed the interview-based study conducted by research commission of the International Association of Conference Interpreters (AIIC). It is worth mentioning that they did not distinguish between experimental and non-experimental studies. Some of the studies that they reviewed are not empirical but experimental ones (e.g. Gile, 1984, 1995; Collados Ais, 1998; Garzone, 2000; Mazza, 2000; Shlesinger, 1994; Cheung, 2000), because the researchers conducted them through the manipulation of some parameters; therefore, they do not represent real contexts. Yet, Collados Ais and Gile's findings are quite insightful. They concluded that quality is a phenomenon depending on many variables; it is not a holistic entity the knowledge of which is the property of the interpreter; there might be some discrepancies between the expectations and the real reactions on the ground; there is difficulty in measuring the real reactions and in extrapolating the expectations. Recently, a similar review of literature on users' expectations versus genuine evaluation was published by De Grogoris (2015) in which the assessment criteria of quality were divided into content-related and form-related ones and compared on that basis. In light of the limits of expectations, a call for a gestaltic model of perception was launched. We will deal with users' criteria in-depth in Chapter Three.

A more synthetic study about the concept of quality is Kellett Bidoli's (2013) review in which she gives account of the evolution of reasoning in quality assessment

of both simultaneous and consecutive interpreting. She states that the definition of the quality of interpretation is much more elusive than that of conference interpreting because quality is deemed to be a degree of excellence, a relative measure, which is perceived by each individual in a unique manner. She explains that prior to the 1980s researchers focused on the interpreters' mental processes neglecting the receivers' views and focused in the beginning only on the linguistic features of verbal communication. Gradually, kinesics, prosody and proxemics, which are the main areas of research in non-verbal communication, attracted their attention, because these paralinguistic features became increasingly recognized as essential components of speech acts. In addition, she extracted data from the literature and tried to provide a classification of the quality criteria by putting them into two categories. The first category concerns the factors which influence the quality of the 'input' and the second concerns those influencing the output. She surveyed 'who did what' by drawing tables where she included most of the quality-related criteria and referred to the researchers who investigated them either as criteria for the input or output both in consecutive and simultaneous interpreting. The entries for these criteria are as follows: Environmental factors, experience-related factors, inter-personal/social factors, linguistic features, non-verbal features, physical and mental factors, gender, prosodic features, situational factors, task-related factors, technical features, textual features and, finally, time factors.

Despite the fact that the recent works of Becerra et al. (2013); Barranco-Droege et al. (2013); Tsagari and Van Deemter (2013); Schwieter and Ferriera (2014) and Zwischenberger and Behr (2015) considered the concepts of competence (Kalina, 2000, PACTE, 2008, Albl-Mikasa, 2012; Klimkowski, 2015), although limited from the angle of evaluation (Louis, 2008), and expertise (Ericsson, 2000; Liu et al. 2004; Sawyer, 2004; Liu, 2009; Albl-Mikasa, 2013), they did not deepen research on quality further, because they neither inform us on how to design assessment tools of distinct educational constructs and criteria of interpreting quality nor on how the exit profile of

the trainees should be constructed and benchmarked to meet the global professional standards and norms (Pöchhacker, 2002, 2004, 2012, 2013).

### **2.1.2. Methodological Considerations in IQA:**

First of all, as far as the methodology of research on IQA is concerned, few studies explore the theoretical implications of the different notions of quality. Grbić (2008) explores quality as a social construct and focuses on the multi-perspectivity of some quality benchmarks. She acknowledges that quality is relevant to the individual assessing it in addition to the context of situation, then stresses the fact that the relativity of quality as an attribute pertaining to an object extends to the benchmarks used to gauge it in three contexts: training, professional practice and interpreting research. She builds her discussion on two heuristic concepts: social construct and social system. In addition, she discusses some divergent definitions of interpreting quality in this framework. Her point is that quality is, as a social construct, subjective like any other element of our reality. She considers quality in the previously-mentioned contexts as an object in distinct social systems which are created by their members due to their aligned views regarding their reality. However, she admits that these systems are interconnected and influence each other.

The interaction of the research system and the professional system are well depicted in the invaluable contributions of Pöchhacker's (2001). In this straightforward study, he surveys the state-of-the-art in interpreting studies in search of conceptual and methodological tools for the empirical study and assessment of quality. Eventhough his review of literature covers conference and community interpreting as mentioned above, he tried to connect the dots by providing a general schema but very illuminating one that would help any researcher to situate himself in the right assessment standpoint and to take into account the necessary methodological considerations and precautions in order to get unbiased results.

With reference to some exemplary studies on quality in conference interpreting (e.g. Gile, 1991; Moser-Mercer, 1996; Shlesinger, 1997; Kahane, 2000) and community interpreting (e.g. the First BABELEA Conference on Community Interpreting, held in Vienna in early November 1999), Pöchhacker sketched the basic assumptions and insights regarding assessment perspectives and quality criteria which can be applied along the entire spectrum of interpreting activity (See his model in Diagram 1 below). He tried to model the relationships between the various positions and perspectives by depicting them within a rectangle in a trapezoid. According to him, quality is assessed from the perspectives of the interactants directly involved in the communicative event of text production/reception, that is to say the interpreter, the speaker who is the source text producer (ST-P) and the listener who is the ST-Receiver (ST-R). But, It is also assessed from the perspectives of the Client (employer) and the Interpreter's Colleague (fellow interpreter/team member). Additionally, he highlights two deep analytical distinctions underlying the study of quality in interpreting. Firstly, the researcher, as an 'external observer', may investigate the various actors' attitudes, needs and views either with regard to an abstract (hypothetical or experienced) interpreting situation or with regard to a concrete communicative event. Secondly, research on quality in a concrete event may be either product-oriented by focusing on the recordable product or interaction-oriented by looking into the overall process of communicative interaction.

It is evident that each type of Interpreting requires specific skills and sub-skills and is practiced in a specific ecology. Thus, one of the criticisms that we can give to most of the previous studies or the subsequent ones (e.g., Schwieter & Ferreira, 2014) in this field is their holistic reviews and approaches which do not help much in deepening reflection over a more psychometric assessment as advocated by Clifford (2005) for example or a more educational assessment which considers all the insights of educational research.

Hence, a crucial consideration in IQA lies in distinguishing the different modes of interpreting before any evaluation endeavour. Thus, it is worth noting that Pöchhacker's suggestion is not detailed, because there are two different modes in conference interpreting which are Simultaneous Interpreting and Consecutive Interpreting. In addition, there are many modes in community interpreting ('USA', or Public Service Interpreting 'UK') which span from Legal Interpreting, Court Interpreting and Medical Interpreting to Police Interpreting. Community Interpreting can be done face-to-face, over-the-phone or via tele-conferencing. Besides, Liaison Interpreting, Whispered Interpreting 'Chuchotage' and Dialogue interpreting are just similar modes to Community Interpreting, because they are performed differently and may be used in SI and Consecutive Interpreting settings. Moreover, other modes of interpreting include Media Interpreting and Business Interpreting, Sight Interpreting and Signed-language Interpreting.

What are the different skills? Consecutive Interpreting for instance requires the use of memory and note-taking skills. As mentioned above, some modes require intercultural communicative skills and the use of technology. Overall, the degree of specialization which is required varies. As a result, assessment of these modes is not commensurable and should not be seen wholesale. Therefore, theories would not suffice, and here comes the role of models as we will see in the next section.

## **2.2. A taxonomy of SIQA Models:**

There are two types of SIQA models: Academe-based models and industry-based models. However, not all the models described below are designed for assessment purposes or for evaluation in educational settings. Still, we believe that they are relevant to our research, because they help in depicting the elements of the interpreting activity. It is up to the assessor to build his personal assessment model by considering his own context and assessment situation. It is worth reminding that sometimes these models encompass both modes of conference interpreting (i.e., Simultaneous and consecutive interpreting).

### 2.2.1. Academe-based Models:

They can be divided into four types of models: The prescriptive models, the descriptive models, the explanatory models and the didactic models.

The early interpretive models such as Herbert's (1952) and Seleskovitch's (1962, 1985) are prescriptive, because they focused on translational processes only without delving deeper into the description of the real cognitive processes. Herbert asserted that interpreting consists of three stages: understanding, conversion and delivery. Seleskovitch claimed that interpreting means '*deverbalisation*' and expression of '*sense*' and not '*transcoding*' of the meaning of words, that is to say the interpreter's understanding of the message is dissociated from any linguistic form in cognitive memory.

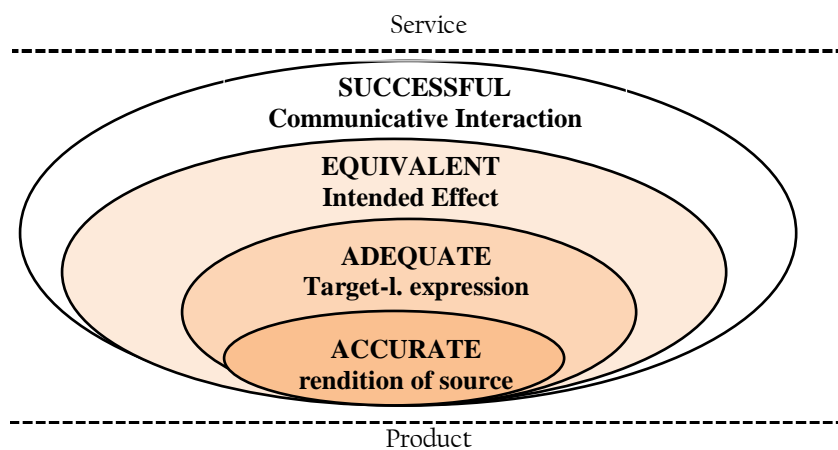
In addition to his discussion of Herbert's and Seleskovitch's works, Pöchhacker (2004) tackled the two other categories, the descriptive and the explanatory models. He explains that modeling in interpreting could be seen at various levels: anthropological, socio-professional, institutional, interactional, textual, cognitive and neural. Yet, he confirms that there is no model to date which has modeled interpreting from all these vantage points to describe and explain the memes of interpreting. Indeed, his comprehensive survey includes the interaction-oriented models and the process-oriented models of interpreting which were introduced from the simplest to the more complex ones. Nevertheless, we will focus on the models relating to conference interpreting only.

To present his own interpreting quality model, which is descriptive, Pöchhacker (2001, 2002) starts by recalling and confirming the fact that the invariable core criteria of quality for most of the researchers are accuracy, clarity 'and/or' fidelity. Nonetheless, for him, the various sets of criteria underlying quality assessment in interpreting pertain to different aspects or even conceptions of the interpreter's task,

ranging from text processing to communicative action for a certain purpose and effect and, most generally, to the systemic function of facilitating communicative interaction.

His holistic model (2002: 96) describes in a gradual way the quality standards which are ranging from a lexico-semantic core to a socio-pragmatic sphere of interaction by using specific adjectives. Hence, according to him, the rendition of the source must be *'accurate'*, the target language expression must be *'adequate'*, the intended effect must be *'equivalent'* and, finally, the communicative interaction must be *'successful'*. We would be interested in looking at teachers' perception of these levels and their instruments in turning it into meaningful assessment for the students.

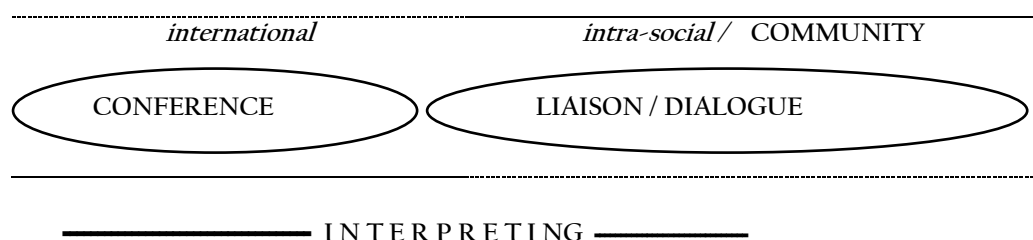
Diagram 1. Quality standards for the product and service of interpreting.  
(Pöchhacker, 2002: 97)



Pöchhacker (2002) complemented this model by suggesting a conceptual spectrum of interpreting which depicts all the modes in two main areas. He suggests that interpreting should be viewed as a conceptual spectrum of different (proto)types of activity which are distinguished by the extent to which they are governed by the constraints of a particular socio-cultural environment or institution. Thus, according to him, the concept of interpreting can be modelled, as shown in Diagram 2 below, in a continuum covering international conference interpreting and community interpreting. It is a bi-polar, inter- to intrasocietal continuum. Pöchhacker (2002) stresses that his model acknowledges the existence of middle grounds, forms and fuzzy boundaries, but

it is not concerned with the boundaries between the different types, and the ‘grey areas’. Rather, it calls for a ‘unified concept’ of interpreting as a sociocommunicative practice. However, this conceptualization is not helpful in classroom assessment where each module is taught separately with some pinpointed educational objectives.

Diagram 2. Conceptual spectrum of interpreting. (Pöchhacker, 2002: 96)



From a ‘quality perspective’, in order to deepen our understanding, we will look below at some explanatory and didactic models which are applied to conference interpreting only, that is SI and CI, because Pöchhacker’s model includes even community interpreting. Hence, one might cite Moser-Mercer’s (1978) Skills model, Gile’s (1995) Efforts Model and Setton’s (1998) Meaning-Assembly Model, which are considered as the main explanatory models in SI (Pöchhacker, 2004). In addition, we will look at some models which are more pedagogy-focused and are more relevant to this study, that is to say Stenzl’s (1983) model and Kutz’s (2010) Leipzig Competence Model (LKM).

Moser-Mercer’s (1978, 1979, 2008) Skills model which focused on memory structures and processing operations in simultaneous interpreting is based on a psycholinguistic model of speech comprehension. The main features of this model are the search for the ‘conceptual base’ and the construction of a prelinguistic meaning structure with the help of various types of knowledge (general knowledge, conceptual network, contextual knowledge, semantic knowledge, schematic knowledge, strategic knowledge). The conceptual meaning base then serves to activate target language elements for syntactic and semantic word and phrase processing on the way to output

articulation. The model discusses a number of decision-making skills, such as prediction and anticipation, which affect the process from Source Language identification to the activation of target-language elements. This model “*assumes a high degree of interaction between bottom-up and top-down processes and discusses trade-offs between the operations or stages competing for available processing capacity* (Pöchhacker, 2004: 101).

The Effort models of interpreting formulated by Gile (1985, 1997/2002, 2009). assumes three basic efforts, labeled “listening and analysis” (L), “production” (P), and “memory” (M). Gile (1985) originally used his effort model of simultaneous interpreting to express the basic tenet that there is only a limited amount of mental “energy” (or processing capacity) available for the interpreter’s processing effort, and that the sum of the three efforts must not exceed the interpreter’s processing capacity:  
 $(L + P + M) < \text{Capacity}$ .

In subsequent refinements of the model, a “coordination effort” (C) was added, and the relationships between the model components were expressed in a set of formulas and relationships as follows (Gile 1997/2002: 165):

*‘Simultaneous interpreting modeled as a process consisting of the three main efforts plus a coordination effort.’*

$$(1) SI = L + P + M + C$$

*‘Total processing capacity requirements are a (not necessarily arithmetic) sum of individual processing capacity requirements.’*

$$(2) TR = LR + MR + PR + CR$$

*‘The capacity available for each effort must be equal to or larger than its and requirements for the task at hand.’*

$$(3) LA \geq LR, (4) MA \geq MR, (5) PA \geq PR, (6) CA \geq CR$$

Setton’s (1998) Cognitive-Pragmatic Model brings together linguistic and cognitive approaches to give an account of discourse processing in SI. His model depicts processes of meaning assembly in SI in a Relevance Theory framework with particular emphasis on pragmatic factors. This model addresses mainly aspects of comprehension, memory and production in SI.

To introduce her model in 1983, Stenzl (as cited in Pöchhacker, 2004) admits that developing a detailed overall model of SI which involves the speaker, interpreter and target text receiver is premature and points out that her work is one of the first attempts to provide a framework for interpreting models. Besides, it was among the first models to use insights from text theory and translation theory to give an account of the communicative information flow in SI which centers on text processing. This model, which is adapted from Stein's model, takes into account the socio-cultural contexts and the situational and textual knowledge of these interactants who process the linguistic, the paralinguistic and the extralinguistic elements (auditory and visual) of the exchanged messages in fifteen stages. The model acknowledges that these stages might interfere and occur simultaneously.

According to Kutz's (2010, as cited in Albl-Mikasa, 2014: 51-81) Leipzig Competence Model (LKM), in interpreting:

*“the aim of competence building is to proceed from the acquisition of conscious knowledge of the task or activity (“Handlungswissen”) via routine abilities (“Handlungsfähigkeiten”) to unconscious skills (“Handlungsfertigkeiten”). The process is compared to learning to drive a car, which, after starting out from carefully thought-out deliberate moves, turns into an automatic routine practice.”*

### **2.2.2. Industry-based Models:**

We will focus on few models that correspond to novice-level expertise. The parameters mentioned in these models are useful; we would benefit from them to check whether the tests and exams of the HAIT reflect this kind of parameters in a planned way and see if the trainee interpreters take them into account in their performance, especially in the achievement tests, if any. Firstly, one may cite Tseng's (1992) model of the professionalization process, which describes the stages from 'market disorder' to 'professional autonomy'. It has been applied in various forms of interpreting in many countries. Professionalisation in institutional contexts has been also modeled by Ozolins (2000) and Agger-Gupta (2001) (as cited in Pöchhacker,

2004). Secondly, Kalina (1998, 2002, 2005) does not draw a model but discusses how the quality issue in conference interpreting can be approached using a scheme for quality assurance. She states that participants in conferences have different roles and not always the same preferences; interpreters depend on speakers and may have to work for heterogeneous audiences. So, on the basis of a model of mediated multilingual conference communication, interpreting processes and conditions can be analysed with the aid of a list of parameters; the list can be used by clients and conference interpreters for their overall and personal quality assurance. The figures below show the dimensions of quality and the factors which might affect it.

Figure 1. Dimensions of interpreters' output quality (Kalina, 2002: 125)

<b>Semantic content</b>	<b>Linguistic performance</b>	<b>Presentation</b>
Consistency logic, coherence completeness accurateness unambiguity clarity reliability	grammatical correctness adherence to TL norms comprehensibility stylistic adequacy terminological adequacy discretion lack of disturbances	voice quality articulation public speaking discipline simultaneity technical mastery conduct

Figure 2. Factors relevant to interpreting quality (Kalina, 2002: 126)

<b>1. Pre-process prerequisites</b>	<b>3. In-process requirements</b>
skills and competences contract specifications task definition preparation	knowledge and presuppositions conditions of ST presentation target language requirements interactional competence
<b>2. Peri-process conditions</b>	<b>4. Post-process efforts</b>
number of participants working languages technical equipment booth position team strength, composition working hours, event duration language combinations relay quantity/quality availability of documents information on proceedings	terminological follow-up documentation quality control further training specialisation adaptation to technical progress

Albi-Mikasa (2012, 2014) added a para-process dimension which refers to business skills, because most of the time the interpreters are required to manage their business.

## **Summary & Conclusion**

This second chapter is a review of the most relevant literature on the issues of Interpreting Quality Assessment (IQA). In the first section, we showed the complexity of IQA to pave the way for a better understanding of what is at stake in SIQA. Then, we delved into the purpose of IQA, its theoretical frameworks and methodological considerations. In the second section, we tackled the suggested taxonomy of IQA models both in academe and industry.

We learnt from this chapter that IQA requires a hard work on methodology in order to avoid bias in research because of the multiple forms that are distinguished along the spectrum. In addition, modeling interpreting in academe and in industry turned to be a nightmare due this complexity and the related norms and standards in each area.

It is worth pointing out that we tried to focus only on the models which deal with conference interpreting, particularly with SI; namely Pöchhacker's (2002) model, Moser-Mercer's (1978, 2008) model, Gile's (1985, 1997, 2002, 2009) model and Setton's (1998) model. Unfortunately, there is scarcity in the models which are didactics-oriented, so we relied only on Stenzl's (1983) model and Kutz's (2010) Leibzig Competence Model.

As far as Industry-based models are concerned, we found Tseng's (1992) model of the professionalization process, Ozolins's (2000) and Agger-Gupta's (2001) models of professionalisation in institutional contexts in addition to Kalina's (2002) scheme of interpreting quality dimensions and factors relevant to grasping the required standards in the development of expertise in the industry.

**Chapter Three:**  
**Criteria, Norms & Standards**  
**of Simultaneous Interpreting**  
**Quality Assessment**

This third chapter deals with the criteria, norms and standards of SIQA. It is divided into three sections. In the first section, we will try to reach at least an operational definition of SIQA. In the second section, we will tackle the educational criteria for SIQA of trainee interpreters for aptitude, ability and achievement. In the last section, we will delve into the industry-based criteria, norms and standards of SIQA from different perspectives: self and peers; certification and accreditation; employers and clients; end-users.

### **2.1. Definitions of SIQA:**

There is no agreed or ready-made definition neither of Simultaneous Interpreting Quality (SIQ) nor of Simultaneous Interpreting Quality Assessment (SIQA), and there is ‘no consensus’ on what these terms refer to. The practitioners, trainers and researchers alike focus in their assessment solely on some aspects of quality from different perspectives by considering specific variables and neglecting other aspects and variables due to their relative contexts and narrow angles of view. One might stress that the term SIQ is vague, relative and complex and the number of factors, variables and criteria included in its assessment is huge as we will see below.

Kellett Bidoli (2013) provided a succinct and somehow chronological review of the works of many authors, such as Barik (1969, 1971), Galli (1990), Altman (1994), Gerver (1971, 1976) in addition to Keiser (1978), McIlvan Parsons (1978), Dornic (1978) and many others, through which she clarified how each one perceives interpreting quality. Kellett Bidoli classified the studies chosen for her review according to their approach; therefore, she commented on their neglect of Non-Verbal Communication features, which were stressed by Laver (1980, 1991), Poyatos (1993), Shlesinger (1994) and Daro (1994). Then, she presented their objectives, variables and criteria in addition to showing the weaknesses of each approach. Kellett Bidoli (op. cit) stressed the relevance of Barik’s (1969, 1971) seminal study on the quality of SI, because it paved the way for error analysis studies in interpreting, which we will talk about subsequently (Gerver, 1971; Galli, 1990; Altman, 1994; Agosi, 1995; Russo & Rucci, 1997). In fact, Barik did not attempt to evaluate the ‘goodness’ or the overall

quality of the interpretation but used a ‘coding scheme’ to quantify the extent of departure of the interpretation from the original text in addition to considering the nature of the departures (omission, addition, substitution, errors). However, the delivery factors such as the voice intonation and the appropriateness of pausing were not considered.

Then, Kellett Bidoli (2013) tried to show the different output errors which were analysed by Galli (1990), Altman (1994) and Gerver (1971a, 1971b, 1976), because these errors were seen as a hindrance to quality in different contexts. Firstly, in a professional milieu, according to three variables which are the direction of language, the type of text and the speed of the ST, Galli (1990) tried to identify ‘departures’ (omissions, substitutions and additions) from the ST to TT by three professionals at a medical conference with the intent of discovering how the interpretation changed. Secondly, referring to educational settings, Altman (1994) took error to mean ‘*inaccuracies*’ leading to lack of fluency and constituting an obstacle to communication such as the omissions, additions, inaccurate renditions of text items and distortions of longer phrases and suggested that it would be possible to predict and forestall the production of errors by trainee interpreters through analysis and classification of the most frequently occurring ones. Gerver (1971a, 1971b, 1976) focused on the detrimental effect of poor listening conditions, noise and anxiety on the quality of the output which was analysed for errors, omissions, corrections, ear-voice span and temporal characteristics.

Moreover, Kellett Bidoli (2013) reviewed many stress-related studies. To mention just few of them, Keizer (1978) highlighted the issue of candidates’ stamina during selection or training, McIlvane Parsons (1978) outlined time and environmental (booth) variables, and Dornic (1978) delved into the deterioration of language. In addition, years later, there were the studies of Cooper et al. (1982), Klonowicz (1994), Agosti (1995) who advanced this line of analysis. Cooper identified four sources of stress: physical environmental factors, task-related factors, interpersonal factors and the home/work interface. Klonowicz stressed the fact that the human being is a limited

processor who responds to varied degrees of load. Agosti, adopting the hypnosis approach, examined the impact of prosodic features on emotional states.

Quality in interpreting is also guaranteed through certification. Stejskal (2003, as cited in Chan, 2011) pointed out that there are three types of certification in the field of translation and interpreting: (1) by a professional association, (2) by a government, and (3) by an academic institution. Below, we will deal with these types.

What is worth pointing out here is that in this research we will adopt an operational definition which sees SIQA as the assessment of the quality of Simultaneous Interpreting in light of different constructs, criteria, professional standards and norms from a given perspective in a given socio-cultural context. Also, the present study focuses on the educational criteria from an insider perspective, that is to say that the researcher's own experience in this field as a novice interpreter is exploited to shed light on the areas which are dealt with in the literature on educational assessment in general. This will allow us to understand and evaluate the data gathered empirically.

## **2.2. Educational Criteria for SIQA of Trainee Interpreters**

In this regard, let us simplify things in order to have a clearer and wider view, because even in the educational context the researcher must pay heed to many considerations. First, It is worth pointing out that we are referring to the assessment of student interpreters and not to their trainers' assessment. That is to say, we are not talking about how to assess educational interpreters who might be in training or for the sake of professional development, but about the trainee interpreters who are enrolled in an academic programme to qualify as interpreters or get a postgraduate degree in interpreting. Their contexts and their degree of expertise are different.

Sawyer (2004) points out that assessment constructs in interpreter education have not been precisely defined. Kutz (1997, as cited in Sawyer, 2004), for example, does not provide any criteria to diminish measurement errors, but lists a few broad areas for performance assessment: overall impression (behavior), information content, language

use, and a summary statement in reference to the interpretation assignment. However, the literature focuses more on the criteria than the constructs. For instance, Riccardi (2001, as cited in Sawyer, 2004) defines four macrocriteria for the evaluation of quality in professional situations. These are equivalence, precision, appropriateness, and functionality. In educational settings, she defines fourteen microcriteria for both summative and formative assessment. For instance, register, omissions, deviations in content, successful solutions. These categories are used as scoring criteria (none, some, many) for a feedback sheet.

Narrowing the research down to focus solely on educational criteria does not leave this research endeavour without ‘problems’ as there are a lot of criteria to be considered combined with a great deal of factors and variables that the teachers take into account when assessing the trainees. Their assessment depends on the context, i.e. educational setting, teaching environment and level of progress. Besides, the trainees can be assessed at different stages: for their aptitude before admission, for their abilities while in training or for their achievement skills in the final examination, if any. Also, the training can be full-time, part-time or, even intensive, over 6 months or less. Therefore, appropriate criteria are difficult to determine for each stage.

Regarding the literature on educational criteria, It is worth noting that there are more studies on aptitude than on ability or achievement. Additionally, there are more studies on diagnostic and summative assessment than on formative assessment. Consequently, the difficulty in this research lies in the dearth of literature which deals directly with our research question. It seems that there has been no progress in this line of research as the dearth was also signaled by Gile (2000) and Peng (2006).

### **2.2.1. Aptitude Criteria**

Russo (2011) provided the most recent, extensive and updated review on interpreting aptitude starting from the early debates in the 1940s and 1950s on ‘nature and nurture’ and the subsequent attempts by professionals and trainers to identify the components of aptitude and how to test them until the latest findings which preceded the publication of her paper. Carrol (1978, as cited in Russo, 2011) exploited his

background research in language acquisition and testing to shed light on admission tests at George Town University where he suggested oral testing instead of written tests. Donovan (2003, as cited in Pöchhacker, 2011b) examined only the selection procedure of European Institutions, and Timarová and Ungoed-Thomas (2009) dealt with admissions tests for conference interpreting courses in Europe only.

As far as universities' aptitude requirements are concerned, we will start with examples from the oldest continent, Europe, where there are a lot of programmes. In the UK, for instance, the 'MA Interpreting' at the University of Surrey is offered in full-time mode and is studied over one year (Brochure). On successful completion of the programme, the student can be awarded a PG Diploma (120 credits), which usually takes 9 months if studied full time, or Master's degree in Interpreting (160 credits). In addition, the Postgraduate Certificate Interpreting is an exit award only and requires students to achieve 60 credits in the practice-oriented modules designated in the programme. The entry requirements are the same for native speakers of English, a minimum of a UK 2.2 honours degree, or equivalent, in the relevant language or another relevant subject (e.g. translation, law, criminology, forensic linguistics, healthcare, business studies). For speakers of other languages, an equivalent to a minimum of a UK 2.2 honours degree in English or another relevant subject. In exceptional cases, alternative qualifications and professional experience may be considered. Non-native speakers of English will also normally be required to have IELTS 6.5 or above (or equivalent), with a minimum of 6.5 in writing and speaking, and 6.0 in Listening and Reading. The university offers intensive English language pre-sessional courses, designed to take the candidate to the level of English ability and skill required for his studies in interpreting. There is possibility for those who did not have language degrees to change their minds and apply for a postgraduate training in interpreting at any time of their lives at their wish, but it is not the case in Algeria. Besides, language proficiency is considered more important than the possession of a higher degree in the UK. To apply for the M.A in Translation and Interpreting programme (SI is an optional module) at the University of Westminster, the standard postgraduate entrance requirements is that the candidate should have a first degree

from a recognised university or qualification or experience deemed to be equivalent, although mature linguists without a degree but with sufficient experience in translation and/or interpreting are invited to apply. Nevertheless, the candidate will need fluent written and spoken English and an IELTS score of 6.5 overall with 6.5 in writing and 7.0 in speaking or equivalent if English is not his first language.

In addition, at the University of Manchester, students who apply from overseas with non-UK qualifications should send a copy of their degree certificate, a transcript of their degree results with an officially authorised translation, and an indication of the marking scale relating to their degree. English language scores of IELTS 7 (with 7.0 in the writing element of the test) or TOEFL 600 (paper-based test), 250 (computer-based test) or 100 (internet-based test) or a Pearson Test of English (PTE) score of 70 overall (with 70 in the writing element of the test) are required for students whose first language is not English.

But, either at the University of Surrey or the University of Westminster, or the University of Manchester, the possession of the minimum entry criteria does not in itself guarantee admission as all applicants take either an ‘entry test’ or an ‘Interpreter Aptitude Test’. At the University of Westminster, the entry test consists of written and oral components, it comprised a personal interview, memory exercises, written translations and a general knowledge quiz. At the University of Surrey, the applicants must succeed in the interpreting aptitude test. At the University of Manchester, the upper second class Honours degree or higher (or its overseas equivalent), in a relevant subject is not the only criterion for admission to the MA in Conference Interpreting (MACINT) as the applicants must complete successfully an aptitude test for conference interpreting.

Meanwhile, in France, at ESIT, in addition to holding a degree and justifying language proficiency, there are two exams for admission, a written and an oral one. The written exam of ESIT consists of answering questions about the content of three long press articles (3-4 pages) which are extracted from Le Monde and The Economist. A student named ‘Alice. S’ answered a question on a forum about the

extent of difficulty of these texts, so she said, « *La complexité des textes eux-même ne veut pas dire grand chose [...], parce que fait la difficulté de l'examen, ce sont les questions, pas les textes. On peut poser des questions faciles sur des textes difficiles ou l'inverse !* » (interpreter.freeforum .org/ Fri. Apr 12, 2013). In other words, the exam requires higher order skills of reading comprehension. In 2013, the oral exam consists of two interviews with two juries with an interval of 40 mn. The first jury is for languages A, B and C. The second jury is for language D. The task consists of listening to two pieces of discourse of about 3-5 min without taking notes then reconstructing them in the second or the first language alternatively. After that, there is a short discussion in C about a discourse (C-A). At ISIT, the kind of questions that can be asked is different. It allows the candidate to express himself about various issues. For instance: “*Parlez-nous en russe de quelque chose qui se passe en ce moment, la Syrie, par exemple, ou la Grèce... Tout ce que vous voulez sauf les elections russes ?*” (Forum).

In Canada, the battery of selection tests at the University of Ottawa, were described by Lambert (Lambert, 1991). She presents in a simple way with a detailed description the screening instruments used at the University of Ottawa to select applicants seeking admission to the Graduate Diploma Programme which provides professional training in both consecutive and simultaneous interpretation. So, the selection exams include these exercises: 1) *Shadowing*, 2) *Cloze*, 3) *Sight translation or sight interpretation*, 4) *Memory test (Wechsler)* and 5) *Interview* (p. 25-32).

The latter for instance is private and lasts approximately 20 to 30 minutes per candidate. It is conducted by two faculty members in addition to the programme director when available. It is carried out in two working languages A and B. The candidate is asked first about his linguistic and educational background then about current events in Canada and in his native country if he is a foreigner. For the sake of standardization, candidates are given four topics, they discuss one in their A language and another one in their B language. Eventhough the candidates are given 20 minutes to prepare the topics before the interview, they are not allowed to read from their notes during the discussion (Lambert, 1991).

A new version of testing by using the latest tools of technology is being promoted at the University of Glendon (Brochure). Thus, the Interpreter Aptitude Test is administered online with the use of a headset through a virtual platform. According to Clifford (Video, 2015), the MA Interpreting Director, the test consists of three steps: The first step is the connection test which lasts 30 minutes to ensure the technical requirements are under control. The second step is round one of the test where the candidate listens without taking notes to a speech on a given topic then writes a summary. Round two consists of two tasks and takes about 60 minutes.

Obviously, entrance tests are not compulsory and depend on the legal situation of each country (Dejean le Féal, 2014), but because the aforementioned requirements are not enough to decide on the aptitude and pre-requisite skills of the candidates, most of the universities and translation schools administer selection examinations to choose the promising applicants only. In this regard, contrary to the basic requirements, the entrance tests or exams are different, because they do not have the same format and content and differ in what Gerver et al (1989: 724-735) termed the '*predictive power*'. In addition, this review confirms the weakness of the descriptions of admission tests available on interpreter training schools' websites which was highlighted by Rosier et al. (2011: 54): "[They] often fail to provide a clear notion of how the linguistic and non-linguistic skills of future interpreters should interlock." That is why these tests and examinations have been criticized as we will see below.

There is abundance of literature on aptitude criteria in the field of interpreting (Keiser, 1965; Lederer, 1974, 1975; Sofr, 1976; Carroll, 1978; Moser-Mercer, 1985; Szuki, 1988, Longley, 1989; Gerver & al, 1984, 1989; Dodds, 1990; Lambert, 1991; Alexieva, 1993; Russo, 1993, 2011; Pearl 1995; Tryuk, 2002; Pippa & Russo, 2002, 2004; Angelelli, 2007; Pöchhacker, 2011). While the practitioners described the components of aptitude only in light of their retrospections, most of the trainers and scholars tried to collect empirical and experimental data in order to set objective criteria for screening potential interpreters. They had all noticed weaknesses in the forms of tests adopted in the academic programmes of their universities or schools and worldwide. There is some agreement on the basic skills and aptitudes required for a

trainee or a novice interpreter (Lambert, 1991; Pippa & Russo, 2002; Russo, 2011) but there is no consensus on how to assess them reliably. For instance, linguistic proficiency was stressed by Lederer (1975) in terms of lexicon, grammar and idiomatic expressions and emphasized in many subsequent publications (Pippa & Russo, 2002). Herbert (1952), Longley (1989), Gerver et al. (1989) and Lambert (1991, 1992) surveyed the articles written by members of the profession as well as transcripts of interviews with working interpreters, and widened the scope and the profile of the ideal professional interpreter to include other abilities recognized as essential or prerequisites for success as a trainee or as a novice interpreter (Gerver et al, 1989; Lambert, 1991; 1992; Pippa & Russo, 2002; Russo: 9):

1. Profound knowledge of active and passive languages (A, B, C) and of the cultures.
2. Ability to grasp rapidly and convey the essential meaning of discourse, what is being said, irrespective of the spoken language.
3. Ability to memorise and recall the links between sequences of discourse.
4. Ability to project information with confidence, coupled with good voice.
5. Wide general knowledge and interests, and a willingness to acquire new information.
6. Ability to work as a member of the team.
7. Ability to work under stress for long periods.

Nevertheless, the selection of trainee interpreters became the most controversial issue in interpreting studies as there was disagreement on the type of exercises to use (Dejean le Féal, 2014) in order to ensure the validity and reliability of the aptitude tests. For example, some scholars advocated ‘shadowing’ (Lambert, 1991; Moser-Mercer, 1985; Schweda Nicholson, 1990) while others rejected it, or even believed it is harmful (Gerver et al, 1989; Alexieva, 1993). Besides, as mentioned previously, assessment of simultaneous and consecutive interpreting were treated in the same way which was problematic too. It is true that professional conference interpreters may have to practice both modes, but we argue that for pedagogic purposes these modes should be assessed separately as each mode involves different kinds of skills.

As mentioned above, after identifying the main features in an interpreting performance and the weaknesses of the candidates, Carroll (1978) recommended oral testing of: 1. Verbal intelligence, 2. General culture, 3. Multiple forms of fluency

(ideational, expressive, rephrasing and associative fluency), 4. Naming facility factor, 5. Shadowing (Russo, 2011). Dodds (1990) questioned the validity and reliability of the Italian aptitude tests in their form of that time, in light of the results of the studies conducted by other researchers who confirmed his assertions, namely Stenzl (1989), Gringiani (1990), C. Russo (1989) and Taylor (1990). He also questioned the fact that the testers assess aptitude on the basis of tasks that most of the candidates have never performed previously and considers the usefulness of shadowing, especially phonemic and lag shadowing, as a myth either as an aptitude test or as an introductory exercise to simultaneous interpreting. Contrary to Lambert (1991) and Schweda Nicholson (1990) who believe that shadowing is useful, Dodds states that nothing has been proven scientifically in this regard.

In addition, Dodds shows the weaknesses of traditional aptitude tests which consist of summary writing activities, sight translation and general knowledge. For him, each of these testing devices has weaknesses and cannot reflect the real aptitude of the candidates, especially that 45% of the successful students dropped in the middle of the training at the University of Trieste and failed to complete the course.

Hereafter, we will present the works of the researchers who went a step further, i.e. instead of remaining at the level of theoretical and empirical description and criticism, they have suggested some aptitude tests through different approaches and called for more experimental studies to check their validity and reliability. Such studies would also check their authenticity and feasibility, especially that there is an increasing number of applications on the translation and interpreting programmes. Indeed, Pöchhacker (2011b) draws a fork picture of two adopted approaches in screening students' performance: test-based assessment and course-based assessment. The main example for the latter is the course proposed by Moser-Mercer (1978), the description of which follows.

At the Monterey Institute of International Studies, Moser-Mercer (1985) developed a monolingual introductory course on simultaneous interpretation, based on her model for SI (1978), which served as a guideline for selecting students for

interpretation in the Graduate Division of Translation and Conference Interpretation. Then, she collected data about the performances of 114 students over a 4-year period (1978-1981) and compared the type of her recommendation with the results of these students at the end of their first and second (last) year of studies. She found a significant relationship between her recommendation and their performance. The hypothesis underlying her study was that students who received a recommendation are more likely to pass the qualifying examinations at the end of their first year of study and the final professional examinations at the end of their second year of study than those students who receive only qualified or no recommendation at all. Her 10-week course was offered in English and included 4 types of exercises which are shadowing, memory exercises, paraphrasing and processing of numbers. She issued three types of recommendations which are (1) Recommendation for interpreting, (2) Qualified recommendation for interpreting and (3) No recommendation for interpreting. This was complemented by other criteria, namely command of the English language, pronunciation and enunciation, ability to cope with stress, assertiveness, resilience and microphone technique. She concluded that before issuing a judgement on interpreting potential, it is desirable to observe the students over an extended period of time. Nevertheless, Pöchhacker (2011b) highlighted that there are no descriptions regarding the scoring method for the task of paraphrasing. In addition, he stated that the proposed course is time consuming and costly in terms of teaching resources, especially in a one-year MA programme.

As for the test-based approach, Carroll (1978, as cited in Russo, 2011) based his test on findings from foreign language acquisition. Adopting a multidisciplinary approach by relying on the findings of research in Interpreting and Psychology, Gerver et al. (1984, 1989) developed a multi-component testing battery. Lambert (1992a) also suggested a shadowing test while Schweda Nicholson (2005) described a test which focuses on personality traits. Other tests were developed by Russo (1993), Alixeva (1993), Russo and Pippa (2002, 2004), Arjona-Tseng (1994), Pöchhacker (2011).

At the School of Languages in London, Gerver et al (1984, 1989) aimed at establishing objective criteria for the selection of students for the postgraduate course in conference interpreting techniques, which was created in 1963 by Patricia Longley. Gerver et al (1984, 1989) devised twelve psychometric tests to facilitate the selection of simultaneous and consecutive interpreter-trainee candidates. The tests, which were based either on text materials (eight tests), linguistic subskills (three tests) or speed-stress (one test), were correlated with judges' ratings of the final interpretation examination. Indeed, text-based tests required either the recall of the auditorily presented information or the completion of individual target words in the text, i.e. there were exercises of text memory and logical memory in addition to cloze and error detection. The subskill-based tests test synonym generation, sentence re-expression and vocabulary selection to be an adequate reflection of the interpreter's task. The speed-stress test is called the Nufferno test: The students who passed the exam had higher mean scores on all tests than those who failed. Test scores were positively inter-correlated between completion-detection tests and simultaneous interpretation ratings and between recall tests and consecutive examination ratings.

At the Scuola Superiore per Interpreti e Traduttori of the University of Trieste, Russo (1993) proposed an aptitude test for an 'experiment' during the academic year 1988-1989 which aimed at highlighting the text-processing strategies of subjects under conditions simulating a simultaneous interpreting task. Twenty second-year Italian students who had never practiced simultaneous or conference interpretation before took part in the 'experiment'. The undergraduate students were asked to paraphrase an oral text from Italian into Italian. Nineteen of them started the two-year training course in simultaneous and consecutive interpreting. To score their performance, the researchers suggested five positive and five negative criteria, that is to say that they gave (+1) point for the semantic abstraction of incoming segments of information via construction or generalization, the active storage in short-term memory of linguistic occurrences, the syntactic and/or semantic anticipation ability, the paradigmatic replacement of a phrase and (+0.5) point for the ability to detect and modify a paradigmatic structure; whereas, (-0.5) points were given for form-linked errors

(wrong collocations, grammar mistakes, ...) and (-1) for loss of cohesion, loss of coherence and for semantic errors such as misinterpretation. In addition, two aspects were considered positively which are the length of the performance (Completion of 12 sentences) and the fulfillment of the task without breaks while two others were scored negatively, the deletion of a whole sentence and the fulfilment of the task with breaks

The result was that 6 students out of 19 failed because they abandoned the course and, statistically, the error probability of the test was 3 out of 100. She claims that this test has an 'accurate predictive value' but only the supplementary data will provide wider-ranging evidence in favour or against its validity. Russo and Pippa (2004) administered the same test to four interpreting advanced students who were asked to paraphrase the same text and interpret it into English. The aim was to observe possible similarities in the processing skills of the two groups, and within the same group for both tasks. On the whole, the advanced subjects performed better than the 2<sup>nd</sup> year students, but there were signs that both groups performed the same operations on given target information nuclei. The researchers doubted the possibility of generalizing their results because of the small size of their sample. Thus, these methodological considerations lessened their confidence about its predictive validity, because the test was repeated in 1991 with twelve 2<sup>nd</sup>-year students only were considered. In 1992, the sample of the students who started the two year interpreting course went down to thirteen (12). (90 students in 2002)

Russo and Pippa (2004) felt a need to develop their '*paraphrase evaluation model*', which is built upon a linguistic-cognitive model, because the assessment paradigm with one numerical score did not clearly differentiate between the many relevant components involved in the task. What is noteworthy is the fact that the researchers found a sound correlation between the test items and the interpreting exams results. In fact, the suggested three-tier model (Syntax and Lexicon, Semantics and Pragmatics) consists of 15 categories divided as follows: 7 categories for syntax and lexicon which are reduction, expansion, syntax transformation, lexico-syntactic transformation, lexical transformation, permutation and production disorder; 4 categories for semantics which are deletion, addition, substitution and loss of

coherence; 4 for pragmatics which are pragmatic loss, pragmatic gain, theme and focus. Russo and Pippa (2004) concluded that the paraphrasing-based model via its linguistic and cognitive criteria provide a reliable operational tool to evaluate interpreting aptitude and assess the candidates' performance. In fact, the limited number of positive and negative parameters correlating significantly with time needed and level of performance achieved are easy to apply. They have concluded that their model has a predictive validity but requires more studies to confirm it.

Alexieva (1993) reminded the reader of the criticism levelled at shadowing as an unreliable exercise for assessing aptitude and even as a training exercise (e.g. Dodds, 1990; Thiéry, 1990). She also emphasised the necessity of including a short-memory test in the entrance examinations. Firstly, she suggested a two-stage testing strategy to test the aptitude for a prompt and accurate recall of information retained from previously received texts by asking questions about the redundant information from Text 1 (Stage 1) in Text 2 (Stage 2). Secondly, she advocated a sight translation exercise or a SI proper test in stage 2 to get more reliable evidence about candidates' aptitudes to use their procedural knowledge, establish intertextuality and recall the stored information in their minds. Thirdly, she argued that a proper selection of the two texts combined with some parameters such as the number of matrix predications, the ratio between the expository and the argumentative predications and the observation of the '*three-to-one*' overlapping relationship between Text 1 and Text 2 can give the examiners the opportunity more accurately to assess the candidates' performance in quantitative terms. Forthly, she claimed that all three versions of the Two-Stage Testing work extremely well as efficient training exercises.

Arjona-Tseng (1994, as cited in Pöchhacker, 2004) developed and implemented a two-day screening procedure for admission to a two-year graduate program. Following a five-part written test for language proficiency and general knowledge, final selection was based on a series of oral tests, including written recall of a recorded passage, error detection, and sight translation. Out of a total of 565 applicants over three years, eleven candidates were selected as 'trainable' in conference interpreting between English and Mandarin Chinese.

In his turn, adopting an experimental design, Pöchhacker (2011b) suggested a SynClose test to stress the importance of understanding and weighing the aptitude criteria which determine how a trainees' probability of success is shaped, especially that they would have to choose between different interpreting specializations which depend on the programmes. In other words, the programme may be general or more focused on one or two modes of interpreting such as conference interpreting and community interpreting. Thus, there is a need for reduction so that the aptitude test involves the most important criteria which are required in any interpreting specialization. The students took this test in the third week of a course for which he was the instructor at the BA level. It leads to access the new MA at the University of Vienna where the two mentioned specializations are available. Pöchhacker's course is a 90-minute generic and monolingual interpreting course over 15 weeks of study. It aims at introducing students to interpreting through a set of activities such as active listening, message analysis, memory and note-taking. The intended outcome is to enable students to interpret a two-minute speech consecutively from German into German. After that, the test comprises an auditory cloze exercise and an exercise on synonymic sentence completion which requires high expressional fluency. The criteria taken into account in the scoring are the degree of accuracy and speed of response. The test was administered to 120 undergraduate students with a control group of interpreting students. Comparison was also made between students whose A language is German and those who have it as their B. As for the results, moderate correlation was found between the test scores and the students' performance on an intralingual consecutive interpreting exam at the end of the course. Therefore, there is a lack of data on the correlation with the performance in SI.

Further, some scholars, namely Szuki (1988), Gile (2001), Pöchhacker (2004), Schweda Nicholson (2005), Rosiers et al. (2011) and Shaw (2011), advocate the integration of personality criteria in the admission tests (see Russo, 2011). Indeed, in the late 1980s, It has been proved statistically by Szuki (1988) that personality traits, centers of interests and achievement motives among translators and interpreters, as aptitude criteria, vary tremendously. Using a questionnaire sent by email to 244

translators and interpreters from Tokyo, he had responses from 93 of them and concluded that: first, the translators have interest in (1) arts, especially in writing novels, screen plays, drama, baiku and poems, (2) looking after others and voluntary work, (3) intercultural contact on the job and in daily life. Besides, they are (4) patient, (5) cheerful and humorous, and (6) active. By contrast, interpreters have interest in (1) deepening insight into people, (2) verbal communication, (3) journalism, mass communication, and social issues, (4) physical labor. And, they are (5) empathetic towards others, (6) progressive and (7) extrovert. They have (8) high achievement motive. Moreover, the 'Dualists', i.e. those who are working as translators and interpreters at the same time, have interest in (1) looking after others and voluntary work, (2) deepening insight into people. They are (3) prompt and they have (4) very high achievement motive.

Furthermore, Timarova and Ungood Thomas (2007), Rosier et al. (2011), Shaw (2011) also dealt with the relevance of 'soft skills or personality traits' to the issue of distinction between translators' and interpreters' profiles by investigating individual difference variables and how they relate to interpreting students' performance. Rosier et al. (2011) stress that researchers have taken up the issue of personality research in second language acquisition studies as well as in interpreting studies since the early 2000s and have harvested widely divergent results. They explain that the assessment of personality characteristics in professional associations such as the AIIC is still left to the appraisal of assessors. In 2008, Shaw (2011) studied the difference between spoken language (SP) and sign language (SL) interpreting students at four institutions in the European Union in terms of their cognitive and motivational aptitudes using a computerized neuropsychological screening test and an achievement motivation instrument. Significant differences between the SP and SL students were found in the areas of visual memory, concentration, and internality (belief that success is due to internal causes), and between the advanced and entry-level students in the areas of concentration and the eagerness to learn new concepts in the absence of external rewards.

Pöchhacker (2004) deplors that despite the usefulness of standardized language grades, such as those used for NATO peacekeeping interpreters or those formulated in the Common European Framework of Reference for Languages, in screening interpreters for their linguistic aptitude, they remain useless in the assessment of cognitive and affective aptitudes. He emphasised that personal qualities such as motivation, learning style, and coping with physical as well as emotional stress have not been objectively assessed in entrance examinations to date.

All in all, there are no conclusive findings regarding where to place the focus, on the already acquired skills or on the aptitude to learn interpreting-related skills (Lederer, 1975; Sofr, 1976; Pearl, 1995; and more recently Tryuk, 2002 and Angelelli, 2007, who talks about ‘interpreter readiness’) (Russo, 2011). Furthermore, Gile (2012) notes that despite the insightful research on interpreting aptitude, none shows the way to efficient high prediction-rate aptitude testing and it is still difficult to design reliable tests with a strong predictive power.

As for the exams, one format only is used which is simultaneous interpreting of short speeches. Nevertheless, little genuine information is available about assessment of abilities and achievement in other universities worldwide. That is to say, we could not gather data about the content and format of real exams.

### **2.2.2. Ability Criteria**

Few studies describe Interpreting curricula in detail. They only discuss the exercises or exams and their aims.

Gile (2001) summarizes his stance regarding the issue of in-training interpreting quality assessment as compared to on-the-job assessment by stating that it is process-oriented:

*A process-oriented assessment is recommended for the early part of the training syllabus in view of its psychological advantages and its efficiency in guiding students. A gradual migration towards product-oriented assessment is however necessary so as to increase the efficiency of the instructors’ intervention in helping students optimize the product and in*

*preparing them to graduation tests. Potential discrepancies between the instructors' norms and market norms are not fundamentally problematic as long as they reflect higher standards in the training environment rather than differences in interpretation strategies and norms. (p. 379)*

At trainee level, Schweda Nicholson (1991, as cited in Kelett Bidoli, 2013) addressed the issue of quality from the linguistic point of view (for both simultaneous and consecutive interpretation) through self correction of language errors. She suggested a number of monitoring strategies to be taught in the classroom to make trainees aware of mistakes before, during and after interpretation and thus overall quality. In a subsequent paper she promoted a 'constructivist criticism model' with teacher-student, student-student interaction to assist "*trainees in identifying their strengths and weaknesses*" and contribute to "*improved performance during subsequent exercises*". She dealt with various types of error in the categories of terminology, lag time, public speaking skills, completeness and accuracy and the handling of difficult points.

Three alternative forms of assessment were presented by Lindquist (2005), Lee (2005) and Iaroslavschi (2012). Lindquist (2005) presents the MRC analytical approach which is a technology-based model for interpreter performance assessment. It relies on applied digital audio, word-processing, and spreadsheet technologies in addition to combining discourse analysis and corpus linguistics techniques to identify specific areas of difficulty for student interpreters. As for self and peer assessment, Lee (2005) explores the usefulness of self-assessment in an interpretation classroom by, first, asking students of the Korean-English Program in a two-year graduate school of translation and interpretation to analyze and assess their interpretation performances and submit a self-assessment report to the instructor. Second, by conducting a survey to find out what aspects of self-assessment the students found positive or negative. The results revealed that self-assessment is a useful tool of learning and is positive in identifying students' weaknesses and strengths, in enabling targeted practice, and in monitoring progress. But students also found selfassessment to be time-consuming and emotionally draining. In this study, three assessment criteria were adopted for self-assessment: meaning (meaning errors, additions, omissions, deviations, excessive

condensations and generalization that distort the message of the source speech), language use (grammar errors, inappropriate sentence structure, word choice, and terminology errors), and finally delivery (unfinished sentence, back-tracking, pause, hesitation, and pronunciation). Iaroslavschi (2012) suggests some solutions to avoid the problem of assessing SI by impression both by the teacher and the trainee interpreters who might face problems (when asked to give feedback about their peers' performance), because it is difficult to listen to the ST, to the TT and take notes of the performance flaws. She suggested that the trainees be split into performers and assessors, then the content of the SL discourse should be explained to the assessors. Besides, the details regarding names and numbers should also be given to them. However, she does not encourage giving them the transcription of the source text. The second method is "*the triangle method*" where the trainee-assessor is asked to take notes and perform and a consecutive interpreting when SI is finished. Furthermore, she discarded the use of a supplementary discourse by the evaluator due to the load and energy required to process the comparison between multiple discourses simultaneously.

### **2.2.3. Achievement Criteria**

Pippa (2009) states that the evaluation of simultaneous interpretation at the end of courses is still a challenge today both theoretically and didactically, because the definition of the level of skill required in graduation compared to the professional level is at the heart of the interpretation evaluation process. Kalina (1994, as cited in Kellett Bidoli, 2013) devised a test to compare student trainee strategies in a mock conference setting with those of professional interpreters at a live conference, to raise student awareness of the standards required. The main problem was the identification of the various strategies implemented, each leading to its own analytical problems.

Gerver et al (1989) pointed out that the final examinations of a six-month postgraduate course in conference interpreting at the School of Languages in London, are oral in both forms of interpretation, i.e. simultaneous and consecutive, in the candidate's principal working languages, and are conducted by a board of external

examiners who are themselves senior practicing interpreters in major international organizations such as the U.N., Council of Europe, International Labour Organization (I.L.O.), and the World Health Organization (W.H.O.) Students were rated out of 10 by each examiner independently (usually by three examiners) for each principal working language combination and for each form of interpretation. In cases of disagreement, the examiners discussed the ratings until agreement was reached . No assessment other than by examiners' ratings was made. A pass mark for the diploma was a rating of 5 out of 10. In addition, Mikkelson (2013) points out that university certification of interpreters is a relatively recent trend in the translation and interpretation profession, but it is rapidly gaining favor as a mechanism for guaranteeing quality for the users of interpreting services. Nonetheless, certification is still carried out through the professional associations' tools and processes as we will see below. But, it is worth mentioning that there is a plethora of criteria, norms and standards which are industry-based. We will delve into this area in the next section.

### **2.3. Industry-based Criteria, Norms & Standards of SIQA:**

There are many other criteria, norms and standards for SIQA that are industry-based. These are related to self and peers, certification and accreditation, employers or service providers (Pöchhacker, 2012) and clients, and finally end-users.

#### **2.3.1. Self & Peers Criteria**

It is worth mentioning that self-evaluation exercises through tape-recording of output of professional interpreters in the booth have been encouraged by Dejean le Féal (1990) to give the interpreter a better insight to his/her performance and increase the motivation for improvement, because he/she hardly profits for self-improvement from the mutual judgements with the colleagues or from the *ad hoc* judgements leveled by his audience in the absence of recordings. The interpreter would be ill-equipped to appreciate his/her own performance objectively (Kellett Bidoli, 2013) due to the multiple criteria which must be considered at the same time. Kurz (1993) delved into the criteria set by the interpreters themselves as professionals by reviewing some studies such as Bühler's (1986) survey. This latter concluded that interpreters, who

were asked to choose among 15 linguistic and extra-linguistic criteria, agree on the criteria applied to sponsor their would-be colleagues in the AIIC. These criteria are: 1. native accent, 2. Pleasant voice, 3. Fluency of delivery, 4. Logical cohesion of utterance, 5, sense consistency with original message, 6, completeness of interpretation, 7. Correct grammatical usage, 8. Use of correct terminology. However, Kurz claimed that a good interpretation is determined only by the interpreters' criteria, because only empirical studies would ascertain the match between their perspective and the users' perspective. According to Kurz, this call was launched also by Cartellieri (1983), Gile (1989) and Stenzl (1983). We will give a more detailed account of the users' perspective in the last section below. Besides, Jesus Sanz (1930) mentions other criteria than the ones highlighted in Bühler's survey such as cognition abilities (intelligence, intuition and memory) in addition to moral and affective qualities (tact, discretion and poise) (Pöchhacker, 2011). Thus, there are different views among the members of the same community as shown in the results of the other studies selected for this review which addressed the relevance and importance of various parameters and criteria.

For example, in a broader perspective which attempts to address several issues concerned with communication during SI in general and not only the linguistic elements, Altman (1990: 23) completed two surveys, with two different groups, on conference interpreters' attitudes to their own contribution to the communication process where she tried to identify the factors affecting their performance and to discover *“under what circumstances do they themselves consider they have failed to establish communication between speaker and listener”*. The two different groups were composed of EC interpreters in Brussels and AIIC members in the British Isles. A high correlation between the views of the two groups emerged. It was found that *“cultural differences are a greater obstacle to communication than linguistic ones”* and that the individual factors *“have the greatest impact on the interpreter's ability to bridge the communication gap”* (as cited in Kellett Bidoli, 2013: 112).

Aside from the meta-analysis as such, the following presentation of survey research findings is limited to the topic of conference interpreters' perception of their

professional role. Scholars relate it to the issue of quality, since conference interpreters' performance is based on their perceptions of the nature of their tasks (Zwischenberger & Pöchhacker, 2010). To add a brick in this line of research, we will concentrate on some works that have not been widely reported.

To begin with, Chiaro and Nocella (2004) conducted a web-survey among 286 conference interpreters who had no chance to express their views on the issue of quality elsewhere previously (as cited in Zwischenberger, accessed online in 2013). Using multidimensional scaling, they devised a perceptual map based upon interpreters' ranking of a number of linguistic and nonlinguistic criteria which reflect their perception of the criterion importance in the interpretative process. Moser (1995, 1996) summarized the findings of a comprehensive survey conducted by the AIIC about the professional interpreters' views. In order to investigate the self-evaluation criteria, i.e. the professional interpreters' criteria, Pöchhacker (2011a) provided an overview on survey research in the field of interpreting with a corpus of 40 survey-based studies among conference interpreters as a way of gaining knowledge about the profession, including the role perception of individual practitioners. Touching on design issues of qualitative research such as sampling and question types, he showed that the 'role' of conference interpreters emerges as one of their dominant concerns as far as their occupational status is concerned. This 'survey of surveys', which was designed as groundwork toward a state-of-the-art survey on role issues in conference interpreting, was carried out as part of a larger research project on 'Quality in Simultaneous Interpreting' at the Center for Translation Studies of the University of Vienna. So, Pöchhacker's (2011: 50) research question is not: "*What is interesting or unique about the profession?*" but "*What have survey researchers been interested in?*". And, on the whole, his focus is not so much on what these survey researchers have found out as on how they have conducted their research. He sketches a quantitative as well as qualitative analysis of survey research with an emphasis on methodological issues and on the topics addressed.

The web-survey on Quality and Role (2008), which was part of the research project on Quality in Simultaneous Interpreting mentioned previously, yielded 704

responses out of 2523 requests sent to the AIIC interpreters via their emails, which were taken from the AIIC Directory. The questionnaire consisted of 41 items, including some follow-up questions, and was divided into three main parts. Part A elicited information on sociodemographic background variables, such as age, gender, working experience, and language combination, while Part B was essentially a replication of Bühler's (1986) study, the AIIC members were asked to judge the importance of 11 output-related quality criteria. Part B also included a web-based experiment for which respondents were asked to give their impression about a one-minute audio sample of a simultaneous interpretation by rating it on a six-point scale. The survey recipients had been randomized into two groups, one hearing the sample with lively intonation and the other with monotonous intonation. Part C was devoted to aspects of the conference interpreter's role. With regard to role, the survey examined how interpreters' sociodemographic and professional background data related to their perception of four behavioral constructs: Intervention in the original, loyalty to the speaker/original, reaction to working conditions and interpreter's detachment.

By using a questionnaire administered to 33 interpreters from Taiwan and a semi-structured interview with six of them, Fan (2012) looked into their views on the importance-based ranking of the required aptitude criteria and tried to identify the ones which can be taught. He believed that the answer to this question would increase the predictive power of admission tests. The 13 aptitudes, the importance of which was highly emphasized in previous research, were introduced to the interpreters according to three dimensions: knowledge (mother tongue proficiency, foreign language proficiency, world knowledge, professional knowledge), skills (source language comprehension, speed of comprehension, public speaking skills) and personality (stress tolerance, mental stamina, fast learning curve). The results revealed that the most important criteria appeared to be language proficiency, comprehension of the source language and speed of comprehension. Additional criteria were highlighted during the interview by the interpreters as pre-requisites, namely common sense,

logical thinking, curiosity and the desire to communicate. Multitasking was seen as the most important criterion which can be taught.

### **2.3.2. Certification and Accreditation Criteria, Norms & Standards**

Chan (2011) pointed out that the terms certification, accreditation, licensure and registration can cause confusion; Stejskal (2003) clarified the difference between certification and accreditation by stating that certification is a voluntary process by which an organization grants recognition to an individual while accreditation is a process by which an entity grants public recognition to an organization. Mikkelson (2013: 66) points out that alternative terms are also being misused such as ‘licensure’ which is an alternative term for the credentialing of individuals. According to her, ‘licensure’ normally refers to *“an authorization granted by a government agency to an individual who has met certain standards so that he or she may engage in an occupation for a fixed period of time.”*. Additionally, she mentions that Bancroft and Rubio-Fitzpatrick (2009) further divide certification into three subcategories: government certification (licensure), professional certification (awarded by a professional interpreters association, as is the practice in Canada and Australia), and program certification (granted upon successful completion of training). Furthermore, professional associations such as AIIC gives ‘recognition’ to the schools which comply with some requirements, one of them being the use of entrance examination (AIIC Training Committee 2010; Russo, 2011).

In this section, we will tackle the criteria, norms and standards required for the certification of interpreters and not the accredited agencies which are providing interpreting services. Nonetheless, it is worthy of note that there are no certification tests just for simultaneous interpreters, because these criteria, norms and standards concern conference interpreters in general. We will try to focus on the tests’ tasks regarding SI only.

Monacelli (2009) reviewed the works which discussed the norms in interpreting as a profession, as a system of shared values and ideas. She states that the norms are set by professional organizations, because in most parts of the world this profession is

not regulated at a central (governmental) level. Thus, the professional organizations play the role of the body which is giving sanctions in this activity both to endorse the admission of new members and to authorize the code of conduct. The professional interpreting associations which operate at a local level are either affiliated to an international association such as the Association internationale des interprètes de conférence (AIIC) and The American Association of Language Specialists (TAALS), or members in a strictly local association that is national in scope such as Assointerpreti, in Italy. Some associations are regional entities as they restrict membership to individual geo-political regions within a country's border (A.I.C.F.V.G., Associazione interpreti di conferenza del Friuli-Venezia-Giulia, from a Region in northeast Italy).

Monacelli (2009: 13) states that an interesting paper entitled '*Discourse Analysis as a Tool for Investigating Norms in Simultaneous Interpreting*' was read by Ebru Diriker (2000) at a conference in Manchester on research models in Translation Studies, in a session dedicated to issues in interpreting research and textual analysis. Acknowledging that Interpreting Studies has been rather reluctant to accept and deal with interpreting as a norm-governed social behavior, she gives an account of how norms are generally conceived in the discipline, stipulating that in the field norms are not considered an 'intrinsic' part of the cognitive processes of the interpreter. They are at best external factors that the interpreter can take into account consciously and willingly but only to a limited extent, due to the impact of the cognitive overload in simultaneous interpreting. Norms are nonetheless assumed to be an intrinsic part of the interpreter's decision-making process to some extent, since they represent social factors in the discipline. Norms in interpreting have also attracted the attention of Kalina (2007) who borrows the concept from TS as we saw in the first chapter. That is to say, from Toury.

Generally speaking, the criteria of these professional associations are: 1) 100 days of practice; 2) Endorsement from members of these associations. In addition to the working experience, Kurz (1993) highlights that adherence to the AIIC for example requires the endorsement of five professional interpreters, who must observe

their would-be colleague performing in a real working situation, to the candidate. Besides, adherence depends on the opinion of a special committee, the CACL. Additionally, the AIIC stresses the knowledge and respect of the Code of Professional Ethics (Code d’Ethique Professionnelle).

In the UK, the Institute of Translation & Interpreting (ITI) has 2 types of membership, individual and corporate. First, Individual Membership which is available to translators and interpreters at every stage of their career and to any individual who has a professional interest in the translation or interpreting industry. There are 6 ‘*career-ladder*’ categories and 2 additional categories (Fellow, Associate, Career Affiliate, Student Affiliate, Academic, Supporter). Second, Corporate Membership which is divided into 3 categories. The first category is for businesses and organisations that provide translation and/or interpreting services. The second category is for universities who provide translation and/or interpreting qualifications. The third category is for any business or organisation that has a professional interest in the translation or interpreting industry. Therefore, the criteria for individual membership are the academic qualification and the number of years of professional experience after graduation.

Setton and Liangliang (2011) (as cited in Liu & Luo, 2005) state that Taiwan launched its official certification exams after several years of intensive government-funded research. On mainland China, certification schemes are still limited and experimental; examples include CATTI, a test required for promotion within the civil service but also open to other applicants, and various locally-based certificates offered by universities and other institutions (e.g., the Shanghai Interpreters Accreditation (SIA)) with varying degrees of recognition by private or government recruiters.

Let us consider the interpreting skills required for the certification of a Court Interpreter in the USA (ALTA Report, 2010:43), i.e. the abilities assessed in a test for community interpreters. They are as follows:

- *Ability to concentrate and focus;*
- *Ability to process linguistic information quickly;*

- *Ability to make quick linguistic decisions regarding word choice or terminology selection;*
- *Ability to apply short-term memory skills in retaining small units of information;*
- *Ability to think analytically;*
- *Ability to utilize predictive thinking skills to anticipate incoming messages;*
- *Ability to convey meaning;*
- *Ability to provide transference from one language to another;*
- *Ability to preserve accuracy;*
- *Ability to select appropriate equivalent for vocabulary or phrases;*
- *Ability to conserve intent, tone, style, and utterances of all messages;*
- *Ability to reflect register; and*
- *Ability to self-monitor and self-correct*

This is a succinct description of the required skills. However, the writers of this report do not provide any criteria with detailed descriptors which reflect the professional performance in the different modes of interpreting. It is needless to say that these skills are performed at different levels of proficiency when we compare the performance in conference interpreting and in community interpreting.

Below is a sample description of a test of SI for the certification of community interpreters in the Candidates' Handbook (2011: 2-3) of the Quality Assurance Test, which was designed by the Ministry of Justice (UK). It describes the mode of interpreting, the content of the prompt and the task. Additionally, it gives details about the rate of delivery, the number of takes and some details about the abilities which are required to ensure quality, i.e. paying attention to specific details in the original speech.

*“Test 2 is a simultaneous interpreting test, from English into the non-English language of your language pair. You will see and hear a speaker of English. They will be EITHER a Probation Officer OR an Immigration Judge, OR a Crown Prosecution lawyer, OR a defence lawyer. During this test, you should interpret into the non-English language of your language pair using the simultaneous mode.*

*Each clip is delivered at a rate of approximately 120 words per minute. You will not be able to pause the video-clip while it is playing.*

*Test Outcome: your ability to understand and render semi-technical English from a legal context into the non-English language of your language pair, using the simultaneous mode of interpreting*

*Advisory notice: the English language video clips do contain reports of obscene language used by defendants and of offensive criminal behaviour.”*

In the UK, the graduate can enter the professional interpreting market. He can also become full member of the Chartered Institute of Linguists and/or the Institute of Translation and Interpreting.

### **2.3.3. Professional Criteria, Norms & Standards: Employers & Clients**

Let us begin with the institutional perspective. For instance, to join one of the EU institutions as a conference interpreter, the required basic criteria after which the candidate must succeed two tests, are:

- Perfect command of one European language and a thorough command of at least 2 others (one of the languages must be English, French or German).
- An appropriate qualification in conference interpreting or a degree plus one year's professional experience as a conference interpreter.
- Many years of experience, to be recruited at a more senior level (Grade AD 7).

As regards the clients' view, Grbić (2008) reminds us that the subjective opinions of the clients are frequently guided by first impressions, perceived ideas and personal feelings due to the strong influence of the norms, the stereotypes and the emotions in addition to the personality traits of the interpreter and his attitudes (Grazzone, 2002; Grbić, 2006; Strong & Rudser, 1992; Edwards et al. 2005).

### **2.3.4. End-users Criteria**

Kurz (1993, 2001) points out that from the early studies there was a call to pay attention for the audience(s)' opinion and the context of communication situation. Mainly, building on the studies of Herbert (1952), Gold (1973), Chernov (1985), Seleskovitch (1986, 1989), Snelling (1989), Déjean le Féal (1990), Thiery (1990) and Kahane (2000). Kurz showed that all these practitioners concur about this call, which was of course confirmed later by most of the scholars interested in quality research as

they advanced this line of research through empirical surveys to highlight the need of considering the criteria set by the audiences or the situational contexts, more generally. Chiaro and Nocella (2004) point out that these studies on users' perception could be subdivided into two types: studies on the reception of fixed interpretation and studies on perception in general. In this regard, to synthesize the perceptions of end-users regarding the criteria of quality, we will give an account of some interesting findings which have been revealed in the works of Moser-mercer (1985), Bühler (1986), Kurz (1988, 1993, 1994, 2001), Gile (1990), Meak (1990), Kopczynski (1994) (Kellett Bidoli, 2013), Chiaro and Nocella (2004).

Surprisingly, Kurz (1988, 1993) shows that the first eight criteria which were rated by interpreters as highly important in Bühler's (1986) survey such as native accent, pleasant voice and correct usage of grammar were not given the same consideration by the users, who were delegates to a medical conference, as they have given them much lower ratings. According to Kurz, this finding was confirmed by Gile's (1990) survey of ophthalmologists' perceptions, because his respondents did not give a high rating to the quality of the voice. Thus, he made the hypothesis that for scientists the voice plays a minor role in ensuring the quality of the interpretation. Collados Aís (1998/2002), points out that delivery-related quality criteria such as intonation, are usually given less importance when rated generically, but play a significant role when the overall quality of a given interpreting performance is to be assessed (Zwischenberger & Pöchhacker, 2010). Additionally, Both Gile (1990) and Meak (1990) have revealed that different user groups differ in their perceptions of quality. Indeed, Meak (1990) questioned ten Italian medical experts representing different disciplines (cardiology, urology, pediatrics, clinical pathology, pharmacology, etc.) (Kurz, 1993; Chiaro & Nocella, 2004). Unsurprisingly, Kopezynski (1994) also concluded that speakers and delegates have different expectations.

Moreover, Kurz (1993, 1994, 1996, 2001) examined the variability of perceptions among users through three surveys. She asked three different user groups to rank Bühler's (1986) criteria to look for any similarities with the AIC interpreters'

perceptions. The first survey was conducted in 1988 during an international conference on medicine, the second in 1989 during an international conference on quality control, and finally, the third during the Council of Europe meeting on equivalences. The 124 users rated these criteria differently, but the average is as follows: 1. sense consistency, 2. logical cohesion of utterance, 3. use of correct terminology, 4. completeness of interpretation, 5. fluency of delivery, 6. correct grammatical usage, 7. pleasant voice, 8. Native accent. Kurz (2001: 394-409) concludes that “*Quality of interpretation services is evaluated by users in terms of what they actually receive in relation to what they expected.*”. Mack and Cattaruzza (1995) and Colladis Aís (1998) followed Kurz’s (1993) footsteps. Nevertheless, Viezzi (1996, as cited in Chiaro and Nocella, 2004) claims that users cannot give an objective account of quality evaluation and provides four parameters for assessing quality: equivalence, accuracy, adequacy and user-friendliness. Users can only assess the delivery-related criteria, because they do not have access to the original. Therefore, they cannot judge its faithfulness to the original.

Kopczynski (1994: 88, as cited in Kellett Bidoli’s 2013: 113) considered that quality does not have ‘*an absolute value*’ but it is ‘*contextually determined*’ by variables such as the setting of a conference or the norms of interaction of a speech community. As for the design of his research, he chose to study two small Polish end-user groups, speakers and receptors at an international conference, who formed three distinct classes: specialists from the humanities (20), specialists from techno-scientific disciplines (23), and diplomats (14). Kopczynski found a general uniformity of opinions among the three groups regarding their preference for content and correct use of terminology over form in addition to their preference for a ghost role of the interpreter. Similar to this results are Moser’s (1995) findings which highlight the importance of content. Moser reported the final results of an interview and questionnaire-based survey (1995) which was conducted by the AIIC about the expectations of end users of conference interpretation. He revealed two types of criteria, the first type is concerned with content and the second with form. Under content-related criteria, he mentions completeness of rendition, terminological

accuracy and faithfulness to meaning. As for form, he reported synchronicity, rhetorical skills and voice.

### **Summary & Conclusion**

This third chapter dealt with the criteria, norms and standards of SIQA. It was divided into three sections. In the first section, we tried to define SIQA. In the second section, we tackle the educational criteria for SIQA of trainee interpreters (i.e. aptitude, ability and achievement criteria). In the last section, we delved into the industry-based criteria, norms and standards of SIQA. So, we tackled self and peer criteria. Then, we dealt with certification and accreditation criteria, norms and standards. After that, we shed light on the criteria, norms and standards of the employers and clients. Finally, we dealt with end-users criteria. We concluded with this summary and conclusion.

We conclude that the term SIQ is vague, relative and complex and the number of factors, variables and criteria included in its assessment is huge. Scholars and professionals alike draw many lists of, both linguistic and non-verbal communication, criteria, norms and standards which must be taken into account. Besides, researchers tried to sift them down into a short list on which there is agreement among the interpreters, senior interpreters, employers, clients and end-users. However, there is little research on weighing these criteria, norms and standards in educational settings in relation to clear constructs.

We focused in our analysis of the educational criteria on their relevance to aptitude (e.g. mastery of languages, quick comprehension, memory, voice, teamwork, work load, stress, etc), ability and achievement. Yet, it is worthy of mention that there dearth of literature on ability and achievement criteria, because the constructs in the field of simultaneous interpreting are not well-defined (Sawyer, 2004).

There is a missing link between educational-based and industry-based criteria, norms and standards, because there is a need for more empirical and experimental studies on change of performance in the various educational and professional settings.

# **Chapter Four:**

## **Methodology, Design and Procedure**

This chapter is devoted to explaining and justifying our choices in terms of methodology, research design and procedure. The first section is devoted to giving the reader an idea about the context of this study. Thus, it deals with the presentation of the Higher Arab Institute of Translation by describing the programmes, types and modes of interpreting offered in this Institute and how SI is currently assessed at the HAIT. Moreover, we will shed light on the paradigm and approach of this study. Then, we will describe the research design of the pilot and follow-up studies. Furthermore, a description of the administration of the research instrument is provided.

#### **4.1. Context of the Study:**

This study investigates the HAIT, which was inaugurated in Algiers in 2005. This Institute is a scientific, academic and research body, amongst other academic institutions of the League of Arab States. It was founded in order to promote research and the translation movement from and into Arabic as well as to ensure high-quality trainings for would-be translators and interpreters at post-graduate level. On its website, the HAIT states that it has many goals which exceed the area of training to reach other goals. First, it aims at the dissemination of research, documentation and exchange with other Arab research centers, not to mention the creation of a scientific regional and international network which might contribute, through the various translations, in the promotion of the Arab-Islamic heritage and the establishment of the Arab interpreter library. The HAIT believes in the importance of contributing as much as possible in the definition of the Arab and Islamic heritage by translating books from Arabic into other languages in order to improve the Arab and Islamic image in the eyes of the other. We will focus solely on the training mission of this Institute.

Since the creation of the HAIT, more than 72 students have been trained in interpreting. In order to apply, the students must hold at least a BA in Translation or in foreign languages; so far, only English or French languages are accepted. The Students' mastery of Arabic is tested through an interview. The first class was composed of 30 students, 20 in written translation and 10 in interpreting. In 2014,

there were more than 234 enrolled students (121 of them graduated, more than 62 are interpreters). We will give further details when we tackle the population of this study.

#### **4.1.1. The Programmes: Degree Tracks & Language Combinations:**

In a globalized world which is characterized by an explosion of communication and international trade, despite the fact that there is a need for more specialization and the need to cover more languages, the HAIT still provides degree qualifications in two majors and two combinations only, trainees choose either translation or interpreting. Focus in the HAIT is only on three languages which are Arabic, French and English. The two combinations are ‘Arabic – English – French’ or ‘Arabic – French – English’. Other languages which are well-ranked worldwide such as Chinese, Spanish and German or Russian for example are not represented. As a result, the Arab world do not have a view from these windows which are still closed. Furthermore, there is a technical track/major which is translation technology; it is open mainly for engineers and technicians in computing but graduates in languages with interests in computing can apply.

The three degree tracks provided by the HAIT (i.e., 1) Written Translation, 2) Interpreting, 3) Translation Technology) are two-year programmes. However, practically speaking, most of the students need more time to complete their dissertations. The first semester of the first year is a core training for all the students. Then, they have to choose their major starting from the second semester. The Admission Test or the Entrance test is an “Interview”. When the level of the selected candidates is not homogeneous, the administration may decide to administer a written test. So, the written test is not compulsory.

After the University Reforms which led to the adoption of the LMD system by the Algerian Ministry of Higher Education and Scientific Research, the curriculum of the HAIT was slightly changed to comply with it. It was divided into four teaching units including a number of courses (modules). These units are labeled: Fundamental, Methodological, Discovery and Transversal. No further details were provided by the administration on the content of these courses. There are no final exams, i.e the exams

are not administered at the end of the semester. Indeed, the exams are administered upon completion of a course. We will later deal with the assessment process and procedures in depth and focus on how simultaneous Interpreting is assessed in S2, S3 and S4.

There is scarcity of information and documentation regarding the detailed description of the instructional syllabus of each module. There is just a one-page document on which there is general info about the labels of the modules, the number of hours per credit, and the “coefficient” to calculate the pass mark. However, one cannot know the exact course content or syllabus (i.e the content which is going to be taught in class). Each teacher is free to choose what to teach and what to leave out. Obviously, there is lack of coordination and of a clear objective for each module, especially that most of the teachers are visiting lecturers. Consequently, teachers would not have clear cut criteria to assess different levels of mastery of the targeted skills or sub-skills.

#### **4.1.2. Interpreting Modes at the HAIT:**

As ‘activity’, Interpreting has a lot of modes, which are taught at the HAIT as modules and not as majors or specializations, such as Sight Interpreting, Consecutive Interpreting and Simultaneous Interpreting. Nevertheless, other important modes of Interpreting have no existence in the HAIT syllabus; therefore, no training is offered, such as:

- i. Community interpreting (USA)/Public Service Interpreting (UK). This mode may take several forms and give room for a more focused training: a) Medical interpreting, b) Court Interpreting, c) Legal interpreting, d) Police Interpreting, e) Business Interpreting. There is increase in the accreditation of these kinds of very specialized training programmes in the USA, Europe and Australia due the needs in their respective societies. These modes may be achieved via: a) Over-the-phone interpreting, b) Tele-conference interpreting or Remote Interpreting, c) Whispered Interpreting/Chuchotage.

- ii. Media Interpreting which may take two forms: a) Broadcasting Interpreting, b) Voice-over Interpreting/Dubbing.
- iii. Sign Language Interpreting

In this research, we will study only one mode provided by the HAIT, which is SI.

#### 4.1.3. Assessment of SI at the HAIT:

This study revolves around this question, i.e. what are the assessment procedures of the trainees at the HAIT? In order to answer it fully, we need to start in this section by a detailed description of what is currently practiced at the HAIT, then move on to complete the picture with more empirical data in the subsequent chapters and their sections.

As informed from the students' transcripts, the teaching programme at the HAIT has known three periods of change, the last one occurred after 2010 to comply with the requirements of the LMD system where the SI module knew a reduction in the hourly volume, as shown below. The first period spans from 2005 to 2007. The second period is from 2008 to 2012. The third period is from 2012 to 2015. Tables 1, 2 and 3 below describe the courses related to SI, which were provided by the Institute between 2005 and 2015. The 'Practicum training in SI', i.e. interpreting in real conferences, was not mentioned, because it does not seem to be assessed.

Table 1: Teaching and assessment of SI at the HAIT (2005-2007)

Semester	Code	Module	Coefficient	N.B. Estimated hourly volume/Semester
S1				
S2	S115	Methodology of SI	03	72H
	S116	Preparation for SI - AB / BA	02	72H
S3	S204	SI - AB/BA	03	72H
	S205	Group Training in SI	01	72H
	S207	Preparation for Conferences	01	72H
S4	S212	SI - AB/BA	03	72H
	S213	Group Training in SI	01	72H
Total	7	10 modules	14	504H

Table 2: Teaching and assessment of SI at the HAIT (2008-2010)

Semester	Code	Module	Coefficient	N.B. Estimated hourly volume/Semester
S1				
S2	S116	Preparation for SI AB	03	72H
		Preparation for SI BA	03	72H
S3	S206	SI AB	03	72H
		SI BA	03	72H
	S207	Group Training in SI	01	72H
S4	S211	SI AB	03	72H
		SI BA	03	72H
		SI CA	03	72H
	S213	Group Training in SI	01	72H
Total	5	9 modules	23	648H

For the third period, (with the adoption of the LMD system), SI is taught and assessed in this way:

Table 3: Teaching and assessment of SI at the HAIT (2012-2015)

Semester	Code	Module	Hourly Volume/Semester	Coef.	Credits
S1					
S2	MU 1 Preparation for SI	Module 1 : Preparation for SI A-B	28H	3	3
		Module 2 : Preparation for SI B-A	28H	3	3
S3	FU 3 SI	Module 1: SI B-A	28H	3	2
		Module 2: SI A-B	28H	3	2
	DU 1 Group Training	Module 3 : Group Training in SI	21H	1	1
S4	FUI SI	Module 1 : SI A-B	28H	3	3
		Module 2 : SI B-A	28H	3	3
		Module 3 : SI C-A	28H	3	3
	DU 1 Group Training	Module 1 : Group and Practicum Training in SI	56H	2	2
Total	5	09 modules	273H	4	2

We notice that SI is not taught in the core training. In sum, Table 4 below gives a summary of data on the number of hours, coefficients and credits per semester in the last curriculum, i.e. the third phase, by including all the modules:

Table 4: Hourly Volume of Credits

Semester	Number of Hours	Coefficients	Credits
S1	336 H	29	30
S2	413 H	49	30
S3	406 H	40	30
S4	357 H	34	30
Total	1512H	152	120

Keeping in mind that we are seeking to provide an accurate description of SI assessment where every detail counts, the HAIT’s academic programme can be read as follows: After completing successfully the core training in the first semester (S1), the trainees who are willing to choose interpreting as a ‘Degree Track’, (as ‘Major’ or ‘Specialisation’) in the second semester (S2) will have either to pass the ‘Interpreter Aptitude Test’ or to get just good impressions from the teachers as the test is not mandatory. How? The trainees will practice Simultaneous Interpreting just with a teacher in the ordinary classroom or by using the SANAKO lab 100 for a short trial period then they will be assessed either by impression or through the Interpreter Aptitude Test. This depends mainly on the availability of experienced teachers. The training period will sometimes last up to 15 days, not less than 6 hours per day. During this trial period, the students are sometimes given the autonomy to use the SANAKO lab at their own pace while being supported just by a technician in case of need. But, some classes confirmed that they were not given a chance to prepare themselves, maybe for time constraints or due to the impressions of the teachers who were in favour of the adequacy of the level of the students to start the interpreter training without further delay.

Eventhough the interpreter students at the HAIT will study in the formal programme three modes of interpreting which are simultaneous, consecutive and sight

interpreting, the Interpreter Aptitude Test focuses on Simultaneous Interpreting only. This mode is very difficult because it requires high-level skills (e.g., Listening and speaking at the same time). Consequently, the students who fail the test or feel that they cannot carry on after their self-assessment will join the translators' group automatically. In case they miss the regular formal exams in the written translation track, catch-up exams are administered to them. We will highlight more empirical data on the IAT and discuss its implications in chapters five and six.

If admitted in 'The Interpreting Class' in the second semester (S2) of the first year, the trainees will be introduced to SI through a course called 'Preparation for Simultaneous Interpreting'. Obviously, its aim is to familiarize the trainees with this kind of activity and its requirements by training them on interpreting small chunks of language. Consequently, there is no formal exam to assess the trainees, they are assessed by impression according to their performance during the different sessions of the course. The teacher assumes that, through the short activities of simultaneous interpreting which are done in class, he can assess the abilities of the trainees just by impression. We will discuss this way of assessment in the subsequent sections of this chapter.

Likewise, in the second year, which is divided into two semesters, S3 and S4, there are 3 exams in SI in each semester, corresponding to the three available language combinations (A-B, B-A, C-A). But, this number does not include all the modules. We have excluded from the description the modules related to "*Team-training*" where there are 3 more exams with the three combinations we have just mentioned. Therefore, in this study, we will focus on 2 tests and 3 exams only with emphasis on the combination A-B, i.e. from Arabic into English. However, there are some constraints that will be discussed later when we describe the size of the sample and its composition, especially regarding the directionality of delivery of the trainees.

According to Sawyer (2004: 1-7),

*"[...] assessment practices cannot provide guidance in the design and implementation of the curriculum if the purpose of the instructional*

*program – its aims and goals – has not been clearly defined. Glaser and Nitko (1971: 632) note: In an educational system, the specification and measurement of the outcomes of learning in terms of observable human performance determine how the system operates.”*

According to Sawyer, despite the increase in the number of translation and interpretation schools worldwide which rose from 49 to 80 between 1960 and 1980, and had reached a total of more than 250 by 1994 and 300 by 1998, a discussion of curricular diversity has been neglected in this community and in the scientific literature, more generally. He points out that the area of curriculum design is under-researched even though these institutions adopt a wide variety of curriculum models which differ substantially from one another, even within Western Europe. In fact, he sees that this neglect is due to the lack of reliable curriculum documents, because it would be a daunting task for an individual researcher to collect them from a range of schools of translation and interpretation, particularly those documents on course sequencing, which may not be available from a central administrative office on the school level. Rather, they are often developed and circulated within individual programs or departments. Besides, comparisons among schools is problematic, because the learning objectives may also differ among language combinations. (Pym, 1998; Caminade & Pym, 1998).

Nonetheless, Sawyer (2004) notes that researchers distinguish between two types of curriculum: Official curriculum and hidden curriculum. Some researchers, such as Freihoff (1995), venture to claim that the hidden curriculum (i.e. the curriculum that exists in the minds of the trainees or the students) is the only curriculum with practical impact on learning and on the educational outcome. He explains that the internal documents which are readily available to the public are often limited to an objectified, abrupt description that provides little information as to how the curriculum is implemented and how it is subjectively experienced by the teachers and the learners (Freihoff, 1995). Sawyer states that:

*“the hidden curriculum instills values and beliefs that shape future members of the professional community. If, for example, simultaneous interpretation into the non-native language is not offered officially in*

*the curriculum and remains in the hidden curriculum, students may come to believe that it is not a legitimate practice.” (p. 2)*

Therefore, our study is an attempt to delve into a part of the hidden curriculum at the HAIT. In order to start the empirical study, we have conducted a pilot study in order to gather some primary data on the active knowledge of the teachers regarding assessment matters. Then, we have elicited the students’ perceptions, views and opinions for the sake of triangulation. We will describe the design of the survey-questionnaires later. In the next section, we are shedding light on the issues of methodology.

#### **4.2. Research Paradigm, Approach and Methods**

In this chapter, we justify our methodology and research design, i.e. the decisions we have taken to align our research paradigm, approach and methods in order to increase the validity of our research and its findings. Indeed, as pointed out by many scholars (e.g. Creswell, 2008; Saunders et al., 2007), research design involves the interplay of philosophy and approaches in addition to the strategies of inquiry, or specific methods, and the practical techniques and procedures of data collection, because when planning a study, researchers need to think through the underpinning philosophical worldview assumptions which guide the study, the approach and strategy of inquiry related to this worldview and used to approach the issue and the research problem, and finally the specific practical methods, procedures and techniques of research which translate this approach.

Thus, the research paradigm or the philosophical underpinnings determine the way the research is carried out and its scope ontologically and epistemologically, and it can give justification for the logical limits of our investigation. Let us start by defining these terms to understand how they are connected and to grasp their impact on the orientation of our research. Firstly, according to Flowers (2009) and Blaikie (1993: 6, as quoted by Mack, 2010), ontology is “*the science or the study of being*”. In other words, it refers to the “*claims about what exists, what it looks like, what units make it up and how these units interact with each other*”. Ontology therefore refers to

the view on the nature of the ‘reality’ which exists ‘out there’ in the world. Secondly, epistemology refers to the conception that researchers have of what knowledge is and how It is acquired in our world, that is to say by experience, by reasoning or by research (Mouly, 1978).

Instead of the term paradigm, Creswell (2008: 6) uses the term ‘*worldview*’ and distinguishes four worldviews which are post-positivism, constructivism, advocacy/participatory, and pragmatism. He sees worldviews as the general orientation about the world and the nature of research that is held by the researcher. These worldviews or types of beliefs of the researcher determine the approach which can be quantitative, qualitative or mixed-methods approach. For instance, the positivist researcher prefers the quantitative approach and tends to quantify everything, because he thinks that the reality that exists “out there” in the world is ‘objective’ and that knowledge is based on its careful observation and numeric measurement. However, the postpositivist thinks that the reality that exists “out there” is ‘subjective’ and it cannot be explained just by numeric measurement, because there are many hidden things that the researcher must explore in the personal experience of the individuals.

The present research is based on a post-positivist research paradigm. It is exploratory and empirical, both quantitative and qualitative. The post-positivist paradigm fits the objectives of our study as we are interested in exploring the views, perceptions and opinions of the teachers and students of the HAIT about the current procedure of assessment of Simultaneous Interpreting quality. We are not interested only in gathering statistical or numeric data but rather in analyzing the textual data produced by the participants’ responses, i.e. their words, utterances, phrases, clauses and sentences, because they reflect their real views, perceptions and opinions regarding the assessment procedure in the way it is currently shaped through the different tests, i.e the admission test, the Interpreter Aptitude Test and the different exams of SI.

We have adopted a ‘holistic mixed-methods approach’ because of the difficulty we encountered in getting the required data for a more narrowed study with a very

detailed description of the curriculum and the assessment procedures, as we will explain later (Cf, Chapter five). Therefore, our approach is both quantitative and qualitative. This strategy of inquiry is appropriate for unexplored grounds. We would like to go behind the statistics, within our reach, to understand the issue under study as a whole by analyzing and interpreting the qualitative data.

Obviously, we relied on the review of literature and on introspection to introduce the topic as it is unexplored, but we tried all possible means to gather as much empirical data as possible at the HAIT through a questionnaire-based survey where the participants or the respondents are the teachers and the trainee interpreters. In addition, we tried to get in touch even with the graduates of the HAIT. Fortunately, we managed to get the responses of some of them.

Hence, our research is, mainly, inductive, because we intend to explore the reality of SI assessment at the HAIT as it is practiced to see how quality is seen and perceived by the different subjects of the study. We would like to extract the meaning they give to the assessment experience in order to go a step further the numeric data which might be gleaned easily.

Gile (1998) states that the main feature of exploratory studies is their endeavors with the analysis of situations and events in the given field without any prior intent to make a specific point, ask a specific question or test a specific hypothesis. This exploratory trend of investigation proves that hypothesis testing is not a sine qua non condition for "scientific work" especially that there are plenty of good results which were obtained through this approach in a wide variety of disciplines such as mathematics, medicine and psychology. Nevertheless, two points must be stressed:

*a. The absence of a well-defined and explicitly formulated hypothesis does not mean that exploration is done at random; some expectations always guide the investigator,*

*b. Exploratory projects can, and in fact most often do lead to precise hypotheses and their subsequent testing. (p, 2-3).*

A concrete example is provided by Pöchhacker (1995a,b) who, without a specific hypothesis to test, undertook an exploratory study of interpreting research productivity in the 1980's and 1990's by collecting publication data, and then analysing it in terms of individual authors, of 'schools' and of evolution over time. Then he drew a number of conclusions.

As for the research methods (i.e. the research strategies and choices, if projected to the research 'onion' of Saunders et al. (2007)), we use multi-methods which are Case Study, Survey and Grounded theory, in particular Content Analysis: Case Study method in order to focus on the teachers and students of the HAIT only; the questionnaire-based survey to collect qualitative and quantitative data from the subjects of our study. In the pilot study, the questionnaire has open-ended questions similar to the structured interview, as well as closed-ended and open-ended questions, because the primary data gave us some insights as we will explain below. Finally, Grounded Theory, because we did not, from the very beginning, intend to test any hypothesis hoping to generate some hypotheses for future research. The use of multi-methods is another way to ensure triangulation. As Olsen (2004) points out triangulation consists in mixing data or methods so that diverse viewpoints or standpoints cast light upon a topic. First, data triangulation helps in validating the claims that might arise from an initial pilot study. Second, a more profound form of triangulation is the mixing of methodologies, e.g. using survey data with interviews.

### **4.3. Research Instruments:**

The research instrument or tool used for empirical data collection is the 'Questionnaire'. Usually, it is used in quantitative research with close-ended questions, but open-ended questions were used for the pilot study in order to give an opportunity to the respondents (the teachers) to express themselves freely. In the follow-up study, open-ended questions were added. These questions were meant to facilitate responses by respondents, who are interpreter students and graduates, as we assumed that they lack awareness of assessment matters, particularly after the results of the teachers' pilot study, which were neither sufficient nor satisfactory. The questions were then

reformulated giving more explicit response options for the trainees and justification of their answers. Furthermore, the trainees were asked to add any ideas that would be beneficial for our research.

#### **4.4. Questionnaires' Design**

We designed five questionnaires to elicit data for different purposes: Two questionnaires for the pilot study and four questionnaires for the follow-up study.

The 'Questionnaire' was used as tool to collect qualitative data about teachers' views, perceptions and awareness regarding testing and assessment procedures the key words of which are 'quality, tests construction and design, constructs and rubrics'. The aim was to find out whether the teachers have background assessment-related knowledge as a field distinct from teaching and whether they are aware of the multiple dimensions in constructing and designing authentic, valid, reliable and practical tests and exams to assess interpreting skills. The 26 questions of the questionnaire used for the pilot study are shown in Table 13 which is appended (Appendix NO. 01).

The questionnaire used for the pilot study is anonymous and has open ended questions, because we did not want to suggest any option or predetermined category to the respondents. The latter were expected to supply their own answer without being constrained by a fixed set of possible responses (Bryant, 2012). We assumed that the teachers are knowledgeable enough to respond with the required brevity, clarity and precision when verbalizing their views, perceptions and opinion.

In addition, the research required concrete data; therefore, after the pilot study, we tried to collect evidence about the tests and exams to see how scoring was done in terms of principles of assessment from students' perspectives. The four student questionnaires are shown in Tables 14-17 appended to this research. The pilot study was a kind of pretest for the design of students' questionnaires, because there was no other way to pretest them due to the limited number of students at the HAIT. Presser et al (2004) state that pretesting is the only way to evaluate in advance whether the questionnaire will cause problems to the respondents and that the minimum of cases in

the pretest to reveal the major difficulties, flaws and weaknesses vary between 12 and 50. Unfortunately, the number of target students involved in this research is 53 only, not to mention the difficulty encountered in administering the questionnaires several times.

Moreover, we tried to design valid and reliable questionnaires by avoiding imposing our perceptions about the attributes of quality (Moser-Mercer, 2008). We gave the population of our study the freedom to define their constructs and to express their views and perceptions in order to explore the various subjective experiences.

Finally, it is worth recalling that three surveys only were devoted to the topic of 'quality' according to Pöchhacker (2011) who surveyed all the surveys on conference interpreting of the period spanning between 1930 and 2008. This highly prominent topic has already been addressed, with a focus on quality criteria, by only three surveys among conference interpreters to date (Bühler, 1986; Collados Aís, 1998; Chiaro & Nocella, 2004).

#### **4.4.1. Design of the Pilot Study (The Questionnaire to Teachers)**

The aim of the pilot study was to determine the degree of teachers' awareness about assessment principles and all the related matters (formative and summative assessment, traditional and alternative forms of assessment, etc). A questionnaire was designed to elicit their views, perceptions and opinions and their awareness of the educational criteria used to assess the aptitude, ability and achievement of the student interpreters. The questionnaire was also meant to assess the extent to which their tests and exams are authentic, valid, reliable and practical. Most of the questions are open-ended following a 'gradual logic' aligned to the assessment process. We assumed that the teachers would be able to spot this logic just by scanning the questions as these latter track the process of training at the HAIT from the phase of enrollement to the phase of graduation, where the trainees' exist profile is completely shaped.

As regards the design of the pilot study questionnaire, it consists of 26 questions divided into five distinct series of questions. The aim of the first series (Q1-Q5) is to

collect data on the background, qualifications and work experience of the teachers in order to see whether they were trained in assessment matters. The second series (Q6) aims at looking into teachers' conception of how aptitude should be assessed and whether they agree on a common vision which reflect the philosophy of teaching at the HAIT. The third series (Q7-Q21) aims at exploring how teachers assess the abilities of their students in their exams, i.e. how they design their exams and to what extent these exams reflect the principles of assessment. The fourth series (Q22-Q23) aims at determining whether the exams are interim or final, i.e. if they assess achievement skills and sub-skills. The last series (Q24-Q26) is meant to collect teachers' personal opinion regarding the procedure of assessment and for any suggestions that would improve the research in this field.

#### **4.4.2. Design of The Follow-Up Study**

The aim of the follow-up study is to elicit the interpreter students' views, perceptions and thoughts about the way they were assessed in the Entrance Test, in the Interpreter Aptitude Test and in S2, S3 and S4 during their training. This follow-up study is based on four questionnaires the design of which was well thought out for this purpose. Accordingly, the details targeted in each questionnaire correspond to the different tests and exams at different levels of degree study. Thus, two questionnaires were designed to the first-year students, one for S1 students (who were still in core training) and the other for S2 would-be interpreter students (those who passed the IAT). The third questionnaire was directed to the second-year students who validated all the modules in S3 and S4 but had not sat for their defense yet (they were still writing their dissertations). The last questionnaire was for the 39 graduates who completed their studies at the HAIT successfully before March 2014. The four questionnaires mentioned above allowed us to track the procedure of assessment from the beginning of enrollment and throughout the training in order to check whether there is a specific logic of assessment of aptitude, ability and achievement and to see whether the perceptions of the students match the perceptions of their teachers in a way which helps in improving the quality of the end-of-training performance and the quality of all the assessment practices.

The design of the four questionnaires is similar. They all include five series of questions. The first series contains questions about the background and qualifications of the students when they enrolled at the HAIT (age, sex, nationality, level of study and the languages they speak). The second series of questions is about the pre-tests. The questions aim at eliciting the students' thoughts before taking the admission test and whether they prepared for it. The third series contains questions about the different tests and exams (i.e. the tasks and exercises, the content, prompts, criteria, etc). The last series was devoted to the post-test or post-exam to see the kind of feedback students get from teachers. The last series is about the personal opinion of the informants.

All the questionnaires aimed at collecting qualitative data with mixed-item formats, that is why space for free expression was added in each question. The questions have options for ticking to help the students understand the aim of the questions, but our focus is on the views, perceptions and thoughts which follow each answer. We asked the students to justify their responses to see the meaning they get from the assessment procedure and how it affects their learning.

#### **4.4.3. Population of the study:**

The Participants of this study belong to two categories (i.e. the HAIT teachers and HAIT students). Most of the teachers are experienced interpreters who chose to have a second career in teaching for various reasons. As far as the 53 trainee interpreters are concerned, most of them hold BA degrees either in Translation, in Interpretation, in Translation and Interpretation or in Languages (Arabic, French or English). The combinations offered at the HAIT include English and French either as B or C, i.e. a second or third language. Due to the limited number of students in the Anglophone group, that is to say English is their B working language, the target population in this research form a mixed group of both Anglophone and Francophone students. In these groups, we may find students who come from 22 countries members of the League of the Arab States. So their mastery of the three languages is not at the same level. For instance, most of the Middle Eastern students do not master French.

## **4.5. Administration of the Questionnaires**

This section provides details on the administration of the six questionnaires (two to teachers and four to students). The administration concerns both the Pilot Study and the Follow-Up one.

### **4.5.1. Administration of the Pilot Study Questionnaire**

The pilot study questionnaire was sent to the teachers via email, because most of them are visiting trainers. The questionnaire was sent on February 6<sup>th</sup>, 2013. The first teacher completed it and returned by email on February 18<sup>th</sup>, 2013. The second teacher completed it and returned on March 2<sup>nd</sup>, 2013. The third teacher sent it back on July 10<sup>th</sup>, 2013. Practically speaking, the administration of the pilot study questionnaire and the collection of data from teachers took over 6 months.

Besides, the fourth teacher who is still teaching at the HAIT did not wish to take part in the research. The three teachers who accepted to respond expressed their difficulty to complete the questionnaire and said that they needed more time to fill in it. We will refer to the teachers who took part in the pilot study by TEA1, TEA2 and TEA3 following the date of reception of their answers.

During informal face-to-face chats, one of them, T2, said: *“We need to do research to do it”*. T1 said, *“I did not understand some questions”*. T3 just said, *“sorry for the delay”*. Thus, the second questionnaire was not administered, because all the HAIT teachers, including T4, did not wish to answer it. Therefore, the follow-up study questionnaires were administered just to the students and graduates.

### **4.5.2. Administration of the Follow-Up Study Questionnaires**

The four follow-up questionnaires were designed and distributed to the informants, because the data gathered from the pilot study seemed insufficient. Questionnaires 1, 2 and 3 were distributed and collected on-site, i.e. at the HAIT; whereas, questionnaire 4 which concerned the graduates was sent via email. The administration of the last questionnaire took longer time as it was difficult to get in

touch with the graduates or to get their emails. Besides, not all of them responded to our emails. Some of them promised to respond but did not.

### **Summary & Conclusion**

This chapter is devoted to methodology, design and procedure. We started by describing the situational context. Thus, we devoted the first section to the presentation of the Higher Arab Institute of Translation. Then, in the second section, a description of the programmes offered in this Institute was provided. In the third section, we gave the reader an idea about the different modes of interpreting as well as the current practices of SI assessment at the HAIT.

This chapter was also devoted to explaining and justifying our choices in terms of methodology and research design. We explained why this study is based on a post-positivist paradigm and mixed-methods approach; and described the research methods and instruments. The Pilot Study and the Follow-Up Study were described. Finally, the steps and problems in the administration of the questionnaires were described in detail.

This study aims at collecting quantitative and qualitative data. Three methods have been used for this purpose. These are Case Study, Survey and Grounded Theory. Six questionnaires with close and open-ended questions were designed, but five only were administered to the teachers, current students and graduates from the HAIT. The HAIT teachers did not wish to take part in any follow-up studies.

# **Chapter Five:**

## **Presentation of the Results**

This fifth chapter reports the qualitative coding and analysis of the data collected through our pilot study regarding teachers' views and perceptions. In addition, we will deal with quantitative data of this study. Moreover, we will report results of the coding and analysis of the follow-up study data concerning the students' views and perceptions.

## **5.1. Qualitative Analysis of the Pilot Study Responses**

In this section, we will deal with the analysis of the results and report the findings drawn from the pilot study. Ritchie and Lewis (2003) points out that data analysis must be marked by two key and distinct stages: 1) Data management; 2) Making sense through description or explanation. Despite the difficulty to separate interpretation from the assignment of meaning during the analytical process, data management should precede; otherwise, the collation, review or understanding of the material collected will be impossible.

Hence, we will deal with the inscription and coding of the raw data in the section below then we will tackle the selective, descriptive and explanatory analysis in the next sections. We will see the interpretation of data in chapter six.

### **5.1.1. Coding of Teachers' Responses**

Coding of qualitative data is a cornerstone in qualitative research. Strauss and Corbin (1990, as cited in Saldana, 2009) identify two types of coding: open coding and axial coding. Axial coding is divided into two kinds, hierarchical or non-hierarchical. Saldana (2009) distinguishes between first cycle coding methods and second cycle methods. The first cycle codes are holistic, descriptive and provisional; whereas, those of the second cycle are more focused, axial and elaborative.

In this research, the coding of the dataset was divided into five steps or stages. Firstly, as a thematic coding had already been selected for the pre-set themes and categories underlying the questionnaires' items, we just re-wrote them side by side with the relevant responses. Secondly, we extracted the key terms from the teachers'

responses to formulate the initial labels through an open coding (Ritchie & Lewis, 2003). Thirdly, we used an axial coding in order to reduce the long list of codes into a shorter list of themes and categories and to handle them into an in-depth analysis. Fourthly, we proceeded to the selective coding which is linked to the main research question. From the rich data that we harvested, many open and axial codes and sub-codes were identified, so we had to reduce some of them into more significant themes and sub-themes. For space constraints, the open codes are not displayed. The emerging axial codes and sub-codes are displayed in Table 13 (appended) and the emerging themes and sub-themes will be discussed in the selective, descriptive and explanatory analysis section. Thereby, we exploited them in the analysis and communication of data to gain more insights and later, in the interpretation and discussion of the findings. Some codes are abbreviated to gain space due to the layout and size of the paper.

It is worth noting that the open coding is part of our data analysis and reflects the decisions made about how to deal with the data. We did not use any computer-assisted software, rather we analyzed the verbatim textual responses of the teachers line by line and chunk by chunk and identified many codes and sub-codes as shown in Table 13. These codes fall within the pre-set themes and categories that are parallel to the process of assessment at the HAIT described in chapter 4, but do not necessarily fit into them. For the sake of brevity and a more focused analysis, we tried to reduce some of these codes into a small number of axial themes and categories which reflect the main patterns of how the teachers view the assessment experience then we condensed them. Hence, in the analysis section, we will comment on how we summarized the prevalence of codes, discuss the similarities and differences in related codes and compare the relationship between one or more codes.

Table 5 below shows the pre-set theme and categories, an example of the moves from the open codes to the axial codes and our own reflective commentary. The responses concern Question 11 about the tasks or exercises given at the SI test or exam to assess interpreting quality.

Q11 is as follows: What kind of task is appropriate to assess interpreting quality as you perceive it? (cf. Table 13).

Table 05: An example of thematic, open and axial coding (using verbatim textual data)

<b>Pre-set Theme &amp; Categories:</b>	<b>Open Codes</b>	<b>Axial Codes</b>
<b>Theme: Tasks &amp; Exercises</b> <ul style="list-style-type: none"> <li>• According to the level of study</li> <li>• According to the level of performance</li> <li>• According to the level of proficiency</li> </ul>	<ul style="list-style-type: none"> <li>• SI test (a)</li> <li>• Synthesizing (b)</li> <li>• Flexibility to unconventional contexts (c)</li> <li>• Team work (c)</li> <li>• Entrepreneurial spirit (b)</li> <li>• Comprehensive research (b)</li> <li>• Sight translation (b)</li> <li>• Listening for gist (b)</li> <li>• Reformulating ideas in SL (b)</li> <li>• Reformulating ideas in TL (b)</li> </ul>	<ul style="list-style-type: none"> <li>• (a) SI activity tasks</li> <li>• (b) Preparation for SI</li> <li>• (c) Real-life tasks that supplement but do not supplant SI tasks</li> </ul>
<b>Commentary</b>		
<ul style="list-style-type: none"> <li>• According to the open codes, the Teachers do not classify these tasks in a specific order and do not seem to have a common vision about how to go about them.</li> <li>• Not all these tasks are relevant to the practice of SI per se. They enhance the interpreters skills but are not linked to the mechanism of 'listening and speaking at the same time' in order to render a message from the source language A to the target language B with the same impact on the TL audience. [As they have said as a response to Q8, code: Quality Assessment]</li> <li>• These tasks can be classified into the categories: (a), (b) and (c). (axial codes)</li> <li>• These tasks do not take into account the set of categories that were established under the label of the theme 'Tasks &amp; Exercises'. They are not planned according to the level of study, performance or proficiency.</li> </ul>		

The full verbatim textual data (teachers' responses) and their suggested codings are appended in Table 13. The aim is to enable the reader to connect the dots between the different pre-set themes and categories and the emerging codes, because the responses are displayed in full and not as selective quotations. Additionally, displaying data in this way enabled us to compare them with the students' responses easily, because the responses are arranged according to the hierarchical pre-set themes and categories, so it is easy to spot the perceptions in any step of the assessment process.

Note that the term 'teacher' is shortened to 'TEA' in addition to a number, '1, 2 or 3', corresponding to the order of reception of teacher responses. In this way, we abide by the ethics of keeping the identity of the teachers anonymous and confidential.

The next section deals with teachers' responses analysis.

### **5.1.2. Analysis of Teachers' Responses:**

For the sake of objectivity and research validity, we strived to ask ourselves about how to analyze the teachers' responses in a way which distinguishes the analysis from our own interpretation as much as possible, because, as it is known in this kind of research, the coding of data is interpretive by nature and involves the researcher by hook or by crook. Hence, the central question which determines the objectivity of the analysis is: How is the data analyzed? To avoid subjective interpretations of the data, we focused on what the data say and not what we think of it. More specifically, we asked these three questions: What are the qualitative data? What are the perceptions, the views, and the opinions of the teachers? What is the subjective experience of each of them? In order to answer these questions, we perused the teachers' responses and tried to see the possible patterns in their answers through the axial codes and sub-codes that were used in the previous section. Then, via a selective coding, we extracted the patterns representing the similarities in teachers' perceptions, views and opinions and dealing with their distinct subjective experiences. In the next sections, the findings will be interpreted by projecting them on the thematic codes and discussing them in light of the available literature on the educational measurement of trainees' performance with the view to generating hypotheses.

According to Brown and Abeywickrama (2010: 346), measurement refers to “*a process of quantifying a test-taker's performance according to explicit procedures or rules*”. This definition has worked as a guideline for our thinking, because we are interested in exploring the extent to which the assessment procedures are explicit and whether they are planned and organized in a well-defined process that is clear to both teachers and students. Besides, we will ascertain whether the tests and exams have a good score in terms of assessment principles.

Our conceptual and intellectual processes in addition to the levels of abstraction that guide our search for the patterns of teachers' assessment practices are based on the literature and on grounded data. They are also based on deduction as a guideline for the process of assessment, but mostly on induction. So, in addition to the projection of

data on the pre-set thematic codes, we looked for the emerging variables and patterns in these responses: What are the similarities and differences in the teachers' responses? It is worth pointing out that we were mainly interested in concepts; so, we stated our own ideas and comments when we had just hunches or thoughts about something to distinguish them from the analysis of data and the arguments extracted from the literature.

Therefore, after scrutinizing the thematic coding, the open coding and the hierarchical axial coding of the empirical raw data, we concluded that they provided insightful evidence about many themes and sub-themes. However, not all these themes and sub-themes will be presented in this chapter. For the sake of conciseness, we drew a schematic shape. Indeed, through the non-hierarchical axial coding and the findings of the literature and our memos drafted while examining data, we described all these themes and sub-themes, or codes and sub-codes, according to their relationships. These are presented below as categories and sub-categories. We created links between them mainly via the central concept of 'test specifications' which is crucial in the construction and design of tests and exams. This term is in turn based on four axial concepts which are 'purpose', 'constructs', 'criteria' and 'scoring'. Our story line (Gipps, 2010) was built upon these main concepts and was further developed around a selective coding of some interrelated categories emerging from the data.

The results are shown below according to the hierarchical and non-hierarchical axial coding.

#### **5.1.2.A. Results of the Hierarchical Axial Coding**

Below, we will deal with the two themes of the hierarchical axial coding: 1) Teachers' background; 2) Construction & Design of Tests & Exams.

##### **a) Theme 1: Teachers' Background**

Teachers' background will enable us to find explanations to their responses and make sense of their assessment experience, because their background shapes their frame of reference in assessment matters. It has a great deal of impact on their

perceptions and views about how the construction, design and implementation of the tests and exams should be. Undoubtedly, the teachers do have a very significant professional experience, which exceeds for each of them 15 years of busy and hard work (Cf, code: Professional experience), because interpreting is classified as the second hardest job the humankind has ever known. The three teachers merit the qualification of expert interpreters. TEA3 has 26 years of experience in which they distinguish between their local and their international one. TEA2 started working as an interpreter in 1979 and has 35 years of experience. However, in our coding, we distinguished between experience as a practitioner and as a teacher. The result is that teachers' experience in teaching SI varies (cf, code: Teaching Experience of SI). For instance, TEA2 does not give a precise duration for their teaching experience, TEA1 has 18 years of teaching experience; whereas, TEA3's experience in teaching SI is 8 years only. Besides, TEA2 does not mention the type of their professional experience in assessment matters (cf, code: Training in Assessment). TEA2 does not specify whether it was within a hands-on training or just through observation in the professional milieu as stated by TEA3. Therefore, as far as the background of the teachers is concerned, the main emerging theme is that the teachers do not have a sound theoretical background in assessment matters; they rely on their experiential background. This result is confirmed later by their answers, as shown via the subsequent codes, which do not reflect well informed perceptions based on scholastic knowledge regarding the construction, design and implementation of tests and exams. Rather, their perceptions are based on their perceptions on how a professional performance should be. Indeed, even TEA1, who had a 'training' in assessment matters during their MA course in Britain, does not seem to have a deep theoretical knowledge of assessment principles. For instance, TEA1 does not supply an accurate answer to the question about ascertaining if someone has the aptitude to be an interpreter or not. TEA1 says, "*Tests in the same activity (Simultaneous Interpreting)*". TEA1 does not say what type of tests is meant. The expected answer was to give a detailed account about the specifications of these tests (Cf, categories: criterion-referenced assessment, norm-referenced assessment, diagnostic assessment, direct assessment, indirect assessment, etc) as we will see in the following theme.

## **b) Theme 2: Construction & Design of Tests & Exams**

### **1. Sub-Theme 1: The Specifications of SIQ Tests & Exams**

The specifications encompass many other terms such as ‘purpose, constructs, criteria, format, content, tasks, items, scoring, etc’. Thus, the sub-themes were described as interrelated categories.

The results of the pilot study revealed the following:

- The teachers did not elaborate tables of specifications where one might find details about the purpose, constructs and criteria for the tests and exams.
- Despite the fact that they mentioned that the aim of the test is a key element in test design, the detailed purpose or objective for each test or exam was not specified.

Indeed, it was difficult to evaluate the usefulness of the test or exam in the absence of an accurate description of the purpose or objective. It was also difficult to understand the role of the exams beyond the fact that they are formal steps in the process of graduation.

- The teachers did not define and determine the construct(s) for each stage in the assessment process. They did not explain whether they see interpreting ability as a ‘unitary construct’, i.e. one global ability (Bachman, 1995) or as a set of distinct constructs (Cf, assessment of abilities).
- The selection of the prompts was based on the teachers’ intuition (Cf, prompts). It was not subject to an accurate and detailed description which corresponds to the different levels of study and to the content of each course. Besides, there was confusion concerning the prompt’s length. It spans between 5 and 15minutes in the exams, but in real life it is between 30 and 45 mins. This was revealed in the Students’ responses as we will see later.
- The teachers agreed that even if their exams are summative, they remain interim and do not reflect a final achievement or any specific exit profile.
- The teachers often used ‘should’ to refer or describe hypothetical situations or exams and not what they really did. For example, the training period outside the Institute was not part of their exams. Besides, they do not assign the students to write research papers and they did not use the silent booth.
- Teachers often mentioned the process of educating the interpreters and did not focus on their particular exams.
- Teachers did not use well-designed rubrics (Cf, sub-theme: Assess. Tool).
- Teachers did not agree on one way of scoring.

Thus, we can conclude that the teachers are not aware of the importance of preparing the blueprints of the tests (Cf, categories under test specifications). These results go in line with Zhang and Burry-Stock (2003)’s who found out that the self-perceived assessment skills are influenced mainly by teachers’ teaching experience and professional training. With reference to Public schools, assessment practices and

skills were found to be related but with different constructs. Assessment practices pertain to assessment activities while assessment skills reflect an individual's perception of his or her skill level in conducting those activities. This may explain why teachers rated their assessment skills as good even though they were found inadequately prepared to conduct classroom assessment in several areas (Gullikson, 1984). In this study, the specifications considered by the HAIT teachers are the external factors (Cf, code: factors) which affect the performance of the students but have nothing to do with the design of the exam per se such as the students' level, abilities, flexibility, knowledge of topic, lexis, terminology and so on. It is true that these factors play an important role but they do not refer to what is meant by the term 'specifications' in the assessment literature. In the next section, we examine categories and sub-categories focusing more on teachers' assessment criteria, but less on interpreting constructs.

#### **i. Category 1: Criteria**

Harvey (2014) defines criteria as the specification of elements against which a judgment is made. At the HAIT, teachers did not specify the elements against which their judgments were made and no details were provided. Instead, the teachers suggested different short lists of broad criteria corresponding to a given hypothetical professional profile of their individual experiential perceptions. Sometimes, as we will see later, it is hard to differentiate between the criteria and the abilities for which they are set. In other words, there is confusion between criteria and constructs. At all the stages of evaluation described below, the teachers' assessment seem to be criteria-referenced assessment, not criterion-referenced assessment (Cf, code: Assessment of abilities). Criteria-referenced assessment refers to the process of evaluating and grading the learning of students against a set of pre-determined criteria (Harvey, accessed online 2014); whereas, criterion-referenced assessment is the kind of assessment which measures a candidate's performance according to a unique predetermined criterion or standard. The latter is "*a test designed to give test-takers feedback, usually in the form of grades, on specific courses or lesson objectives; the distribution of students' scores across a continuum may be of little concern*" (Harmer,

1983: 347), because no numerical marks are assigned to the performance. Only a 'pass/fail' basis or, in modern parlance a 'competent/not-yet-competent' basis" is used (CELT, 1998). Therefore, the main difference is in the number of criteria shaping the intended profile, or the criteria which are set to describe the abilities at a given level of study or performance.

### ✓ **Sub-Category 1: Assessment Criteria for Interpreting Aptitude**

Teachers focused on assessment criteria rather than the constructs. However, their responses in this theme are vague. TEA1 had no clear idea about the format and content of the suggested test to assess aptitude in SI. TEA2 suggested broad criteria and did not explain how they can be implemented in the Interpreter Aptitude Test. The criteria concern the assessment of basic and cognitive skills in addition to general knowledge but the construct for SI was not clear and the criteria not relevant to practice. Unlike TEA2, TEA3 stated that assessment of aptitude to interpreting is separate from ascertaining whether the student has the pre-requisite skills but TEA3 did not say how to assess it. Consequently, we concluded that the teachers did not have a common perception about interpreting aptitude or aptitudes and they did not share a common view on how to assess that trait. It is clear that these diverging views in the first stage of the assessment of aptitude(s) was likely to affect the assessment of IQ in the other stages.

The next section deals with the SIQA criteria in order to determine whether the teachers consider them from a pedagogical perspective.

### ✓ **Sub-Category 2: SIQA Criteria**

Teachers' responses showed that they paid little attention to the issue of SIQA. TEA1 gave a procedural definition of quality which refers to a broad hypothetical perfect performance without linking it to any context. TEA2 gave no definition. TEA3 limited 'quality' to 'faithfulness' only, which is also a broad definition. By answering Q4, TEA1 defined Interpreting Quality as "*the ability to respond at the right time in a different language and to the point according to the task*". This definition is operational but it is neither clear nor comprehensive, because it does not describe the

constructs' benchmark, the task and the criteria for a successful performance of the task in an educational setting at a specific level of study. It does not reveal the exact reality of the job of interpreting and it is not related to any educational or learning objective. There is no indication of the skills, sub-skills or the criteria needed for this implicit hypothetical ability to respond at the right time. In addition, it is not clear whether this perfect response can be attained or not, especially due to the fact that Interpreting Quality is not linked to a particular context of situation. As a result, these vague definitions of quality are likely to make its assessment difficult and this implies that quality is not well assessed.

The main criteria evoked by the three teachers are 'consistency' and 'impact of the message'. However, this perception is not process-oriented (Cf, code: product-oriented) and it does not distinguish between the learner level and the professional level. Even TEA3 who referred to the classroom setting gave a broad perception about rendering the gist of the message and not its words. This corresponds to Kurz's (1986) users' perception criteria. Teachers do not explain the nature of the product and how faithfulness or consistency can be achieved (Cf. sub-code: theoretical, classroom and market products). TEA1 said that the response to Q5 is the same as to Q4, which means that the point of the question was missed. TEA1 added one criterion only which is "*the ability to convey the message in order to have the same impact*". The aim of the question was to know how each teacher assesses Interpreting Quality starting from their definition. That is to say, at postgraduate level of study, the teacher was expected to mention some specifications of assessment such as its purpose or objective, its constructs and criteria based on psychometrics or on educational assessment and not just on impression. TEA2 listed a set of broad criteria in addition to sense consistency which are cohesion, correct grammar, completeness of interpretation and a pleasant voice. TEA2's assessment is based on an integrative assessment style. As a result, the emerging theme which deserves further investigation is the impact of adopting different scoring styles.

### ✓ **Sub-Category 3: Other Unassessed Criteria**

In addition to the assessed criteria, the teachers were asked about the unassessed ones. TEA1 did not have a view. TEA2 gave a list including the booth's behavior, interaction with colleagues, calm and delivery, and finally, avoiding embarrassments when dealing with jokes and cultural references. TEA3 mentioned the code of conduct, stamina and dealing with accents. However, the teachers gave no reasons for dismissing these criteria, which we consider as 'abilities', from their tests and exams. Hence, what was their priority in choosing the skills and criteria in the training and the procedure of assessment? How did they guarantee authenticity?

### ✓ **Sub-Category 4: Authenticity & Assessment Criteria**

According to TEA1, the authenticity of an interpreting test is determined by the aim of the test only. For TEA2, the test is authentic if it is adapted to the specificities and shortcomings of the students. It is true that, in educational assessment, one of the first and most important steps in evaluation is to assess the adequacy of course objectives to students' needs, background characteristics, and goals (Genesee & Upshur, 2001), but if the assessment criteria are relevant to high-order expectations in order to meet a perfect professional profile, it would be difficult to accept the argument of ensuring authenticity by adapting the test to the prompts. Only TEA3 believed that the use of authentic texts ensures the authenticity of the test. Therefore, the main emerging theme which is debatable is adaptation, because in the real-life job, the students will not have this indulgence, especially as far as the prompts' length, content and rate of delivery are concerned.

### ✓ **Sub-Category 5: Certification Criteria**

TEA1 believed that the test results would be sufficient for certification and for getting the qualification degree. TEA2 explained that certification is determined by the requirements of interpreting in the given country in addition to qualifications and the comprehensive standards. TEA3 stressed the satisfaction of the conferees, the peer-to-peer evaluation and membership in specialized associations. No teacher mentioned the need for proficiency tests. According to them, university qualifications, peers' and users' opinion, in addition to associations' membership would be enough as proof of

certification in the interpreting industry. TEA3 went further by saying that just by satisfying the clients, one would consider himself certified. Nonetheless, in reality endorsements are still a key in this sphere. These responses raise the confusing question of the difference between qualification, certification and accreditation. These terms are sometimes used interchangeably, but in most cases they lead to confusion.

ii. **Category 2: Constructs**

This category contains 6 sub-categories.

✓ **Sub-Category 1: Interpreting Aptitude Assessment Constructs**

Surprisingly, no constructs were mentioned regarding the assessment of interpreting aptitude.

✓ **Sub-Category 2: SIQA Abilities Constructs**

TEA1 and TEA2 mentioned pre-requisite skills such as mastery of languages but said nothing about the SI constructs, neither the simple nor the complex ones. TEA2 mentioned also broad personality traits that are not limited to the interpreters and to the interpreting activity. TEA3 did not answer the question.

The questions one might raise with respect to the competences and performance skills assessed by the teachers can be divided into two categories. Firstly, is there an interpreting competence or competencies? Is translation competence different from interpreting competence? Is interpreting competence assessed? How? In language assessment, competence refers to *“one’s hypothesized (empirically unobservable) underlying ability to perform language.”* (Brown & Abeywickrama, 2010: 347). In the context of interpreting, it might refer to one’s hypothesized underlying ability to perform interpreting. Secondly, to what extent is the difference between interpreting competence and interpreting performance considered in an educational setting? By looking into the other codes (e.g., Cf. Theme: Perception of Interpreting Quality; code: Product-oriented assessment; Process-oriented assessment, etc), we see that the teachers’ perception of competence is holistic:

- The teachers focused on one conception of interpreting competence which consists of being consistent in rendering the message and producing the same impact on the TL audience.
- Despite the fact that the teachers' responses reflected some of the competences mentioned in the assessment models, they did not refer explicitly to specific models of interpreting competence such as the ones described in the review of literature.

### ✓ **Sub-Category 3: Task-based Assessment Constructs of SIQ Exams**

As far as the task-based assessment constructs are concerned, TEA1 gave a broad answer in which there is no distinction between the test and its tasks. In language assessment, a task is “*a set or subset of linguistic actions that accomplish a real-world purpose, problem, or demand*” (Harmer, 1983: 353). If we replace the term ‘linguistic actions’ by the term ‘interpreting actions’, one might devise a definition for ‘interpreting tasks’. TEA2 mentioned a number of interesting tasks, e.g. synthesizing, working in teams and writing research papers. However, TEA2 did not describe any task on the real SI activity. That is, SI was not broken down into distinct tasks corresponding to the different levels of study and performance of the action of interpreting. Instead, TEA2 suggested some tasks which can prepare the students to undertake interpreting. TEA3 mentioned sight translation and reformulation in SL and TL, but these are additional tasks which are relevant to the preparation period and not the real activity of SI.

TEA1's response to Q8 clearly shows their little knowledge of the assessment jargon as Q8 was about the tasks which are appropriate to assess Interpreting Quality. What was expected in the response is a description of the tasks used to assess the aptitude or abilities of the trainees according to the different levels of study. TEA1 mentioned “*Tests in interpreting in the field of knowledge*”, but the term ‘test’ refers to a type of assessment which is part of the whole evaluation but not the tasks of the test itself. We can infer that TEA1 means Interpreting as a task, but even in interpreting, there are many possible tasks.

It is worth noting that there is reference to the level of SI proficiency and not to the level of SI performance in the classroom that is based on the completion of specific tasks. As for the prompts' length(s) which is/are chosen for the tasks, we notice the

broadness of the responses as the teachers did not refer to a specific context and situation of assessment. According to TEA1, the prompt's length is determined by the aim of the test and the tasks of the real-life job. TEA2 mentioned the trainees' abilities and standard time in conferences. TEA3 gave no details.

#### ✓ **Sub-Category 4: Performance-based Assessment Constructs of SIQ**

All teachers speak about their expectations by referring to the task of 'rendering the message'. It is not clear whether this refers to the meaning of a sentence or the whole speech. The literature on language assessment distinguish between task-based assessment and performance-based assessment. Performance-based assessment refers to the "*assessment that typically involves oral production, written production, open-ended responses, integrated performance (across skill areas), group performance, and other interactive tasks*" (Harmer, 1983: 351). That is to say, a set of performances are measured to complement each other. But, this distinction does not appear to be made by the teachers questioned.

TEA1's expectations are the competence in two languages and the command of the topic. This describes the requirements but not the performance expectations. TEA2 mentioned five expectations, namely mastery of terminology, communication skills, accuracy, sensitivity to culture and self-confidence. TEA3 listed accuracy, concentration, speed of analysis, having enough lexis and controlling delivery (voice, pace, body language, etc). But, these are broad expectations which are not divided into high-order expectations and low-order expectations. Besides, no teacher suggested a taxonomy of performance expectations according to the level of study or performance.

#### ✓ **Sub-Category 5: Skill Development & the Training Period' Constructs**

The teachers were asked these questions: how are the exams viewed by the teachers? Are they Mid-term, i.e. Interim tests, or Achievement tests? The responses revealed that all the exams are mid-term or interim exams, because they take place at the end of a fortnight period of training in interpretation for a given combination. But, the objective of these questions was to determine the construct(s) of the exam(s) and

the teachers' perceptions regarding skill acquisition, and finally to see whether there is a difference between the first year and the second year exams in terms of quality assessment of these skills.

The teachers' responses are not clear. TEA1 said that "*It depends on the period of training and the results obtained*", TEA2 referred to the whole training and TEA3 mentioned the requirements to avoid possible defects such as faulty grammar, poor lexis or lack of self-control before the beginning of the training period. Obviously, either the teachers did not understand the question, because they are referring to holistic assessment and the whole process, or they believed that when the requirements are met, the global interpreting ability can be developed and assessed in all the exams.

#### ✓ **Sub-Category 6: Exit Profile Performance Constructs**

According to the teachers questioned, the exit profile can be assessed through tests (TEA1), conferences (TEA2), summative and formative assessment in addition to mentoring (TEA3). Hence, this can be grouped in three categories: Tests & exams, practicum and mentoring. Besides, teachers' criteria for the final performance are keeping calm, being faithful and meeting the audience expectations.

#### **iii. Category 3: Scoring**

The literature on the factors affecting scoring mentions the following: Constructs, criteria, tools of assessments, tools of archiving students' performance, assessment and exam type, etc.

#### ✓ **Sub-Category 1: Scoring & Assessment Criteria**

Teachers' criteria were not classified in any given order and were different from the criteria mentioned in the literature in terms of importance. For example, TEA1 listed knowledge of language, fluency, factor of time, reception of audiences (i.e. impact of the performance). TEA2 mentioned accuracy, grammatical coherence, cohesion, pleasant voice and time. TEA3 limited the criteria to rendering the key ideas. Amazingly, these criteria are completely different from the criteria that they set previously to assess SIQA (e.g., presence of mind, information and nuance analysis, faithfulness, adaptation to changing contexts, ability to respond at the right time, etc).

Besides, the results of the study revealed two types of scoring: impression and a 10 point scale. In impression, TEA1 used holistic scoring and TEA3 used analytic scoring where the sum of sub-scores for delivery, content, voice, etc. were taken into account.

✓ **Sub-Category 2: Scoring, Assessment Tools & Archiving of Students' Performance**

TEA1 used marks and 'notes' (i.e. TEA1 took notes while listening to the students' performances then gave them verbal comments), TEA2 used recordings of performance and TEA3 used additionally checklists. Therefore, only one teacher used recordings. TEA1 did not record the students' performance at all; whereas, TEA3 used recordings just when the lab is available, i.e. TEA3 did not rely on recordings as an essential archive tool. Instead, they used notes and checklists to keep track of their evaluation. We asked these questions: How did they measure the previously-mentioned constructs and criteria? How did they compare the performances of the students with the view of a criteria-referenced and a norm-referenced assessment just by impression? As we have mentioned above, they used mainly holistic and analytic scoring.

✓ **Sub-Category 3: Scoring, Type of Assessment & Exam type**

TEA2's responses indicated that their assessment is summative only; whereas, TEA1's and TEA3's assessment is both formative and summative. However, TEA2 did not explain what unit is meant to be ended. Besides, TEA3 regarded assessment as formative when covering one full academic year and not a fortnight period. TEA3 explained that "*every performance should be graded to enable students to appraise their own learning process*" (Cf, code for Q17). However, in practice, there is no connection between the different courses, because each teacher seems to deal with course content independently from the other teachers. So, it is not clear how teachers ascertain the formative aspect of their direct formal assessments.

All the teachers confirm that their exam is interim despite the fact that TEA3 does not use this term. According to TEA2, the conference would be the final

assessment. Yet, TEA3's response is problematic, because there is confusion between assessment type and exam type. TEA3 states that their assessment is summative and formative. An exam can be summative with the purpose of being 'interim', so summative does not mean final as long as it is not connected to a final performance achievement at the end of the training period.

## **2. Sub-Theme 2: Alternative Forms of Assessment**

As far as alternative assessment is concerned, the teachers questioned did not use any of the forms mentioned previously. TEA1 stated clearly that no alternative form was used; whereas, TEA2 used research papers and performance in unusual situations such as interpreting for a demanding audience and TEA3 used interpreting in conferences in the silent booth. TEA2's response is consistent with their response to Q11 concerning the different tasks used for assessment. However, TEA2 did not describe the main tasks used to assess SIQ. Besides, research papers do not involve performances using the spoken language; therefore, the construct does not seem to be appropriate. In practice, it is difficult to predict the number of conferences the students should attend in order to be assessed by this alternative form especially that booth use in conference interpreting requires hard training.

### **5.1.2.B. Results of the Non-Hierarchical Selective Coding:**

The non-hierarchical selective coding consists of linking data extracted from different places in the process or the responses to the questionnaire. This type of coding should provide evidence concerning teachers' awareness of purpose, constructs, criteria and scoring. The following sub-sections examine these four themes:

#### **a) Theme 1: Purpose**

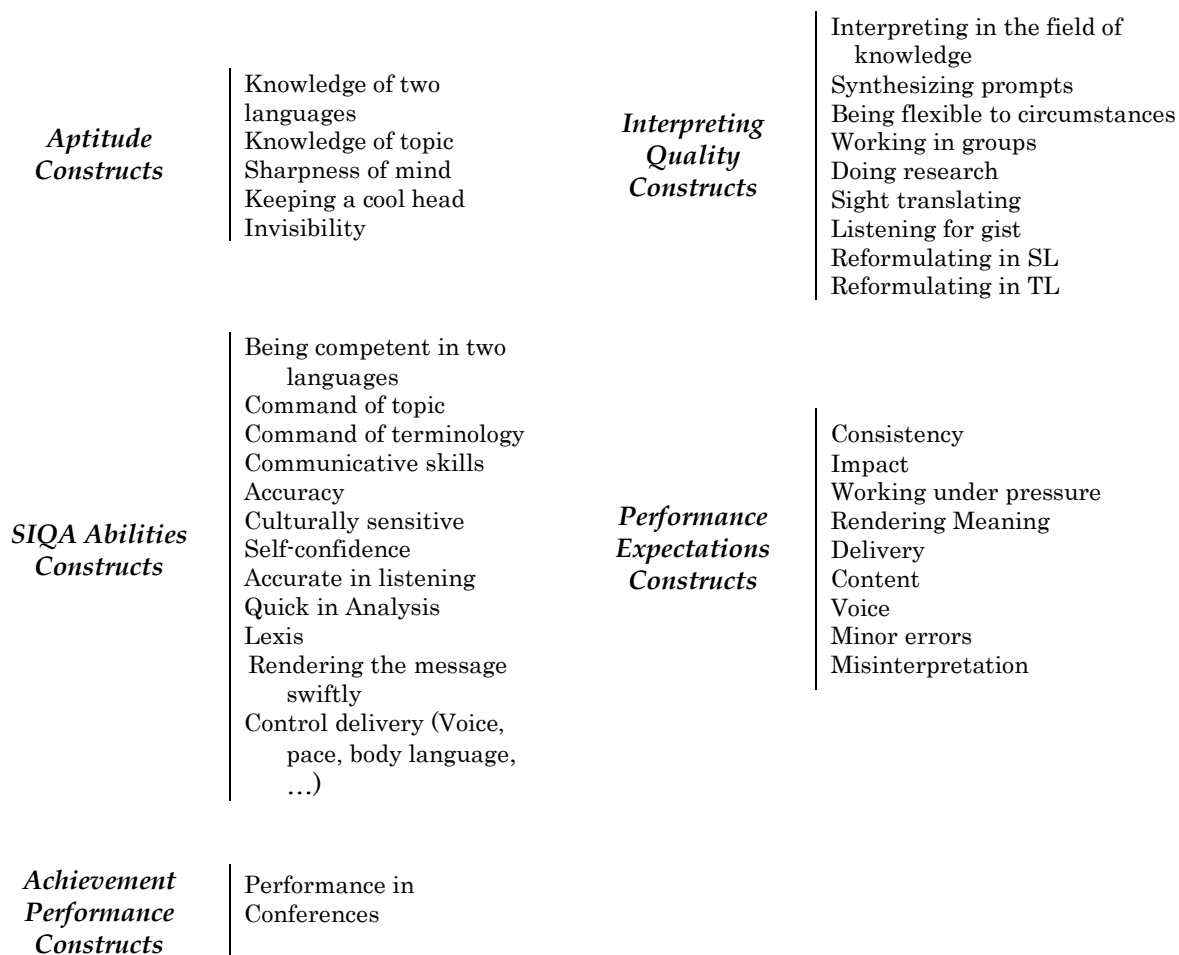
Looking at teachers' data, we can infer that the aim of the IAT and the different exams is not well formulated. Even if we take it for granted that the aim of the IAT is to screen promising candidates in interpreting, this test does not inform the teachers about the students' needs and abilities.

Besides, it is not clear from the teachers' responses that their criteria-referenced assessment is appropriate for the exams.

**b) Theme 2: Constructs**

The constructs mentioned by the teachers reflect the five key ideas or 'memes' that geered theoretical inquiry into interpreting: Processing, verbal transfer, making sense, text/discourse production and mediation (Chesterman, 2000). However, from the teachers' responses, SI is not broken down into meaningful educational constructs for building the students' interpreting competence. The teachers listed at random some broad and mixed abilities which do not seem to reflect distinct targeted levels of study and performance in SI. These abilities were provided by teachers wholesale without any pre-determined order. From their raw data, a taxonomy of constructs can be extracted. It is displayed in Diagram 3 below.

Diagram 3: Taxonomy of Assessment Constructs from the Teachers' Perspective



One may notice that this taxonomy contains no benchmark. The assessment constructs are introduced wholesale. A benchmark is “*a point of reference against which something may be measured.*” (Harvey, accessed online 2014). Indeed, in a benchmark, one might identify landmark points on a continuum or a scale toward a given goal or standard. These benchmarks may be used to describe content standards when interim targets of development have been set by age, grade or level. In educational assessment, the term is used interchangeably with ‘anchor papers’ or ‘performances’ which illustrate points of progress on an assessment scale (i.e., student performances which exemplify the different levels of a scoring rubric) (SCASS/AEAC Glossary of Assessment Terms). In higher education, a clear benchmark statement must be established in order to provide “*a reference point against which high outcomes can be measured*”. This statement refers to “*a particular specification of programme characteristics and indicative standards*”. (Harvey, accessed online 2014). Therefore, benchmarking is “*a process that enables comparison of inputs, processes or outputs between institutions (or parts of institutions) or within a single institution over time.*” (Harvey, accessed online 2014).

For example, ‘*invisibility*’ cannot be regarded as an assessment construct to screen the aptitude of the candidates for SI (Cf. code: Aptitude Constructs). ‘*Invisibility*’ should be taught in the different courses, because it requires specialized knowledge and training. In addition, the two broad psychological constructs ‘sharpness of mind’ and ‘keeping a cool head’ require deconstruction and narrowed criteria and cannot be assessed in a short time at once by impression.

Why is ‘voice’ assessed without its prosodic features? Prosodic features such as tone, stress, intonation and pitch may be assessed, but not natural voice which is a prerequisite. Besides, even if we take its assessment for granted, students should be trained to use it and change it according to the context of situation. We wonder whether TEA3’s course includes such training.

Q7 aims at delineating the construct for Simultaneous Interpreting in the 2<sup>nd</sup> year level exam; however, the teachers’ responses sounded ambiguous. It is true that

knowledge of two languages and topic are a requirement, but they are not the construct for SI. Does the teacher test the students to check their knowledge of two languages or to check their knowledge of a given topic in a SI Exam? This kind of test may be considered as tasks in the teaching process but cannot be regarded as the main construct for second year exams. Language is used to test interpreting, not to test language mastery.

The second taxonomy which reflects our selective understanding of these constructs is described in chapter six, because it shows our own interpretation of the results in light of the literature on standards-based assessment.

### **c) Theme 3: Criteria**

Forty macro-criteria were mentioned by the teachers at the HAIT to assess aptitude in the IAT, students' abilities in the different exams of Simultaneous Interpreting (first and second year exams), achievement performance and certification. Diagram 4 below displays an exhaustive list of criteria that we inferred from teachers' responses, because some of these criteria were not stated explicitly. We sifted them into categories with different headings (Aptitude Criteria; Interpreting Quality Criteria; Performance Expectations Criteria; Scored Criteria; Unscored Criteria; Certification Criteria) in order to establish a more analytical taxonomy in order to reach in-depth analysis of the assessment process and its requirements. It is worth reminding the reader that these criteria do not describe any interpreting constructs but they were stated by the teachers wholesale as features for performance.

As can be seen in Diagram 5 below, there is high variability among these criteria. Grbić (2008) also stresses the variability of the criteria used to gauge quality and Kellett Bidoli (2013) mentions various lists of quality criteria in her review of questionnaire-based surveys conducted by Moser-mercer (1985), Bühler (1986), Kurz (1988), Gile (1990), Meak (1990), Kopczynski (1994) to study end-users' perceptions. But, the issue is not the variability of criteria but how they are used to describe the constructs of interpreting for educational purposes.

Diagram 4: Taxonomy of Assessment Criteria from Teachers' Perspective at the HAIT

<p><b><i>Aptitude Criteria</i></b></p>	<p>Test in SI Complete mastery of the active language(s) Presence of mind The ability to analyze information Ability to spot the different nuances in the utterance The ability to adapt to changing contexts and situations Pleasant voice Public-speaking skills Powers of concentration regardless of the situation General Knowledge Checking pre-requisites (Grammar, lexis, self-control)</p>	<p><b><i>Interpreting Quality Criteria</i></b></p> <p>Sense Consistency (2) Impact on audience (2) Rendering the gist Cohesion of utterance (2) Correct grammar Completeness of interpretation Pleasant voice (2)</p>
<p><b><i>Performance Expectations Criteria</i></b></p>	<p>Knowledge of language Fluency Time Impact Accuracy Grammatical coherence Cohesion Pleasant Voice Rendering key ideas</p>	<p><b><i>Scored Criteria</i></b></p> <p>Mastery of languages Accuracy Fluency Consistency Impact of performance Time</p>
<p><b><i>Unscored Criteria</i></b></p>	<p>Booth's behaviour Stress Code of conduct Interaction Ability to deal with jokes &amp; cultural references Dealing with accents Stamina</p>	<p><b><i>Certification Criteria</i></b></p> <p>Test results Interpreting requirements Qualifications Standards Clients' satisfaction Peers' opinion International Associations' membership</p>

#### **d) Theme 4: Scoring**

When the exams are norm-referenced with numerical scores, this implies that scoring is dichotomous. As pointed out by Brown & Abeywickrama (2010: 348), scores are based on “*a method of scoring that allows only one correct response so that*

*a test-taker's response is either right or wrong.*” However, scoring in SI cannot be always dichotomous. One cannot have right or wrong simultaneous interpretation unless the choice of content is studied thoroughly and focused on a very technical topic in which errors have a significant impact on the rendition of the message.

At the HAIT, there seems to be no scoring guide in which one might find directions for scoring and for interpreting the scores, scoring criteria are used without descriptors, which are short explanations that define the levels of scoring scales. (SCASS/AEAC Glossary of Assessment Terms). No guide provides the general instructions for raters, training notes, rating scales, rubric, and assessing student work (SCASS/AEAC Glossary of Assessment Terms). The scoring criteria which must be used as rules or guidelines for assigning scores in the form of number or label to indicate the quality of a performance must be accompanied with descriptors. In the analytic scoring of a performance, multiple rules may be applied to the different dimensions or traits of the performance (SCASS/AEAC Glossary of Assessment Terms). The teachers questioned did not mention that they use descriptors for these scoring criteria.

According to Brown & Abeywickrama (2010: 349), grading means “*assigning a score to a test or a composite set of recorded assessments, usually by means of a letter (A through F)*”. There are two types of grading: Absolute grading and Relative grading. Absolute grading is “*a score on test-taker's performance which is empirically calculated by predetermined measures of achievement of learning objectives*”. Relative grading which is also known as ‘grading on the curve’ is “*a score on a test-taker's performance [which] is compared to other test-takers and sometimes altered to suit instructional needs*”

In a qualitative analysis, as a mode of inquiry, dealing with the subjective experience of each teacher alone in order to gain more insights is important. The teachers' subjective assessment experience enables us to conclude that they focused on training more than assessment.

Thus, TEA1 focused on the aim of the exam, the job requirements and reality. Indeed, TEA1 mentioned some interesting points needed for the blueprints (exam specifications: aim of the test, the abilities of the applicants and the would-be job expectations), but did not mention the type of content, criteria, types of tasks, description of the items and their format, length of the material, particularly the discourse prompts. Also, there is no description of the expected scores and how the performances are marked. TEA2 focused on the psychological threats mainly stress, load and embarrassments. TEA3 compared his/her performance as a professional interpreter to the performance of his/her students, but did not point to the mismatch between the educational setting and the professional one. If the exam takes place during the mid-term, are the students to be penalized for their weak performance?! Is the training period enough to judge their performance and compare it to that of a professional interpreter, especially that all teachers seem to agree that their exams are just interim exams.

The next section deals with the qualitative analysis of the collected data.

## **5.2. Quantitative Analysis of the Follow-up Study Responses**

It is generally agreed that a study sample must represent at least 10% of the target population. In this research, the sample represents 38%, because there are 27 respondents out of a gross total of 72 students under study. They belong to different levels: 38 students are graduates, 9 are in the second year ( they have completed S3 and S4 and are preparing their dissertations in partial fulfillment of their Magister degree), 6 students are in S2, 19 students are in S1 (i.e. 25 are in their first year). The sample is representative, because it represents almost all the classes from the creation of the HAIT in 2005 until 2014.

Statistically speaking, we notice that the number of students in the Anglophone groups is larger. Nevertheless, the collected questionnaires are from mixed-groups. In the second-year level, we collected 2 questionnaires from the Francophone group and 1 questionnaire from the Anglophone group. As for the graduates, 5 graduates have English as their B and 2 only have English as their C. One graduate did not mention

his working languages. In the first–year levels, the students study together, either in the core training in S1 or in the interpretation programme, in S2, due to the small size of the groups. Hence, they were not asked to mention their working languages.

Table 6 below shows the number of students per class and the number of participants who responded to the survey.

Table 06: Number of Collected Questionnaires per Class

Class	University Year	Number of First Year Students	Number of Interpreter Students in the English Group(s)	Number of Interpreter Students in the French Group(s)	Number of Distributed Questionnaire/Mixed Groups	Number of Collected Questionnaires
1	2005-2007		07	03	10	03
2	2006-2008		03	03	06	02
3	2007-2009		03	02	05	01
4	2008-2010		06	04	10 (-2*)	01
5	2009-2011		02	01	03	00
6	2010-2012		06	NK	06	01
7	2011-2013		NK	NK	07	03
8	2012-2014		03	03	06	03
9	2013-2015	19			19	13
Total					72	27

Key: NK=Not Known

The tables below display the respondents' biographic data such as gender, age, nationality, degree needed, mastery of languages.

Table 7 below shows the respondents' gender per level of study and per semester:

Table 07: Gender per Level of Study

Gender	S1		S2		S3 + S4		Graduates		Respondents	
	Num.	%	Num.	%	Num.	%	Num.	%	Num.	%
Male	2	11%	0	0%	1	11%	2	5%	5	7%
Female	11	58%	3	50%	2	22%	6	16%	22	31%
Total Resp.	13	68%	3	50%	3	33%	8	21%	27	38%
Population	19	100%	6	100%	9	100%	38	100%	72	100%

There is a predominance of females at all levels. 22/27 respondents are females. They represent 31% of the target 72 respondents. In the first-year level, the increase in the number of females studying translation and interpreting is apparent, because there are 11 females and 2 males only. One of the latter intends to study interpreting compared to five females who chose it.

Table 8 below shows the age of the respondents per level of study and per semester:

Table 08: Age of the Students per Level of Study

Age	S1		S2		S3 + S4		Graduates		Respondents	
	Num.	%	Num.	%	Num.	%	Num.	%	Num.	%
20....30	8	42%	3	50%	3	33%	4	11%	18	25%
30.....40	4	21%	0	0%	0	0%	3	8%	7	10%
40....50	1	5%	0	0%	0	0%	1	3%	2	3%
Total Resp.	13	68%	3	50%	3	33%	8	21%	27	38%
Population	19	100%	6	100%	9	100%	38	100%	72	100%

We notice a majority of young learners, only one student and one graduate are aged above 40 years old, which is not surprising. Interpreting is known to be the second most difficult job or activity; therefore, the learner should be young and ready to make a great deal of effort as explained by Gile (1991) in his model of effort. Nevertheless, in the core training, 12 students are under 40 years old. They study translation and specialize in translation in S2. Thus, it seems that either the professionals do not apply for the HAIT's degree to enhance their skills or the Institute opted for recruiting young students only.

Table 9 below shows the respondents' nationality, and not their origin, per level of study and per semester:

Table 09: Nationality of the Students per Level of Study

National.	S1		S2		S3 + S4		Graduates		Respondents	
	Num.	%	Num.	%	Num.	%	Num.	%	Num.	%
ALG.	10	53%	3	50%	3	33%	6	16%	22	31%
Other	3	16%	0	0%	0	0%	2	5%	5	7%
Total Resp.	13	68%	3	50%	3	33%	8	21%	27	38%
Population	19	100%	6	100%	9	100%	38	100%	72	100%

We notice that 22 respondents are Algerian and come from different places in Algeria. Only 5 respondents come from other Arab countries. Thus, the students have different sociocultural backgrounds, coming from Algeria or the Arab world.

Table 10 below shows the degrees submitted by the students to enroll at the HAIT:

Table 10: Degrees of the Students enrolled at the HAIT (ISAT)

Degree	SI	S2	S3 + S4	Graduates	Respondents	
	Num.	Num.	Num.	Num.	Num.	%
BA	11	3	0	6	20	28%
MA	3	0	3	1	7	10%
Other						
Engi.	1	0	0	0	1	1%
Add. BA	1	0	0	0	1	1%
Add. MA	0	0	0	0	0	0%
Field						
Translation	4	2	0	2	8	11%
Transl. & Interp.	1	0	2	1	4	6%
English	5	0	0	0	5	7%
TEFL	2	0	0	0	2	3%
Tech.	1	0	0	0	1	1%
Total	13	3	3	8	27	38%
Target Populat	19	6	9	38	72	100%

Table 11 below shows the students' mastery of languages:

Table 11: Students' Mastery of Languages at the HAIT (ISAT)

Lang. Mastery	S1	S2	S3 + S4	Graduates	Respondents	
	Num./19	Num./6	Num./9	Num./38	Num.	%
Level(s)	Arabic				Arabic	
1	0	0	0	0	0	0%
2	1	0	0	1	2	3%
3	4	3	2	1	10	14%
4	3	0	0	1	4	6%
5	5	0	1	5	11	15%
Level(s)	English				English	
1	0	0	0	0	0	0%
2	1	0	0	0	1	1%
3	8	2	2	3	15	21%
4	3	0	1	4	8	11%
5	1	0	0	1	2	3%
Level(s)	French				French	
1	2	0	0	0	2	3%
2	2	0	0	1	3	4%
3	3	2	1	4	10	14%
4	5	1	2	2	10	14%
5	1	0	0	1	2	3%
Level(s)	Other: Spanish				Spanish	
1	2	0	0	2	4	6%
2	4	0	0	0	4	6%
Level(s)	Other: Italian				Italian	
2	1	0	0	0	1	1%
Level(s)	Other: German				German	
1	0	0	0	2	2	3%
<b>G. TOTAL</b>	<b>19</b>	<b>6</b>	<b>9</b>	<b>38</b>	<b>72</b>	<b>100%</b>

It is worth mentioning that some spaces or boxes were left blank in the questionnaires by the respondents. One student did not mention his degree. So, as shown in Table 10 above, only one student has a university degree in Electrical Engineering. One student only has a BA in Economics in addition to the BA in Translation and Interpreting. There are 7 students who hold an MA and 25 students hold a BA degree. Their degrees are in Translation, Interpreting, Translation and Interpreting, English or English Linguistics and TEFL. Four students studied

translation and three students studied either interpreting or translation and interpreting. Seven first-year students hold degrees in English, English Linguistics or TEFL; whereas, five students hold degrees in Translation or Translation and Interpreting.

When asked about their degree of mastery in Arabic, the graduates said it is just ‘Good’, which is unexpected and quite problematic. If their mastery of the target language is not ‘Excellent’ or at least ‘Very Good’, this means that the Institute’s objectives are not attained, especially that the teachers say they focus on this criteria from admission until graduation as we have seen previously.

It is worth pointing out that most of the students’ questionnaires were returned incomplete, because many of them did not take the Admission Test and the Interpreter Aptitude Tests (IAT). Table 12 below shows the number of students who took the Admission Test and the IAT per level of study:

Table 12: Number of Students who took the Admission Test & the IAT (HAIT)

	SI Q	S2 Q	S3+S4 Q	Grad. Q	Total	Remark
Admission Test	09	/		01	10	The 2 <sup>nd</sup> -year students are not concerned
Interpreter Aptitude Test		03		02	05	
Total Resp.	13	03		08	24	
Population	19	06		38	63	

It is worth mentioning that second-year interpreter students were not asked about the Admission Test in order to help them focus entirely on the IAT, the first-year exam in SI and on the second-year level exams. The graduates were asked to respond to the questions on the two tests and all the exams with more details.

Seemingly, the majority of respondents did not take the Admission Test. In S1, four students did not take it and one graduate only took it. All in all, ten students took the Admission Test out of 24 respondents, which is strange. As for the Interpreter Aptitude Test, only five interpreter respondents took it!

The first-year students were questioned about the easiness of the Admission Test. Four students found it ‘Easy’ and three found it ‘Very easy’. In addition to the

graduate students who passed the Admission Test, two graduates found it ‘Easy’, even though they did not take it. So, they might be referring to the impressions of their classmates about it, or maybe to the tests of other classes. This is statistically significant, especially that two first-year students only found it ‘A little bit difficult’. Nobody found it ‘difficult’ or ‘very difficult’. This is significant even more by looking at the responses given concerning the easiness of the IAT as we will see in chapter six.

As far as the occupation of the eight graduates is concerned, six of them said they are in the translation business. But, three only are sworn translators and the other three translators are freelancers. In addition to translation, two of them said they have other occupations such as consultant or journalist. Moreover, one is an English teacher and the other is administrator.

### **5.3. Qualitative Analysis of the Follow-up Study Responses**

In this section, we report the results and findings of the follow-up study under three sub-sections according to the three steps followed for data processing. Firstly, we perused the verbatim responses of the students in order to code them. Secondly, we analyzed them in light of the emerging codes in search for possible patterns. Thirdly, we interpreted and discussed the findings in light of the available literature on assessment.

The next section deals with the first steps, i.e. the coding of students’ responses.

#### **5.3.1. Coding of Students’ Responses**

In Tables 14, 15, 16 and 17 (appended), the following codes were used: Thematic codes (the pre-set themes, sub-themes and categories), axial codes and sub-codes. These codes enabled us to analyze, interpret and discuss the results of this study. The open codes were not mentioned owing to space constraints and to keep the same layout used to present the coding of teachers’ responses. These tables show the pre-set themes, the verbatim textual responses of the respondents and the qualitative axial codes and sub-codes that emerged from the responses. These codes symbolize the

subjective experiences of the current students and the graduates of the HAIT. Therefore, any detail that we could spot was taken into account. Sometimes, as we will see in the analysis, some patterns will be formed. We will try to shed light on them before discussing the more specific details, because in qualitative analysis the focus is sometimes on the detail rather than the general patterns.

We will refer to the students by the shortening ‘STU’ plus a number. These symbols were assigned at random in order not to harm any respondent. Some codes are used in full once than abbreviated in the rest of the table for space constraints.

As mentioned earlier, there are four tables (appended) corresponding to the different levels of study. Three tables are devoted to the current students and one for the graduates. We will present below the results of the students’ responses analysis.

### **5.3.2. Analysis of Students’ Responses**

After analyzing and presenting the results of teachers’ responses, we will consider the following question: What are the results of the hierarchical axial coding? And, what are the results of the non-hierarchical selective coding? What patterns have emerged?

To begin with, it is worth pointing out that the questionnaires did not harvest responses for all the questions, because the students left a large number of questions unanswered either by skipping them (Cf. code: DAQ) or by responding with fewer details than required and expected. Therefore, the answers did not help in forming the full descriptive picture that we aimed at.

The section below deals with the results of the hierarchical axial coding and the non-hierarchical selective coding.

### **5.3.2.A. Results of The Hierarchical Axial Coding**

#### **a) Theme 1: Students' Background**

The background of the students is not homogeneous, because some of them come from English-speaking Arab countries (e.g. Egypt); whereas, others come from countries that are French-speaking (e.g. Algeria). Therefore, their mastery of the languages taught at the HAIT, their 'B' and 'C' languages, varies. Moreover, the follow-up study revealed that the students have an intermediate level, because they do not have a very good mastery of these foreign languages (Cf. sub-theme: Mastery of languages; code: intermediate level; the learner is weak in one of the required languages; STU needs remedial courses). This finding is similar to Carroll's (1978) finding who studied admission testing procedures at Georgetown University. He was struck by interpreting candidates' poor knowledge of their foreign languages too (as cited in Russo, 2011). It is clear that this kind of training requires more efforts than students might be able to manage until its end, because its aim is not to teach students the foreign languages or to consolidate them but mainly to develop their translation and interpreting-related skills. The question is whether the reinforcement courses in languages, which take place over two years (in S1, S2, S3 and S4) are enough for all the students. Further, their strategic, cultural and world knowledge is limited regarding, at least, one of the languages offered at the HAIT (i.e. English and French), because their mastery is lower than expected.

But, even their mastery of their 'A Language', i.e. Arabic, their native language, is questionable. Five first-year students only out of 13 said that they have an excellent level of Arabic. Furthermore, even the graduates said they do not have a very good mastery of their active and passive languages. GSTU5 for example said that his/her mastery of Arabic is just average (Cf. code: GSTU still needs remedial courses).

Besides, not all the students had background translation-related input. Some of the students did not have a BA in translation but in English and French Arts and Languages. This is of course the minimum requirement to embark on postgraduate studies in interpreting (Q4; Code: Minimum Requirement). However, one wonders

whether the suggested curriculum meets their needs to compensate for their weaknesses in this matter, especially that they will add on to their knowledge more advanced material of postgraduate level and will be assessed equally with other students who have already acquired some translation skills and knowledge. This is part of the issue of fairness which will be discussed when we interpret the results. In this regard, it is worthy of mention that one student only (STU13) came from Electronics, a different field than languages and translation.

### **b) Theme 2: Perceptions about the Purposes of Tests and Exams**

As mentioned earlier, the students did not answer all the questions of the questionnaire (Cf. code: DAQ) for a number of reasons, some of them are obvious. Firstly, the majority of the students did not take either the Admission Test or the Interpreter Aptitude Test (Cf. code: Admitted without being tested), if not both. This practice at the HAIT is different from other institutions of higher education. Secondly, despite their weak mastery of languages, the students who took the tests (Cf. Q6) found them either ‘easy’ or ‘very easy’. The first-year students in S1 justified the easiness of the Admission Test by the opportunity it provided to them to study, by its direct questions in addition to the easiness of the language used by the teachers. The Admission Test was also easy for the graduates. One of them said that its easiness was due to its questions which were about personal ambitions only (Cf. code: self-centered questions). Consequently, because of the easiness of these tests the students did not have more views to express.

It is worth mentioning that only two of the first-year students in S1 said the Admission Test is ‘A little difficult’ and just one of the first-year students in S2 found the IAT ‘difficult’, because s/he did not have any experience in interpreting. The other two students found the IAT ‘A little difficult’ for almost the same reason. STU14 said, “*We did not have a background*” and STU15 said that s/he was not ready. The graduates also found that the IAT was ‘A little difficult’ because of background reasons (Cf. code: I studied translation, I wasn’t prepared for interpreting). More examples about the relevance of background and preparedness will be seen later.

It is obvious that the students are not aware of the specific purposes of these tests, and had no idea about the tests before taking them (Cf. sub-theme: information about the Admission test). For them, these tests seem like compulsory hurdles but meaningless, because they could not ascribe a specific purpose to each of them. There is no information on the website nor in any brochure, which can provide what the students need to be well-informed and get prepared. We will discuss the issue of preparedness below.

Overall, even the S2, S3 and S4 exams were not very difficult for the majority of the students and the graduates, which is quite unexpected. The second-year students found the Simultaneous Interpreting exams in S3 and in S4 ‘A little difficult’ because of the intricate topics, the difficult vocabulary and the high rate of delivery. The graduates were divided into two groups, the first found the S3 exam easy and the second (GSTU1, GSTU5, GSTU7) found it a little difficult and challenging. In addition, just two of the graduates found S4 exam a little difficult. Yet, despite this slight increase in difficulty for some students, they expressed their familiarity with interpreting and a sense of improved command of language after training. So they could cope with the difficulties of this level which was geared up. Almost all of them found that the second-year exams were different from the first-year exam in SI in terms of topic, content, style, prompts’ length and speed (Cf. code: Yes, S3 and S4 exams were different from S2 exam in SI), but GSTU5 did not find any difference.

Apart from GSTU5, GSTU6, GSTU2, both graduates and second-year students thought that the exams reflected the S3 and S4 courses as they dealt with the same topics or texts (Q11; Q12). Notwithstanding, some of the graduates (GSTU4, GSTU7) stated that the exams used the same speeches seen in class (Cf. code: same texts; same text type). Nonetheless, most of the graduates and the current students saw no difference between S3 and S4 exams (Cf. thematic code: there is no difference). Some of them mentioned a certain difference in terms of duration, text types, rate and vocabulary.

### **c) Theme 3: Assessment Constructs and Tasks**

This theme is about the question: Are the students aware of the skills or sub-skills that are being assessed in each step of the assessment process? Their responses revealed that they are not aware of them, because they did not list the same skills mentioned by the teachers. For example, second-year students in S3 and S4 mentioned language skills, comprehension, and vocabulary for S3 and grasp, reformulation and speed in addition to brevity, vocabulary and comprehension for S4. Therefore, they do not seem to be aware of the translation and interpreting skills.

As far as the constructs mentioned by the graduates are concerned, they are as follows: For the Admission test, the graduates mentioned mastery of languages, speed, terminology, rendition of the gist and fluency. As regards the IAT, they cited calmness, comprehension, accuracy, fluency and reactivity. The constructs of the exams are: tone, fluency, accuracy, comprehension and completeness (in S2); fluidity, rapidity, fluency, accuracy, reactivity, comprehension, completeness, calmness, stress management (in S3); ability to interpret for 35mn, fluency, accuracy, speaking like the speaker in another language (in S4). Therefore, two codes emerge. Some constructs match the teachers' constructs, others do not. We will discuss this point in the next sections dealing with assessment criteria and scoring.

The students seemed confused with the questionnaire jargon. When asked about the test tasks, some of them (STU1, STU2, STU3, STU5 and STU13) tried to describe the content of the interview itself by saying that it includes comprehension, listening, speaking, vocabulary, grammar and culture components. They did not describe the tasks, but the activities of the interview targeting their knowledge and performance. In addition, the second group confirmed that the task in the IAT was a simultaneous interpretation of a political speech from B into A and from A into B. But, STU16 seemed confused and added consecutive interpreting and sight translation to the main task.

Three codes emerged concerning the tasks. Code #1: Learner confuses the different modes, #2: STU did not do an exam, and #3: STU did not give details. The

graduates mentioned Simultaneous interpreting and sight translation for S3 and S4 exams. However, their answers seemed random, because doing sight translation in a SI exam is rather odd. Besides, GSTU6 stated that the tasks consisted of interpreting official speeches, TV shows and documentaries. But, the question for which we did not find any answer is: How many prompts did these tasks involve?

#### **d) Theme 4: Assessment Criteria**

In order to analyze the students' responses regarding assessment criteria, three sub-themes were considered: Sharing of assessment criteria, duration of tests and length of prompts.

##### **1. Sub-theme: Sharing of Assessment Criteria**

Only two students out of 13 students in S1 were informed about the criteria for a good performance (Cf. sub-theme: Criteria for a good performance), the other students were completely unaware of the criteria concerning the Admission test itself (Cf. sub-theme: Sharing of assessment criteria; Q16, code: No, the TEAs did not share their assessment criteria; Q17, code: No St provided a list of criteria). In addition, the second group, which is composed of three students only in S2, stated that they were informed about the criteria for a good performance and mentioned fluency, rapidity and short memory. Nevertheless, two students gave a 'No' answer to the question regarding teachers informing students about assessment criteria before taking the IAT (Cf. code: No, I did not have an idea about the IAT). In addition, STU14 added perfect comprehension and vast culture. Also, the second-year students were not given any assessment criteria for taking the exams. Thus, they were seemingly meaningless to them. We will discuss this issue further below.

Finally, most of the graduates stated that there was no sharing of assessment criteria (Cf. thematic code: No, TEAs did not share the assessment criteria). GSTU3 stated that s/he did not feel that the teachers base their evaluations on specific criteria and wondered why they were not shared with the students or displayed to them. Thus, they were not meaningful to them although some of them were interested in having more knowledge about them (Cf. Q23).

## **2. Sub-theme: Duration of the Tests and Exams**

The duration of the Admission test was 15 mins for 6 students. Only one student out of 13 said his test lasted 30 mins. The other students in S1 did not answer this question. Similarly, the whole IAT lasted between 10 mins and 15 mins for each of the three students.

As regards the exams, the students' and graduates' answers were different. Nonetheless, if we add up the length of the two prompts in each exam, it will last between 10 and 45 mins including the instructions and directives (Cf. thematic code: duration). In the next sub-theme, we will examine the details of prompts' length.

## **3. Sub-Theme: Length of the Prompts**

In the IAT, without any prior negotiation, the first-year students in S2 were asked to interpret from A into B and from B into A two different prompts, the length of the first one being of 5 mins and of the second of 10 mins. The graduates and second-year students responses also show that they were exposed to prompts which span between 5 mins and 15 mins throughout the process of assessment. According to the responses, in each exam, they were allowed to interpret one or two prompts.

### **e) Theme 5: Scoring**

Some students were not satisfied with the way the results were communicated to them as they were neither scored nor ranked after passing the Admission Test (Cf. code: No score ; No rank ). Four students only were satisfied, because ranking did not mean much to them. Besides, the first-year students in (S2) did not have their average score for the S1 and did not have any score or rank after passing the IAT (Cf. sub-code: No score ; No rank ; No results displayed). As for the graduates, two of them only mentioned they were scored and ranked (Cf. GSTU2 and GSTU7) after passing the Admission Test and the IAT. But this information seems to be known by some of them only. GSTU7 stated that his/her rank is '1' in the Admission Test. In the IAT, GSTU2 stated that his/her score is '16/20' and his/her rank 'very honorable' ; GSTU7 said his/her score was '13/20' and his/her rank '1'.

However, only the exams were scored. In S2, the second-year students' scores in simultaneous interpreting were above average, most of them obtained between 13/20 and 15, 5/20. Their marks in S3 were between 12 and 16. Thus, we notice some progress in S4 where the marks were between 14 and 16. As regards the graduates, their marks in SI in S2 exams range from 13 to 16. Their marks in S3 and S4 range between 11 and 17.

Moreover, it seems that marks are the most important communication channel in this Institute, because the second-year students obtained some feedback on speed and use of terminology only. Additionally, the graduates had feedback on tone, voice and coherence. Some of them did not consider these remarks as feedback, may be because it was not very specific. GSTU3 stated that "*feedback use needful all the time*". However, the students did not receive enough feedback (Cf, Theme: Feedback).

### **5.3.2. B. Results of the Non-hierarchical Selective Coding**

The non-hierarchical selective coding led to identify four main themes: Preparedness, mismatch between teachers' and students' criteria, subjective experience and dissatisfaction.

#### **a) Theme 1: Preparedness**

Most first-year students in S1 were not informed about the characteristics of the Admission Test, because eight out of thirteen stated they had no idea about it before taking it (Cf. Information about the Admission Test), and most of them did not prepare it (6/13, 5 did not answer) (Cf. sub-theme: Preparedness for the Admission Test). Also, in S2, they did not prepare for the IAT (Cf. No, I did not prepare for the IAT) and did not have a specific training period before sitting for this test (Cf. code: No, we did not have any training). As for the usefulness of the courses in preparing the students for the exams, STU17, STU18 and STU19 (Second-year students) found them useful and helpful (Cf. Q31 ; code: training; helpful). GSTU3 said, "*S3 course: approximately no preparation; S4 course: just a bit*". This lack of preparation could also be linked to the students' perceptions about Language Assessment (LA) and IQA, as will be discussed in the next chapter.

## **b) Theme 2: Mismatch between Teachers and Students' Constructs and Criteria**

There seems to be a mismatch between teachers' and students' constructs and criteria. The students did not refer to the same skills, sub-skills and criteria mentioned by their teachers.

As far as the Admission Test is concerned, the first-year students in S1 mentioned only three broad criteria of good performance which are language acquisition, fluency and self-confidence. One student in S2 mentioned perfect comprehension and vast culture. The second-year students mentioned quick comprehension, interpreting the message and rich vocabulary in addition to topic-related knowledge, speed and fluency.

The graduates were asked about the difference they make between first-year criteria and second-year ones. The results are as follows: The 1<sup>st</sup>-year level criteria are speaking, simple syntax, word order, booth use, accuracy, comprehension, synthesis, gist rendition, load management and discipline. The 2<sup>nd</sup>-year level criteria are speaking, complex syntax, word order, diction, accuracy, conference interpreting, hesitation management, self-corrections, fastness, completeness, intensive reading and hard work.

However, the teachers highlighted a wider range, namely, criteria for aptitude, quality, performance, achievement and certification.

We will next look at the students' and graduates' distinct subjective experiences.

## **c) Theme 3: Students' Distinct Subjective Experience**

STU3 expressed interest in having enough information about the criteria (Cf. code: STU interested in the criteria). S/he said, *"Yes, these criteria could clarify for me the image of a good accepted student in the institute"*. The second group (students in S2) also expressed their interest in knowing the assessment criteria, because they make assessment meaningful to them (Cf. Q29 ; code: Yes, the criteria are meaningful).

It is worth mentioning that 6/13 students intended to study interpreting (STU1, STU3, STU5, STU6, STU10, STU11) and students suggested more translation-related tasks in the Admission test. STU13 suggested a written examination for the Admission Test. Besides, even though some first-year students (STU1; STU12) stated that speaking in Arabic was the most difficult part in the Admission Test, they passed it (Cf. Q19 ; code: None ; Arabic, Listening). The second group did not find the Admission Test difficult at all (Cf. Q30 ; code: The beginning, None). Moreover, GSTU3 gave an example of how an exam was postponed because of the difficulty it posed to some students (Cl. Thematic code: Tasks of S3 exam in SI). This reveals a kind of negotiation between the teacher(s) and the trainees.

#### **d) Theme 4: Unsatisfactoriness**

A number of codes revealed that the students seemed quite unsatisfied with the method or procedure of assessment. They did not think that it was meaningful (Cf. code: No idea; The same test is repeated; Not informed; No, I did not prepare, Unsatisfied). The students' responses revealed that the Admission Test was short and easy, so they were satisfied with its easiness, because it enabled them to enroll at the HAIT. However, STU11 was not satisfied (Cf. code: No, It was not sufficient).

Although STU18 and STU19 agreed that the exams were similar to the conditions in the real-life job due to the use of booths, STU17 stated that their exams did not reflect the real-life job characteristics, because *“in a real situation, SI is more difficult and more stressful”*. GSTU1 stated that an exam of 5 minutes has nothing to do with a whole day of consecutive/simultaneous interpreting. GSTU7 also stated that exams are not representative of real situations. S/he said, *“[It] is not like when you know that there are people who are listening and depend entirely on you in order to understand what the speaker is saying”*. This explains his/her reaction to the exam results. S/he said, *“Not really, sometimes we feel like some teachers give us more credit than we deserve”*.

Still, other students (Cf. GSTU2, GSTU3, GSTU4, GSTU5) were satisfied and believed that the training helped them do well their jobs in the translation business. It

is worth mentioning that three of the graduate respondents (Cf. GSTU1, GSTU2, GSTU4) are presently sworn translators and two (Cf. GSTU5, GSTU7, GSTU8) are freelancers. When asked about the directives and instructions in tests and exams (cf. Q25 in Table 15), first-year students in S1 mentioned the questions which were asked, but did not focus on the directives or the instructions. In S2, first-year students were asked to be calm, follow the news, read the newspapers, and do a lot of training.

### **Summary & Conclusion**

This chapter reports both the qualitative coding and analysis of the pilot study data on the teachers' views and perceptions and the coding and analysis of the follow-up study data concerning the students' views and perceptions. One section only dealt with the quantitative analysis of the students' responses, the other sections focused on the qualitative coding and analysis of the teachers' and students' responses. One cannot deny that descriptive statistics are crucial, especially in this unexplored area of research. It would certainly help us with some insights to understand the qualitative issues, but quantitative data analysis was kept to a minimum.

The main results emerging from the qualitative and the quantitative analyses are highlighted below:

- The HAIT teachers do not seem to be aware of how to construct and design effective tests and exams tied to assessment principles.
- The students have an intermediate level in languages (Cf, Sub-theme: Mastery of Languages)
- They do not have the same background.
- Many of them found the IAT difficult, because they did not have a background in interpreting.
- The students were not prepared to take the Admission Test and IAT, because they were not well informed about their contents, constructs and criteria.
- There is a mismatch between the perceptions of the teachers and those of the students regarding the constructs and criteria of assessment.
- The teachers do not share their assessment criteria with the students.
- The students do not negotiate the prompts' length with their teachers. These prompts do not correspond to real-life job characteristics.
- The assessment at the HAIT is not standards-based.
- There is a general feeling of dissatisfaction (Cf, Sample) regarding the available information, communication of the results and design of the curriculum.

# **Chapter Six:**

## **Interpretation and Discussion of the Results**

This sixth chapter reports the findings of the pilot study and those of the follow-up study concerning the perceptions, views and opinion of the teachers and the students, both the current students and the graduates of the HAIT. Based on the literature, these findings are the result of the interpretation, synthesis and discussion of the data presented in the previous chapter.

### **6.1. Interpretation of the Results:**

In this section, we attempt to find answers to our main research question and its sub-questions. It is worth reminding that we wondered whether the current assessment procedure of SI at the HAIT is based on valid, reliable, authentic and practical tests and exams. Four sub-questions have been explored:

1. Is the Interview an adequate and sufficient test for admission?
2. Is the Interpreter Aptitude Test (IAT) a valid, reliable and practical test to assess the trainees' aptitudes?
3. Do the exams target the competencies needed for future SI at the 2<sup>nd</sup>-year degree level?
4. Are the tests and exams built on a well-defined rationale?

To interpret the results of this study in light of the literature reviewed, this section focuses on the results' meanings, the possible inferences and explanations that we can make out of the results regarding the principles of educational assessment, and on the evidence that the tests and exams are well constructed and designed or not.

Underneath, we will try to show evidence from the analyzed data of the extent to which the HAIT tests and exams meet the principles of assessment described in chapter one, and the criteria, norms and standards of SIQA which were described in chapters two and three.

#### **6.1.1. Validity**

The most salient result of this study is that the tests and exams at the HAIT lack validity, because the inferences made might lead to conclude that they are neither appropriate nor meaningful or useful regarding their purpose as tests and exams. Lack of validity is apparent throughout the study. For instance, as far as the Admission Test

is concerned, the analysis of students' responses has revealed that the Admission test has no face validity, because the students found it short and very easy with similar and repetitive questions targeting personal information which the students could answer without preparation (Cf. code: STU finds the test easy). Nonetheless, the Interpreter Aptitude Test seems to have some face validity to the extent that the students found it a little difficult, especially those who had no background in interpreting (Cf. code: A little difficult).

As far as construct validity is concerned, a lot of scholars such as Messick (1993) and Sawyer (2004) see that it is the unifying force for the other aspects (face, content, criterion, etc). Sawyer (2004) argues that assessment constructs in interpreter education have not been precisely defined, so a description of assessment criteria is necessary in order to validate the measurement and determine its possible errors. The analysis of teachers' responses revealed that they did not define clear constructs reflecting well defined abilities. As shown in the diagrams (chapter five, section 5.1.2.C), we had to deduce these constructs from their responses. Besides, their constructs were not benchmarked. That is, these constructs did not correspond to different levels of study and were not tied to clear standards as explained by assessment experts such as Wiggins (1993). This implies that the teachers were not aware of the implications of construct validity in ensuring high-quality teaching and testing. Additionally, second-year students' responses revealed that teachers focused only on comprehension, vocabulary reformulation and speed. Thus, there is a mismatch between teachers' and students' constructs (Cf. code: Skills do not match the skills assessed by teachers).

As far as criterion-related validity, the results revealed that the pre-set objectives of teaching at the HAIT are different from the criteria for tests and exams. The objective of teaching is mainly to provide the students with basic training, but the objectives of the exams are to check whether the students' performance meets the professional level which corresponds to a high-level of expertise. In theory, the exam results should give the teachers an idea about what is missing in the student's profile.

For instance, the exams should answer the following question: Does the profile correspond to the level of expertise required in the market? It is true that teachers' holistic quality criteria can serve in the development of constructs, but they are not 'productive' in this stage due to the teachers' apparent lack of awareness of their importance in pedagogical assessment.

As for content-related validity, the results revealed that the teachers did not focus on one type of discourse. For instance, during the training, teachers seem to adopt an eclectic approach by dealing with discourse (i.e. topics) relevant to politics, technology, business, media, medicine, court or even daily life. Then, for the exam they would choose samples from one of these types. To ascertain content validity, we checked whether the content of the test or the exam is representative of the content of the course, the students' responses revealed that the content is chosen randomly. Selecting contents randomly will surely reduce the exam content validity, as the exams, whether short or long, will unlikely represent real-life content. But, this aspect is not enough. So, we agree with Hughes (1989) that achievement tests must be based on objectives rather on content in order to be truthful.

As far as concurrent validity is concerned, the results revealed that all the exams at the HAIT are formal and direct performance-based assessments. Therefore, we cannot say whether the teachers have tried to ensure concurrent validity for their own exams. To determine whether tests and exams have concurrent validity, we need more data to be able to compare the tests and exams of the HAIT with the tests and exams of other institutes and schools of interpreting which explored other ways to increase the concurrent validity of their tests and exams. Unfortunately, little information is available in the literature on concurrent validity of interpreting tests and exams.

As regards consequential validity, we do not have data about the results of the graduates of HAIT in other interpreting-related academic or professional tests and exams. Therefore, we cannot draw conclusions about the tests' and exams' impact.

The scope of our research was limited to exploring graduation achievement at the HAIT only.

### **6.1.2. Reliability**

The second interpretation which stems from the salient results is that the tests and exams may lack reliability. For example, as far as student-related reliability is concerned, the unreliability of the tests and exams is due to the lack of what scholars call ‘test-takers’ test-wiseness’, i.e. the test-takers’ strategies for efficient test-taking (Mousavi, 2009: 804, as cited in Brown & Abeywickrama, 2010: 28). This is the natural result of lack of preparedness which was mentioned by the students.

As for inter-rater reliability, there seems to be a lack in the consistency of scoring across raters. In other words, teachers’ current practice diminishes the consistency of scoring across raters. The results revealed that the teachers of the HAIT use different scores because of the different criteria they use (Mastery of languages, accuracy, fluency, consistency, impact of performance, time) (Cf. Chapter Five; code: Sub-theme: Assessment Criteria). So, there is a need to train them in order to establish a ‘uniform basis for scoring’ (Sawyer, 2004: 102). It is possible to discuss the issue of inter-rater reliability for the admission tests where there is a board of assessors (jury) but not exams, because each exam is rated by one teacher only. The results of the study regarding the tests confirm Wu’s (2010) findings that most of the tests taken by students lack inter-rater reliability, because they depend on the behaviours of jury examiners.

As for parallel forms reliability, only one version of the exams was administered to the students (i.e. interpreting simultaneously). Thus, there is no possibility to explore it further by comparison with other versions. However, one might confirm that the teachers did not consider performance at different levels of study, because only one form was used in the first and second-year exams. In this regard, it is worth noting that Snell-Hornby (1992, as cited in Sawyer, 2004: 28), who did not focus on interpreting per se, draws attention to the need for training in a professional context and advocates

the implementation of alternative forms of testing that reflect the realities of future professional life.

To conclude, one might argue that, by considering all these aspects, the tests and exams reliability at the HAIT is questionable. Furthermore, even the internal consistency reliability is debatable, because, as mentioned earlier, in all the exams, there is one task and one item only which is interpreting. As a result, tests' and exams' administration must be reviewed in order to improve their administration reliability, because the responses revealed that the conditions in which the test is administered (Brown & Abeywickrama, 2010: 28) did not reflect professional settings, as we will see below.

### **6.1.3. Scoring and Grading**

The scores at the HAIT are holistic, so their reliability is not based on edumetric measures. Hughes (1989: 86-110) believes that total score reliability should be the result of many factors: total score must be the sum of many scores, descriptions of criterial levels must be clearly recognizable, irrelevant features of performance should be ignored, each performance must be scored by more than one trained scorer. In addition, the results revealed that no criterion-referenced cut-scores are determined. This means that the students may be unaware of performance standards, of what they should know or do at a minimal level in the real job. It also means that passing and failing is meaningless in the absence of milestones against which performance along the continuum can be measured with clear descriptors such as 'Basic, Proficient and Advanced' (Horn et al., 2000; Zieky & Perie, 2006).

Surprisingly, the students' responses revealed that they get numerical scoring. It is worth pointing out that this type of scoring is connected with a norm-referenced assessment which *"measures how the performance of a particular candidate compares with that of (an) other candidate(s) whose score is taken as a norm."* (Hatim & Mason, 1997: 181-182). Indeed, Harmer (1983: 350) points out that a norm-referenced test refers to *"a test in which each test-taker's score is interpreted in relation to a*

*mean (average score), median (middle score), standard deviation (extent of variance in scores), and/ or percentile rank*” (Cf, code: Aptitude Assessment). Consequently, in this case, we are in front of an awkward assessment practice; it is both criteria-referenced and norm-referenced. It creates problems in the interpretation of scores, their fairness and usefulness in the learning and educational outcomes. We dealt with these points in chapter five when we presented the results of students’ responses and we will interpret them below to get more insights.

#### **6.1.4. Practicality:**

The HAIT has a very sophisticated lab, so recording is feasible and scoring by a rubric is possible. However, TEA3 said that they used the lab only when it was available. Clearly, it is not a problem of equipment scarcity. We wonder whether it is just a question of management. Is it a problem relating to the courses’ schedules especially that most of the teachers are visiting lecturers? We could not interview the managers about this problem. Nonetheless, even if these are the reasons, they are not acceptable for scoring by impression, because teachers’ work load must not lead them to choose the easy way and think that recording the students’ performance is not practical.

#### **6.1.5. Feedback, Washback**

We intended to get more insights about assessment practices at the HAIT regarding the type of feedback the students are used to get from the teachers after the exams in addition to their washback or backwash, fairness, usefulness and meaningfulness, but the students’ awareness of the specifications of the tests and exams led to little interaction and did not generate rich feedback. At a micro level of analysis, even though some students had, informally, some feedback, one might deduce from the body of the results that it was negative, because the objectives of the courses were not clear to the students and the tests and exams were not well constructed to fit a specific purpose. The tests and exams seemed as similar hurdles which turned to be the source of students’ anxiety and dissatisfaction. The general

feedback that some graduate students mentioned was in the form of appraisal and encouragement. Others mentioned shadowing, coherence, voice control and terminology (Cf. code: Yes, we did have some feedback). Therefore, it seems that little attention is given to the impact of feedback on students' performance.

#### **6.1.6. Fairness, Usefulness, Meaningfulness:**

The results revealed that the most important constructs in the literature are not meaningful to the students. For instance, no student mentioned that they worked on developing some competences or sub-competences which are seen by the professional interpreters as key to success in ensuring quality such as the 'intuitive competence' (Albl-Mikasa, 2014: 51-81) which consist itself of many sub-competences. None of these sub-competences was indicated to be a construct for an exam neither in the first year level nor in the second despite their relevance to the criteria reflecting the degree of expertise.

Furthermore, assessment criteria are not meaningful to the HAIT students, because the teachers did not share explicitly these criteria with them (Cf. code: No, TEAs did not share the assessment criteria). In addition, the second-year students, for instance, were not given any assessment criteria for taking the exams. Therefore, the students do not seem to be satisfied with tests' and exams' fairness (Cf. chapter five).

#### **6.1.7. Authenticity:**

The tests lack authenticity, because the results revealed that the prompts were not representative of real-life job realities neither in term of length nor of content. In addition, there are two other findings which mean that the tests and exams lack authenticity. First, the students did not share the same views, perceptions and opinions with their teachers. Second, the students did not have the same perception about to the exit profile.

If we apply Angelleli and Jacobson's (2009) findings mentioned in the literature on simultaneous interpreting, we would ask, for example, the following question: Is a test the task of which is to interpret a prompt of five-minutes long authentic? Obviously, it is not as it does not reflect a '*real-life*' task. What if the prompt is ten-minutes long but the students listen to it three times or more before recording the test performance which is going to be marked, is this procedure authentic? It is definitely not authentic, because in the real life the interpreter will not have time even to think, he has just to act. As a result, we conclude that this practice at the HAIT would not be appropriate to provide authentic assessment of postgraduate students.

#### **6.1.8. Specifications of Tests & Exams:**

We mentioned in the review of literature that the TB must include all the details such as the purpose, construct, criteria, questions, items, scoring scale, etc. We will focus below on tests' and exams' purposes, format, content, construct and criteria. Despite the absence of teacher-made TBs, we conclude that:

**i. Purpose:** There are different purposes for admission tests worldwide, because they are intended to test and measure capacity or general ability a priori to predict success in a specific undertaking (e.g., before taking a foreign language course or a translation course) (Brown & Abeywickrama, 2010). As far as the Interview is concerned, it is intended only for admission, but not for placement because the students are not allowed to enroll in the second year whatever their performance. Its purpose is not clear, because it is used for both would-be translators and would-be interpreters. Nevertheless, contrary to the Admission Test, The IAT has a clear purpose which is to select only would-be interpreters.

In exams, the teachers use direct testing and their tests are integrative. In fact, teachers use subjective testing, because they are called for making judgements on students' performance. So, their scoring is subjective (Hughes, 1989: 14-19). The purpose is also to rank students against one another because the students' responses revealed that they obtain numerical marks.

Nonetheless, there is no end-of-training achievement test. No test was designed or is used for this purpose.

**ii. Format and Content:** In order to compare the HAIT with other institutions, we have widened the scope of our review of literature by considering admission tests and exams from different continents. So, for aptitude assessment, as described in the review, the other institutions use multiple formats with varied contents contrary to the HAIT which uses just an Interview for admission. Therefore, there is a need to supplement it with some of these tests or tasks: Carrol (1978) (Oral Test), University of Surrey (memory exercises, written translation & general knowledge quiz), ESIT (Written Test & Oral Test), University of Ottawa (Shadowing, Cloze Test, Sight Translation, Memory Test).

As far as the IAT is concerned, the literature delved into many test formats such as shadowing (Lambert, 1991; Moser-Mercer, 1985; Schweda Nicholson, 1990), psychometric testing battery (Gerver et al., 1984, 1989), personality traits test (Schweda Nicholson, 2005), Short Memory, Sight Translation & Reading Test (Alexieva, 1993), Oral test (Aryona-tseng, 1994), paraphrasing (Russo & Pippa, 2004) and the SynCloze Test (Pöchhacker, 2002). These tests are recommended because of their high predictive validity. Additionally, Moser-Mercer (1978) suggested a course-based assessment which includes shadowing, memory exercise, paraphrasing and processing of numbers.

The students' responses revealed that only one format is used in SI exams, interpreting a prompt. The content is selected at random. It may or may not correspond to the content presented in class. This means that the students are not exposed to the range of formats which might enable them to use different sub-skills to develop their intuitive competence and expertise.

**iii. Constructs and Criteria of Quality:** The results of teachers' responses revealed that they do not rely on any of the models described in the review of literature, neither the academe-based ones, namely Pöchhacker's (2002) model,

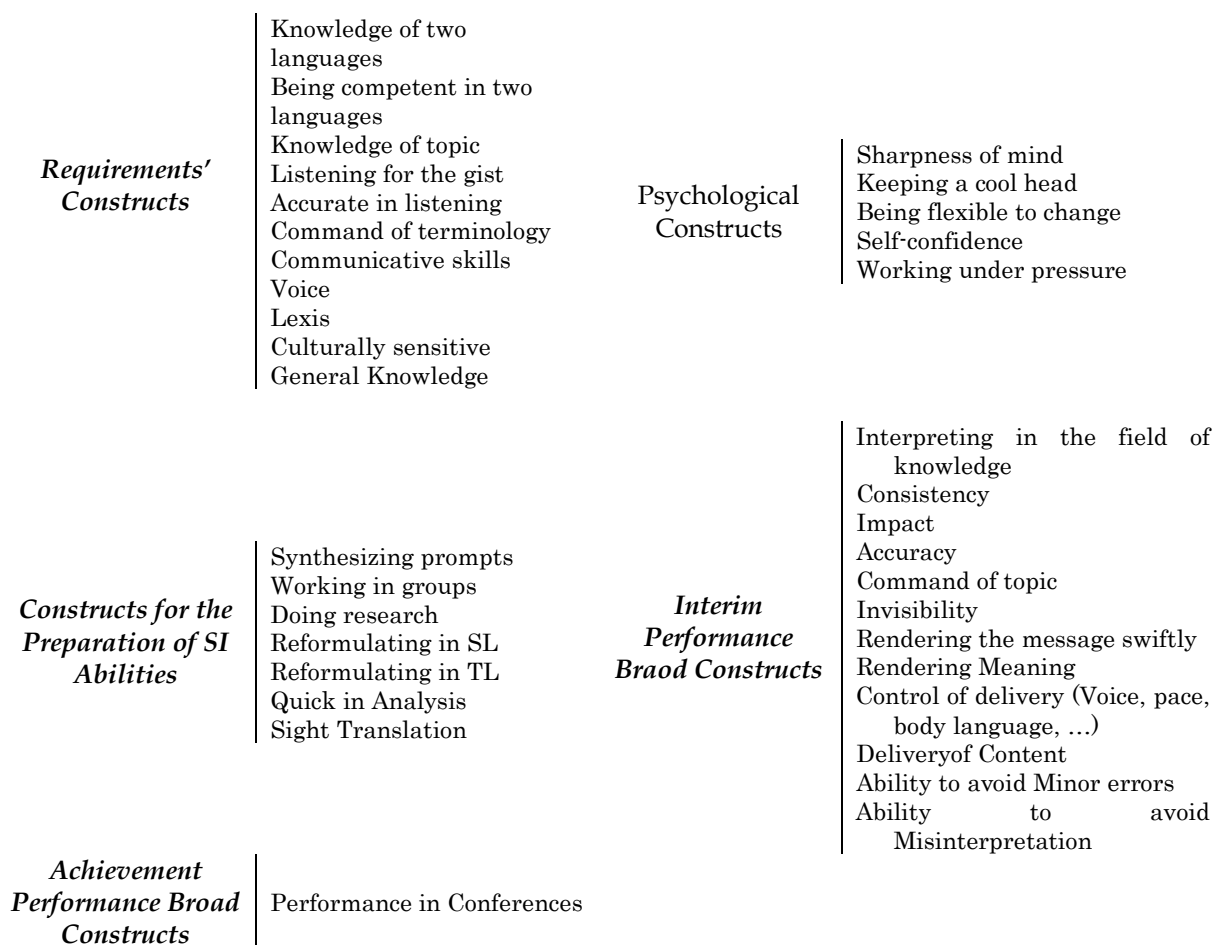
Moser-Mercer's (1978, 2008) model, Gile's (1985, 1997, 2002, 2009) model, Setton's (1998) model, Stenzl's (1983) model and Kutz's (2010) Leibzig Competence Model, nor the industry-based ones such as Tseng's (1992) model of the professionalization process, Ozolins's (2000) and Agger-Gupta's (2001) models of professionalisation in institutional contexts in addition to Kalina's (2002) scheme of interpreting quality dimensions and factors.

The teachers' responses revealed that they have a broad definition of interpreting quality (Cf. broad hypothetical perfect performance) and they use constructs and criteria which are not related to any model but to perfect products rather than products of educational stages of learning (Cf. code: process-oriented assessment; university graduation levels; professional mastery levels). Grbic's (2008) work on the different levels of attainment and achievement where quality is viewed as excellence, perfection and attainment helps in disagreeing with teachers' conception of quality. Quality must be broken down into components with benchmarked constructs and criteria.

The results enabled us also to take a clearer idea about the constructs and criteria that are taken into account by the teachers (chapter five). It is worth mentioning that the variables or criteria which are considered by the teachers to measure quality neglect a great deal of additional variables which are stressed in the literature (Chapter Three, educational and professional criteria and norms) to reach 'optimum quality' (Moser-Mercer, accessed 2015). For instance, Moser-Mercer points out that additional internal and external variables must be considered such as professional variables (team size, length of a turn, load of a working day, morning and evening work load, etc); employer's or agencies' variables or criteria (e.g. price, rhetorical skills, team discipline, schedule flexibility, etc). This means that the assessment procedure especially in the final phase undermines the real-life criteria which are considered as essential criteria for optimum quality in the professional setting. The dimensions and criteria suggested by Kalina (2002) are a case in point.

Diagram 5 below shows a hierarchical taxonomy of teachers' assessment constructs.

Diagram 5: Suggested Hierarchical Taxonomy of Teachers' Assessment Constructs



We notice that the above constructs target an overall global and integrative interpreting ability. They seem as general features for quality and they are more or less similar to the broad constructs and criteria suggested by Sawyer (2004: 266) in the criterion-based assessment rubric for proficiency testing which is below (Diagram 5). That is to say, the teachers are not aware of the need to differentiate between educational performance-based testing, which must be deconstructed and benchmarked with clear standards, and proficiency testing, which corresponds to a norm-referenced assessment. A criterion-based assessment must cover and represent specific course content and performance.

Diagram 6. A criterion-based assessment rubric for proficiency testing (Sawyer, 2004: 266)

Assessment constructs: The examinee should demonstrate his/her ability to:

- Interpret with faithfulness to the meaning and intent of the original
- Interpret in a manner linguistically appropriate to a given communicative situation
- Apply world knowledge and knowledge of subject matter
- Perform with resilience under stress and demonstrate acceptable platform skills

Assessment criteria: Criteria fall into three main categories:

- Meaning, e.g. accuracy, omissions, overall coherence
- Language use, e.g. grammar, expression, word choice, terminology, accent, and diction
- Presentation, e.g. pace, voice, non-verbal communication such as posture, eye contact, and appropriate gestures

Scoring criteria: High pass

Outstanding work: extreme accuracy in meaning, superior command of language, and highly polished presentation

Pass

Acceptable work: the interpretation is accurate; language use is appropriate; presentation is convincing; subtle shifts in nuance, minor slips in language use, slightly flawed presentation, or inconsequential combinations thereof possible

Fail

Unacceptable work: the interpretation does not render the original accurately and convincingly due to any of the following:

- Meaning has been altered, e.g., failure to convey information or distortion of information
- Language use is incorrect, e.g., faulty grammar, incorrect word choice or terminology, thick accent
- Presentation is poor and undermines credibility, e.g., long pauses, slow pace, lack of eye contact, inappropriate non-verbal communication, or extremely shaky or inaudible voice

In light of these categories, one could ask the question about whether these abilities are enough to build interpreting competence. The list of constructs provided by the teachers sounds very limited. Sawyer (2004: 99) highlights other constructs which were revealed by previous surveys. For example, the AIIC Survey on Expectations of Users of Conference Interpretation distinguishes between content match (completeness of rendition, terminological accuracy, faithfulness to meaning) and formal match (synchronicity, rhetorical skills, voice) (Moser, 1995). One could also cite other constructs which are relevant to these variables: Memory (Short, long), mechanisms (Hesitations, false starts, threads, etc), coping Strategies, screening (Story-telling, visualizing, intertextuality, etc), prosodic features (Intonation, stress, tone, pitch, accent, rhythm, melody, etc), and the like.

Nonetheless, all these constructs are not well defined. In other words, the formulation of constructs should be more expressive and they should be presented in the form of clusters which complement each other as skills and sub-skills so that the target profile takes a clear shape. For instance, the formulation should be as follows:

- Ability to render a complete media-story

- Ability to memorize long cued chunks of a media discourse.
- Ability to use media-relevant terminology
- Ability to render specific details in a media-story such as dates, places, names, numbers, etc.
- Ability to use appropriate prosodic feature for a given media-story

Or, as follows:

- Ability to interpret short/long speeches in a given topic (general/technical) from A into B according to the level of study and performance using the simultaneous mode of interpreting.
- Ability to understand and render simultaneously semi-technical Arabic from a legal context into English.
- Ability to understand and render technical Arabic from a legal context into English using the simultaneous mode of interpreting.
- Ability to interpret simultaneously a ten-to-twenty-minute discourse on global economy issues from Arabic into English.

These constructs can be complemented by other constructs and divided into two series which fit for the first or the second year of study with a specific series for the first semester and another for the second according to well-thought-over order of priority. It is not enough to have the pre-requisite skills and to focus on the development of the basic cognitive skills. Alb-Michaela (2014) points out that the professional interpreters develop more intuitive skills upon which they tap more than they rely on cognitive and rational skills and strategies. The results of her interviews show that the non-analytic and non-deliberate processes take prominence in professional interpreters. However, one cannot make explicit links between the different constructs suggested by the HAIT teachers. They do not seem to exist.

This issue is complicated even more if we consider the archiving tools of the trainees' performance.

#### **iv. Archiving tools:**

The results revealed that the teachers do not use the Lab and they do not use rubrics, neither a holistic rubric nor a standards-based rubric (Stiggins, 2001) (Cf. code: Assessment without a tool). Therefore, their practice means that their assessment

is subjective and might lead to unfairness, which was signaled by some students in their responses in a form of dissatisfaction from the procedure of assessment and the communication of their scores and grades.

## **6.2. Synthesis & Discussion of the Results**

The results reported above, that is to say the analysis and interpretation of teachers' and students' responses, give us some answers to our research questions. Below, we will discuss the implications for each test and exam along the process in light of the adumbrated points. The conclusions that we have reached about each stage and the emerging hypotheses are as follows:

### **6.2.1. Theme 1: The Admission test**

The first sub-research question (RQ1) is whether the interview is an adequate test. The research tool that we have used to answer RQ1 is Questionnaire 1. The HAIT teachers and students were expected to give answers to the following questions about the admission test which is an interview: What kind of interview is it? How is it conducted? Is it enough to test the aptitude of the would-be translators and interpreters? If the objective is to test just language proficiency, is an interview the appropriate form to test it? Is it fair? That is, do the interviewers keep record of the interview? Do they use 'checklists' or 'rubrics'? What is the construct of 'translation aptitude' for them? Does it correspond to research findings? Is this test similar to what is practiced elsewhere, that is in other universities of the world?

It is worth noting that as an Entrance Test, the Interview could be used for selection instead of being used only for admission. We wonder why it is not used as a placement test. In light of the results, the admission test (i.e. the interview) is intended to test construct(s) common to translation and interpreting skills or abilities. However, its current format does not seem to be appropriate to achieve this purpose. Basically, two or three teachers form the jury or panel and conduct the interview. They start by welcoming the interviewee then they ask him/her to introduce himself/herself. After

that, the jury members ask him about his linguistic, educational and, sometimes, cultural background and his objectives of studying at the HAIT. Some of them ask technical questions, either academic or professional, related to the field of translation. Most of the time, the interviews last up to half an hour. The interviewee is requested to answer either in Arabic, French or English according to the language used by the interviewer when asking him.

According to Sawyer (2004), the guidelines in curriculum design are educational philosophy, aims, goals, and objectives and the current curriculum models describing the relationship between instruction in translation and in interpretation are being reviewed. Inès (2009) comments that even those who are advocating the competence-based approach in the discipline of translation and interpreting have not agreed on the competences specific to them. The questions are: What are the skills and sub-skills that are common to translation and interpreting in the Interview? Are they really tested? The results revealed that SI aptitude is not assessed in the current Admission Test but at the end of S1, in the Interpreter Aptitude Test only. The students' first assessment experience with SI is through the IAT and the courses taught in S1 do not prepare them to take it.

The results revealed that this plain interview lacks predictive validity. In this regard, Timarová and Ungoed-Thomas (2009) argue that all the admission tests have very weak predictive validity because little is known about their efficiency. So, there is no difference between the well-known European admissions tests and the HAIT's interview as they have the same purpose the function of which has not been checked yet.

Nevertheless, regardless of its predictive validity, our hypothesis is that it is not an adequate test, because as the results demonstrate the 'Interview' does not discriminate between future translators and interpreters, especially that the previous studies have already shown the difference between the aptitudes of translators and those of interpreters (Szuki, 1988). Subjective ratings of oral interviews and writing

samples are prone to measurement errors as shown by the G-theory (Bachman, 1995); therefore, it must be supplemented by other edumetric tasks or tests.

The responses revealed that there are no tests to assess the level of proficiency in Arabic. We wonder whether the interview is enough to test it in the absence of standardized psychometric tests. Even the IAT that we will discuss below does not test it.

### **6.2.2. Theme 2: The Interpreter Aptitude Test**

When asking the second sub-question about whether the Interpreter Aptitude Test (IAT) is a valid, reliable and practical test to assess the trainees' aptitudes, we were interested in eliciting explicit answers to these questions: Is the IAT as a Screening/ Selection tool authentic? Is it a valid and reliable test? Does it test all the skills needed to reflect the aptitude of the would-be interpreters? What is the construct? Is the short period of each course enough to acquire a skill or even a sub-skill? What is its backwash effect?

Even though the IAT has some face validity, increasing the level of difficulty would improve validity and give significance to preparedness and meaningfulness for selection and training.

As for construct validity, the conclusion that one might draw from the responses is that the IAT does not test the trainees' abilities which were acquired neither during the training in S1 nor over a specific training period. This IAT does not achieve the purpose of selecting the promising interpreter. This is not an adequate test, because it is meant for professionals, not trainees. It should include discrete point tasks relevant to S2, S3 and S4 levels of performance.

We agree with Odendahl (2011: x) who points out that *“even the highest-quality testing elicits only samples of student performance, not complete information—and test scores are only ways of interpreting these samples, not absolute truths.”*

However, it is important to stress that the sample of interpreting performance must be well-thought over to ensure high predictive and concurrent validity.

In terms of scoring, the results revealed that the three teachers use a criterion-referenced measurement in the IAT, because they did not display lists which show the ranking of the students after the IAT, but they did it after the exams. Their practice in exams corresponds to a norm-referenced test but is still based on a criteria-referenced assessment.

We wonder why some students only were scored and ranked, and knew about their scoring and ranking.

### **6.2.3. Theme 3: Exams**

The third sub-question is: Do the exams target the competencies needed for future SI at the 2<sup>nd</sup>-year degree level?

The results of the analysis revealed that all the exams – whatever the interpreting courses - are in the form of integrative tests. These tests are different from discrete point tests which in language assessment refer to “*assessments designed on the assumption that language can be broken down into its component parts and that those parts can be tested successfully*” (Brown & Abeywickrama, 2010: 348). Sawyer (2004: 29-31) reports studies stressing the difference between expert-level and novice-level of expertise (e.g. Moser-Mercer, 2008, video accessed 2015; Liu, 2001, etc). These studies (e.g. Liu 2001, Becerra et al. 2013, etc) have shown that there are qualitative and quantitative differences between expert and novice performance in interpreting. The results revealed that the criteria used by the teachers do not reflect a clear understanding of the stages in the development of students’ expertise while acquiring interpreting skills as advocated by educational assessment scholars such as Wiggins (1996), Guskey (2006), Stiggins (2007), Jung and Guskey (2007), Jung (2009) or interpreting scholars such Moser-Mercer (2008, video accessed 2015).

In real-life professional contexts (e.g., in conferences), professional interpreters may be asked to interpret all day with breaks. Each performance is continuous for '30 or 45 mins'. Therefore, teachers' responses did not reflect what they really did in the classroom or in the laboratory, even though they targeted a given level of proficiency. We may conclude that their answers referred to hypothetical situations and not to the reality of their practices. Besides, teachers did not take into account the students' level of study and performance, course content, target skill, rate of delivery of the SL, and authentic 'genuine' material. For example, some students were unable to express themselves in English accurately (Cf. code: STU needs remedial courses; STU is weak in one of the languages).

It is worthy of emphasis that teachers' quality assessment criteria should go hand in hand with the assessment constructs. As mentioned in chapter two, Kalina (2002) discussed the dimensions of quality and the factors which might affect it, but these quality criteria remain vague and need more clarity through the use of clear descriptors.

As for the 1<sup>st</sup>-Year Exam (S2 Exam in SI) and the 2<sup>nd</sup>-Year Exams (S3 and S4 Exams in SI), one might conclude that:

- The exams did not target the competencies needed for future SI at the 2nd year degree level.
- The skills and sub-skills were not well identified.
- The students did not regulate their learning through self-assessment.
- The students did not graduate with a common profile.

We could not collect data about the students' real performance, because we could not get any recordings. We intended to record them thrice in order to transcribe the recordings and analyze them. But, we were not authorized.

#### **6.2.4. Theme 4: The Whole Procedure of SI Assessment**

We will attempt in this section to answer the fourth sub-question: Are the tests and exams built on a well-defined rationale? Thus, what are the implications regarding assessment principles and process?

As far as the construction of tests and exams is concerned, the responses revealed that no model related to the academe or the industry was applied (see Pöchhacker's (2004) Model, Moser-Moser's (2008) Model, Gile's (2009) Model, Setton's (1998) Model, Stenzl's (1983) Model, Kutz's (2010) Model, Tseung's (1992) Model, Kalina's (2005) Scheme in the literature review).

Moser-Mercer (2008) points out that developing expertise in any skill is crucial, in particular in SI which requires high-level cognitive skills. The development of expertise focuses in particular on progression, regression, anticipation of errors and the role of practice. She depicts the stages through which a learner must go in order to learn a skill or a sub-skill and pinpoints that progress can be achieved through three stages: the cognitive stage (text analysis, paraphrasing, etc), the associative stage (learners learn from errors) and the autonomous stage (learners learn to use fewer resources).

Teachers' responses did not reflect any difference between novices and experts concerning types of knowledge (i.e. factual, semantic, schematic and strategic) and performance strategies (comprehension, planning, monitoring, work load management) which were described by Moser-Mercer (2008). Additionally, students' responses revealed that they are unaware of the various cognitive skills and sub-skills required for an expert-like performance (e.g. emphasis on input/output; on the known or the unknown, segmentation of the message). For Moser-Mercer (op. cit), the last stage is that of autonomy/automation which requires practice (5000 hours to become an expert) through many drills to be familiar with conference terminology and phraseology.

As for assessment for learning, the results revealed that the current assessment tools are used just for summative purposes (assessment of learning). Angelo and Cross (1993) state that teaching effectiveness cannot occur in the absence of learning, because teaching will be just like talking in vain. Similarly, students' effectiveness, self-assessment and self-regulation cannot be reached, if teachers are not trained in

classroom assessment, because the central purpose of classroom assessment is to empower both teachers and students towards qualitative learning in the classroom.

As far as achievement testing is concerned, the results revealed that the exams cannot be considered as achievement tests, because they are not bound by the courses (not course-related) and there is no end-of-training test at the HAIT. Consequently, one might say that the procedure of SIQA requires more reflection on the inclusion of well-designed achievement tests to improve both learning and assessment. As early as 1951, Lindquist described an educational achievement test as “*a device or procedure for assigning numerals (measures) to the individuals in a given group indicative of the various degrees to which an educational objective or set of objectives has been realized by those individuals.*” (Lindquist, 1951: 9). Therefore, the list of educational objectives at the HAIT must include statements about the assessment of achievement at the end of the training and a final end-of-training test must be designed for this purpose.

#### **6.2.5. Theme 5: Emerging hypotheses**

What hypotheses can be formulated from this exploratory qualitative study? As far as the emerging patterns are concerned, a quantitative study on the correlation of variables is needed as well as more qualitative ones regarding the relationship between performances in the different exams. For example, relationships must be established between performance in the IAT and exams, task-based assessment and performance assessment, performance in exams and the accreditation exams.

Through this study, we have been able to generate the following hypotheses:

- H 1: Training teachers in educational assessment can improve students' attainment and achievement.
- H 2: The interview alone is not an appropriate test for admission to the HAIT.
- H 3: The IAT does not address the trainees' abilities which were acquired during the first semester of the training.
- H 4: The exams of SI do not target well-benchmarked competencies for the two years of the study.

- H 5: Assessing students' achievement with a final test can increase students' performance in international professional certification tests.

We concluded that the data we obtained is not enough. Unfortunately, all the teachers refused to take part in the follow-up study, which reflected students' perspective only, and claimed that they do not have enough time for more detailed studies.

### **6.3. Limitations of the study**

The research we conducted at the HAIT was quite informative, but it was also inconclusive owing to a number of factors that are described in this section.

#### **6.3.1. Lack of Data**

This study is exploratory and not intended to investigate fully one issue or aspect of assessment at the HAIT. Its aim is to give a bird's eye overview of all the issues involved in the process of assessment. Initially, it aimed to compare teachers' and students' perceptions in relation to classroom practices by analyzing the tests and exams, but, due to unavailability of the data (actual tests and exams administered at the HAIT) we could not have full grasp of: 1) Tests' and Exams' Specifications (Purpose, format & content, etc), 2) Tests and Exams' Constructs, 3) Tests' and Exams' Construction & Design (Tasks, prompts, etc), 4) Scoring matters for Summative & Formative Tests & Exams, 4) Discussion of Traditional & Alternative Forms of Design, 5) Competence, Performance & Expertise issues. So, we limited our study to analyzing perceptions inferred from the participants' responses.

It was difficult, and sometimes impossible, to have access to the detailed description of the specifications and content of the different tests and exams, i.e. the Admission Test, the Interpreter Aptitude Test, the Exam N°01 (S1), the Exam N°02 (S3), Exam N°03 (S4). Therefore, an in-depth analysis regarding their validity, reliability, authenticity, practicality and washback was almost impossible to do. Additionally, there is no information on how teachers' feedback affected the trainee's learning and progress.

- Generally, most of the students were reluctant to take part in the study or to express themselves. After discussing with them, they agreed to collaborate. But, they did not answer some questions, maybe because they did not understand them and were embarrassed to ask for explanations.
- They did not give many details, maybe because they could not remember them in their long memory and did not keep assessment diaries either.
- We were not invited or allowed to attend the Admission Test and the IAT as an external researcher or observer.
- We were not authorized to observe classes during the exams.

In addition, there is dearth of data about the curriculum and the procedures for its implementation. It is not clear what the teachers do in classes and how they integrate assessment as a tool for formative learning.

### **6.3.2. Time Constraints for a Qualitative Study**

Both teachers and students took a long time to respond to our emails and return their completed questionnaires.

For time and space constraints, the scope of the research was more restrained than expected in considering all the emerging patterns. No more follow-up studies were carried out as advocated by qualitative researchers. This study is one shot study, because there was no opportunity to include a test-retest component in the questionnaire design to elicit more responses in relation to the following themes:

- Teachers' abilities: The effectiveness of educational assessment depends greatly on teachers' knowledge and skills to implement it in the classroom, because they must be able to deal with changing instructional objectives, plans and practices. Their abilities must be reflected in clear domain-related objectives which correspond to students' needs and are assessed with the appropriate methods and equipment. Hence, the teachers must master the required technology (Genesee & Upshur, 2001; Brookhart, 2004; Alkhusari et al., 2012). It depends also on their language proficiency (Genesee & Upshur, 2001).
- Teachers' Use of Authentic Recordings & Authentic Checklists.

It was difficult to conduct an additional follow-up study to investigate the results collected on the ground and therefore deepen our understanding of the issues from a qualitative perspective.

## **Summary & Conclusion**

This chapter reports the findings of the pilot study and of the follow-up study concerning the perceptions, views and opinion of the teachers, current students and graduates of the HAIT. Based on the relevant literature, these findings are the result of the interpretation, synthesis and discussion of the data presented in chapter five. First, we dealt with the interpretation of the results. Then, we tried to provide a succinct synthesis and discussion of the results regarding the admission test, the IAT, the SI exams and the procedure as a whole. Finally, we explained the limitations of this Study.

In a nutshell, by projecting the results of the survey on the principles, one might say that the tests and exams are not well constructed or designed, because the teachers did not elaborate tables of specifications for their tests and exams. Second, the teachers did not seem to be aware of the implications of their assessment method on the validity, reliability, authenticity, practicality in addition to the washback, feedback, usefulness, fairness and meaningfulness of the tests and exams. In terms of principles of assessment, the results revealed that the current practices of construction and design of the tests and exams need reviewing.

We found that the IAT had a weak predictive validity, because it did not test the abilities required for interpreting activities. It tests SI only and cannot provide information to predict the academic performance of the students in the different modes of interpreting. As for the exams, they are not interrelated in terms of constructs, criteria and well-benchmarked standards, so we need further research to explore their predictive validity.

Moreover, there is no achievement test at the HAIT, so neither the teachers nor the students have a clear idea about the student interpreter exit profile. We recommend the construction and design of a final graduation test similar to the available professional certification tests based on clearly defined constructs, criteria and standards.

## General Conclusion & Recommendations

This study focused primarily on the procedure of Simultaneous Interpreting Quality Assessment (SIQA) at the Higher Arab Institute of Translation (HAIT) and raised a number of pertinent questions in order to generate hypotheses. The research findings have implications for teaching, learning and assessment of Simultaneous Interpreting. Hence, we have tried to answer the main research question we posed and its sub-questions. The research question is: Is the current assessment procedure of SI at the HAIT based on valid, reliable, authentic and practical tests and exams? Four sub-questions have been explored:

1. Is the Interview an adequate and sufficient test for admission?
2. Is the Interpreter Aptitude Test (IAT) a valid, reliable and practical test to assess the trainees' aptitudes?
3. Do the exams target the competencies needed for future SI at the 2<sup>nd</sup>-year degree level?
4. Are the tests and exams built on a well-defined rationale?

The review of literature was divided into three chapters. The first chapter dealt mainly with the principles of assessment that would ascertain valid, reliable, practical and authentic assessments of the real attainment and achievement of the learners. Both traditional tests and alternative forms of educational assessment aim to achieve high quality results. Thus, the construction and design of teacher-made tests (e.g. course exams) must follow a long validation procedure which touches upon the following: 1) Purpose; 2) Objectives; 3) Specifications; 4) Content and format; 5) Construct; 6) Criteria; 7) Scoring and grading; 8) the long-standing issues in assessment which are related to accountability, comparability of the results, fairness, ethics, and equity in addition to educational outcomes.

In the second chapter, we critically reviewed the literature on IQA to gain more insights and advance research in this area. We focused particularly on SIQA models. We divided them into academe-based models and industry-based models. In the third chapter, we dealt with the criteria, norms and standards of SIQA. We tried to define SIQA but it seems that it is vague. In the second section, we tackled the educational criteria for SIQA of trainee interpreters (i.e. aptitude, ability and achievement criteria).

In the last section, we delved into the industry-based criteria, norms and standards of SIQA; namely self and peers' criteria; certification and accreditation criteria, norms and standards; employers' and clients' criteria, norms and standards; end-users criteria.

As far as methodology, design and procedure are concerned, after presenting the HAIT and describing the current assessment procedure according to the curriculum, we explained and justified our choices by explaining the need for a post-positivist paradigm and mixed-methods approach to answer the research question. This study aimed at collecting quantitative and qualitative data to answer the RQs. As regards the methods and instruments, three methods were used for this purpose, which are Case Study, Survey and Grounded Theory. Furthermore, regarding our design of the Pilot Study and the Follow-Up Study, after dealing with the sampling issue, five questionnaires with close and open-ended questions were designed and administered to the teachers, current students and graduates of the HAIT.

We dealt with the qualitative coding and analysis of the data collected through our pilot study regarding the teachers' views and perceptions in addition to the coding and analysis of the follow-up study data concerning the students' views and perceptions. One section only dealt with the quantitative analysis of the data concerning students, the other sections focused on the qualitative coding and analysis of the teachers' and students' responses. One cannot deny that descriptive statistics is crucial, especially in this unexplored area of research. It helped us in giving us some insights to understand the qualitative issues.

The findings of the pilot study and those of the follow-up study concerning the perceptions, views and opinion of teachers, students and graduates of the HAIT are highlighted below:

- Teachers do not seem to be aware of how to construct and design effective tests and exams tied to the principles of educational assessment (validity, reliability, authenticity, practicality, washback, feedback, usefulness, fairness and meaningfulness). Specifically, they do not elaborate tables of specifications (purpose and objectives; constructs, criteria and standards; content and format; scoring and grading methods; etc) for their tests and exams.

- Teachers adopt a broad assessment model(s) of interpreting at the HAIT. They do not rely on the available models described in the review of literature. Therefore, teachers' constructs, criteria and standards are not benchmarked.
- They do not share their assessment criteria with the students.
- They do not negotiate the prompts' length with their students. Besides, these prompts do not correspond to real-life job characteristics.
- Students at the HAIT have a medium level (i.e. Intermediate or Upper-Intermediate) in languages. So, this raises many questions on their real reactions to the instructional syllabus which is supposed to target an advanced level benchmarked with professional-level standards.
- Students have different backgrounds (in Translation and Languages) and those who do not have a background in interpreting find the IAT more difficult.
- Students were not prepared to take the tests for admission and for interpreting aptitude, because they were not well informed about their contents.
- Tests' easiness has reduced their face validity.
- There is a mismatch between the perceptions of the teachers and those of the students regarding the constructs and criteria of assessment.
- The assessment at the HAIT is not standards-based, as it should be.
- There is a general mood of dissatisfaction regarding the available information, the communication of the assessment results and the design of the curriculum and the instructional syllabus.

As far as the implications of this research are concerned, it is even essential to improve curriculum design, because it provides useful information for comparison with other programmes. Moreover, this study gives evidence to the managers of the HAIT, and probably of the other faculties of Interpreting, that the whole procedure of assessment must be reviewed by considering the necessity of designing authentic, valid and reliable simultaneous interpreting tests and exams with positive backwash. Therefore, the recommendations of this study are:

- Teachers' awareness must be raised by training them in assessment. Reading the available literature would not be enough for the assessor to achieve his/her goals without bias unless he/she enhances his/her skills in practice taking into account the specificities of the field of interpreting in which the forms of assessment are implemented.

The research revealed the importance of training the teachers in assessment matters and it would raise their awareness that testing language mastery is different from testing interpreting competence. Our hypothesis is that training teachers in educational assessment will improve students' attainment and achievement performance.

- The assessment procedure must be reviewed by considering the following points:

Regarding the admission test, teachers must test not only student language proficiency in the three languages but also their requisite translation and interpreting skills, because communicative competence is necessary but not sufficient. The testers must check whether the test taker has the aptitude to interpret. Hence, our hypothesis is that the interview alone is not an appropriate test for admission and it must be supplemented with other forms of assessment.

The IAT does not have content validity and it does not reveal whether the student is apt to interpret. The IAT has a weak predictive validity, because it does not test the abilities required in the continuum of interpreting activities. It tests SI only. It cannot provide information to predict the academic performance of the students in the different modes of interpreting. Then, teachers must consider the trail training period length for skills acquisition before setting the students to the IAT. Our hypothesis is that this IAT does not test the trainees' abilities which were acquired during the first semester of the training.

As for the exams, they are not interrelated and do not reflect a well-thought over benchmarking of the constructs, criteria and standards. So, we need further research to explore their predictive validity. Our hypothesis is that the exams do not target well-benchmarked competencies for the two years of study. Interpreting competence must be broken down into specific competences that can be assessed in the early stages of the training and at the end of it with different tasks. Ultimately, the task should look like a real-life one and reflect all the high-level constructs, criteria and standards related to the realities of the job.

What test takers do during a professional certification test must be an accurate depiction of what they will do in their professional lives. Thus, in order to get ready for such tests, the trainee interpreter must complete successfully a training programme which ends with achievement tests reflecting the professional constructs, criteria and standards. Our last hypothesis is that assessing achievement performance by a test will increase student' performance in international professional certification tests.

## Bibliography:

### Primary Sources:

- Ackermann, D.; Hannelore, L.; Redmond, M. (1997).  
Between Three Stools: Performance Assessment in Interpreter Training. In E. Fleischmann; W. Kutz and P. Schmitt (eds). *Translationsdidaktik. Beiträge der VI. Internationalen Konferenz zu Grundfragen der Übersetzungswissenschaft*, (pp. 262-267). Leipzig 11-13.9.1996. Tübingen: Narr.
- Albl-Mikasa, M. (2012).  
The importance of being not too earnest: a process- and experience-based model of interpreter competence. In B., Ahrens; M., Albl-Mikasa; C. Sasse (Eds), *Dolmetschqualität in Praxis, Lehre und Forschung. Festschrift für Sylvia Kalina* (pp. 59-92). Tübingen: Narr.
- Albl-Mikasa, M. (2014).  
Receptivism. An intertraditional approach to intuition in interpreter and translator competence. In L. Zybatow and M. Ustaszewski (Eds.), *Bausteine translatorischer Kompetenz oder Was macht Übersetzer und Dolmetscher zu Profis* (pp. 51-81). Innsbrucker Ringvorlesungen zur Translationswissenschaft VII. Forum Translationswissenschaft - Band 18. Frankfurt am Main: Peter Lang
- Alexieva, B. (1993).  
Aptitude Tests and Intertextuality in Simultaneous Interpreting. *The Interpreters' Newsletter*, 5: 8-12.
- ALTA Report (2010).  
Judicial Council of California/Administrative Office of the Courts. [Report]. CA: Administrative Office of the Courts.
- Altman, J. (1990).  
What helps effective communication: Some interpreters' views. *The Interpreters' Letter*, 3: 23-32
- American Translators Association (2009).  
ATA Certification Program: Eligibility Requirements – List of Approved Translation and Interpreting Schools. [Report]. Retrieved from: [http://www.atanet.org/certification/eligibility\\_approved.php](http://www.atanet.org/certification/eligibility_approved.php) (accessed May 15, 2009).
- Angelelli, C. (2007).  
Assessing medical interpreters: The Language and Interpreting Testing Project. *The Translator* 13 (1): 63–82.
- Angelelli, C. V. and Jacobson, H. E. (Eds) (2009).  
*Testing and Assessment in Translation and Interpreting Studies: A call for dialogue between research and practice*. USA & The Netherlands: John Benjamins B. V.
- Angelilli, C. (accessed in 2012).  
Testing and assessment in Translation and Interpreting Studies. [E-Paper]. Retrieved from [http://ANG-TES\\_000.doc/google.doc](http://ANG-TES_000.doc/google.doc)
- Archbald, D. A.; Newmann, F. M. (1988).  
Beyond Standardized Testing: Assessing Authentic Academic Achievement in the Secondary School. [Report – Evaluative/Feasibility (142)]. National Association of Secondary School Principals. Reston, Va; National Center on Effective Secondary Schools, Madison, WI.

- Ayers, L. P. (1918).  
History and present status of educational measurements. In G. M. Whipple (Ed.), *The seventeenth yearbook of the National Society for the Study of Education: The measurement of educational products* (Part 2, pp. 9–15). Bloomington, IL: Public School Publishing Company. [E-Book]  
[http://www.archive.org/stream/measurementofedu00whiprich.](http://www.archive.org/stream/measurementofedu00whiprich)] (accessed 06/11/2015)
- Banta T. W.; and Associates (2002).  
*Building a scholarship of assessment*. San Francisco: Jossey-Bass, John Wiley & Sons.
- Barik, H.C. (1971).  
A description of various types of omissions, additions and errors of translation encountered in simultaneous interpretation. *M.E.T.A.*, 16(4): 199-210.
- Barranco-Droege, R.; Collados Aís, Á.; Fernández Sánchez, M. M.; Becerra, O. G.; Fernández, E. I.; Pradas Macías, E. M.; Stévaux, E. (2012).  
Quality in interpreting. In *Encyclopedia of Applied Linguistics*. Uk: Blackwell Publishing Ltd (accessed online 2015).
- Barranco-Droege, R.; Becerra, O. G.; Pradas Macías, E. M. (Eds) (2013).  
Quality in Interpreting : Widening the Scope. Vol. 2. (Series: Interlingua). Granada: Editorial Comares.
- Becerra, O. G.; Pradas Macías, E. M.; Barranco-Droege, R. (Eds) (2013).  
Quality in Interpreting : Widening the Scope. Vol. 1. (Series: Interlingua). Granada: Editorial Comares.
- Black, P. J. (1998).  
*Testing: Friend or Foe? The theory and practice of assessment and testing*. London/ Philadelphia: Falmer Press, Taylor & Francis Group.
- Black, P.; Harrison, C.; Lee, C.; Marshall, B.; Wiliam, D. (2005).  
*Assessment for learning: Putting it into practice*. 6<sup>th</sup> Ed. England: Open University Press.
- Blaikie, N. (1993).  
*Approaches to Social Inquiry*. Cambridge: Polity Press; Oxford: Blackwell Publishers.
- Brennan, R. L. (2001).  
An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, 38(4): 295–317.
- Brennan, R. L. (accessed 2015).  
Perspectives on the Evolution and Future of Educational Measurement. [PPT Slides]. Center for Advanced Studies in Measurement and Assessment (CASMA). University of Iowa. Retrieved from: <http://cms.education.gov.il/NR/rdonlyres/473547C9-D26E-4880-8462-B6848829D1B8/57729/Brennan.pdf>
- Broadfoot. P.; Black, P. (2004).  
Redefining assessment? The first ten year of assessment in education. In *Assessment in Education*, 11(1). March 2004: 7-27. Taylor & Francis Ltd.  
DOI: 10.1080/0969594042000208976
- Brookhart, S. M. (2004).  
Assessment theory for college classrooms. *New Directions For Teaching and Learning*, no. 100. Wiley Periodicals Inc. (2005). (accessed 12/06/2015). Retrieved from: [http://www.gwu.edu/~fellows/GTAP/Online%20Makeup/T-L%20Presentation%20Readings/Assessment%20Theory\\_2004-Winter\\_p5.pdf](http://www.gwu.edu/~fellows/GTAP/Online%20Makeup/T-L%20Presentation%20Readings/Assessment%20Theory_2004-Winter_p5.pdf)
- Brown, H. D.; Abeywickrama, P. (2010),  
*Language Assessment: Principles and Classroom Practices*. 2<sup>nd</sup> Ed. NY: Pearson Education, Inc.

- Bryant, L. (2012).  
Structured questionnaires [*E-Book*]. Retrieved from  
[http://www.historylearningsite.co.uk/structured\\_questionnaires.htm](http://www.historylearningsite.co.uk/structured_questionnaires.htm) (accessed  
12/01/2013)
- Büler, H. (1986).  
Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of  
conference interpretation and interpreters. *Multilingua*, 5(4): 231-5.
- Campbell, S.; Hale, S. (2003).  
Translation and Interpreting Assessment in the Context of Educational Measurement. In  
G. Anderman; M. Rogers (Eds), *Translation Today: Trends and Perspectives*.  
Clevedon, UK: Multilingual Matters Ltd.
- Centre for the Enhancement of Learning & Teaching (1998).  
Postgraduate Certificate in Tertiary Level Teaching. [*Module 1*]. The Robert Gordon  
University. Retrieved from:  
<http://www2.rgu.ac.uk/celt/pgcertlt/assessing/assess5.htm>. (accessed, 18/12/2015)
- Chan, A. (2011).  
Effectiveness of translator certification as a signaling device: views from the translator  
recruiters. In R. Sela-Sheff; M. Shlesinger (Eds), *Identity and Status in the  
Translational Profession*. Amsterdam & Philadelphia: John Benjamins
- Chapelle, C. A. (2012).  
*The Encyclopedia of Applied Linguistics* [Online]. Wiley-Blackwell,  
<http://onlinelibrary.wiley.com/book/10.1002/9781405198431> (accessed 20/04/2014)
- Chiaro, D.; Nocella, G. (2004).  
Interpreters' Perception of Linguistic and Non-linguistic Factors Affecting Quality: A  
Survey through the World Wide Web. *Meta* 49 (2): 278–293. URI:  
<http://id.erudit.org/iderudit/009351ar>. DOI: 10.7202/009351ar.
- Clifford, A. (2005).  
Putting the exam to the test Psychometric validation and interpreter certification.  
*Interpreting*, 7(1): 97–131. Amstrdam//Philadelphia: John Benjamins,
- Collados Ais, A. ; Gile, D. (2002).  
La qualité de l'interprétation de conférence : une synthèse des travaux empiriques. In  
C., ShiaoHong (Eds), *Recent Research into Interpreting: New Methods, Concepts and  
Trends (In Chinese)*. (pp. 312-326). (accessed in 2013).
- Craft, A. (Eds) (2003).  
*Primary Education: Assessing and Planning Learning*. 2<sup>nd</sup> Ed. UK, USA & Canada:  
Routledge in association with the Open University.
- Creswell, J. W. (1998).  
*Qualitative Inquiry and Research Design: Choosing among five traditions*. London:  
Sage Publications Series.
- Creswell, J. W. (2008).  
*Educational Research: Planning, Conducting, and Evaluating Quantitative and  
Qualitative Research*. USA: Pearson/Merrill Prentice Hall.
- Davies, A. (2013).  
Fifty years of language assessment. In A. J. Kunnan (Ed). *The Companion to Language  
Assessment*. 1<sup>st</sup> Ed. USA: John Wiley & Sons, Inc. 1:1:1–21. DOI:  
10.1002/9781118411360.wbcla127. Retrieved from:  
<http://onlinelibrary.wiley.com/doi/10.1002/9781118411360.wbcla127/abstract>

- De Grogoris, Grogorio (2015).  
The limits of expectations vs assessment questionnaire-based surveys on Simultaneous Interpreting Quality: The need for a gestaltic model of perception. [*Research Paper*]. Trieste: EUT Edizioni Università di Trieste. Retrieved from: [https://www.openstarts.units.it/dspace/bitstream/10077/11204/1/De\\_Grogoris.pdf](https://www.openstarts.units.it/dspace/bitstream/10077/11204/1/De_Grogoris.pdf)
- Dejean Le Féal, K. (accessed online in 2014).  
Non Nova, Sed Nove. [*E-Paper*]. (ESIT, Université de Paris 3). Retrieved from <http://www.openstarts.units.it/xmlui/bitstream/handle/10077/2490/13Dejan.pdf?sequence=1>
- Dodds, J. M. (1990).  
On the Aptitude of Aptitude Testing. *The Interpreters' Newsletter*, N° 3, 17-22.
- Ebel, R. L.; Frisbie, D. A. (1991).  
*Essentials of Educational Measurement*. 5<sup>th</sup> Ed. USA: Prentice-Hall.
- EMCI (2006),  
The Future of Conference Interpreting: Training, Technology and Research. [*Conference Proceedings*]. University of Westminster, 30 June- 1 July 2006. Graz.
- Ericsson, K. A. (2000).  
Expertise in Interpreting: An expert-performance perspective. *Interpreting*, 5(2): 187-220. Retrieved from: <http://dx.doi.org/10.1075/intp.5.2.08eri>
- European Masters in Conference Interpreting (2008).  
Graz Assessment seminar: Conference Interpreting Assessment- From novice to professional. [*Conference Proceedings*]. Proceedings of the Conference. April 5-6, 2008.
- European Personnel Selection Office (accessed 2015).  
Careers With European Union. [*Accessed Online*]. [http://europa.eu/epso/apply/jobs/perm/2014/translators/index\\_en.htm](http://europa.eu/epso/apply/jobs/perm/2014/translators/index_en.htm)
- Ewell, P. T. (2002).  
An emergency scholarship: A brief history of assessment. In T.W. Banta and Associates (Eds). *Building a scholarship of assessment*. San Francisco: Jossey-Bass, John Wiley & Sons.
- Fan, D. (February 2012).  
The development of expertise in interpreting through self-regulated learning for trainee interpreters. [*PhD Thesis*]. School of Modern Languages. University of Newcastle upon Tyne. Retrieved from: [https://theses.ncl.ac.uk/dspace/bitstream/10443/1741/1/Fan%2012%20\(12mth\).pdf](https://theses.ncl.ac.uk/dspace/bitstream/10443/1741/1/Fan%2012%20(12mth).pdf)
- Flowers, P. (2009).  
Research Philosophies-Importance and Relevance. *Research Philosophies-Importance and Relevance*, Issue 1, Jan. 2009
- Garzone, G. (2002).  
Quality and norms in interpretation. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21<sup>st</sup> Century: Challenges and Opportunities: Selected papers from the 1<sup>st</sup> Forlì Conference on Interpreting Studies, 9–11 November 2000*. Amsterdam/Philadelphia: Translation Benjamins Library.
- Garzone, G.; Viezzi, M. (Eds) (2002).  
*Interpreting in the 21<sup>st</sup> Century: Challenges and Opportunities: Selected papers from the 1<sup>st</sup> Forlì Conference on Interpreting Studies, 9–11 November 2000*. (pp. 107-119). Amsterdam/Philadelphia: Translation Benjamins Library.

- Genesee, F.; Upshur, J. A. (2001).  
*Classroom-based Evaluation in Second language education*. (5<sup>th</sup> ed.). Cambridge: Cambridge University Press.
- Gerver, D.; Longley, P. E.; Long, J.; Lambert, S. (1989).  
 Selection Tests for Trainee Conference Interpreters. *Meta: Translators' Journal*, 34(4), pp. 724-735
- Gile, D. (1985).  
 Le modèle d'efforts et l'équilibre d'interprétation en interprétation simultanée. *Meta: Translators' Journal*, 30 (1): 44-48.
- Gile, D. (1998a).  
 Observational Studies and Experimental Studies in the Investigation of Conference Interpreting. *Target 10(1)*. 69-93. <https://benjamins.com/catalog/target.10.1.04gil>
- Gile, D. (1998b).  
 L'évaluation de la qualité de l'interprétation en cours de formation. In *Meta: Translators' Journal*, 46 (2): 379-393. Retrieved from URI: <http://id.erudit.org/iderudit/002890ar>
- Gile, D. (2000).  
 Opportunities in conference interpreting research. In A. Beeby; D. Ensinger; M. Presas, *Investigating translation: Selected papers from the 4th International Congress on Translation*, Barcelona, 1998. Amsterdam/Philadelphia: John Benjamins.
- Gile, D. (2001).  
 L'évaluation de la qualité de l'interprétation en cours de formation. *Meta* 46(2): 379-394. Retrieved from : URI: <http://id.erudit.org/iderudit/002890ar>
- Gile, D. (2009).  
*Basic Concepts and Models for Interpreter and Translator Training*. Revised edition. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Gile, D. (February 2012),  
 Contribution of Research to interpreter training. [*Presentation*], Rome. Retrieved from: <http://www.cirinandgile.com/RomeContribresetointerprtrain.pdf>
- Gipps, C. V. (1994).  
*Beyond Testing: Towards a Theory of Educational Assessment*. London/Washington DC.: The Falmer Press. Published in the Taylor & Francis e-Library, 2003.
- Gipps, C.; Stobart, G. (2004).  
 Fairness in assessment. In *Persepectives on pupil assessment: GTC Conference: New Relationships: Teaching, Learning and Accountability*. (pp. 31-35). 29 Nov. 2004. London: General Teaching Council for England. Retrieved from: [http://dera.ioe.ac.uk/14022/1/1104\\_Perspectives\\_on\\_Pupil\\_Assessment\\_New\\_Relationships\\_Teaching\\_Learning\\_and\\_Accountability.pdf](http://dera.ioe.ac.uk/14022/1/1104_Perspectives_on_Pupil_Assessment_New_Relationships_Teaching_Learning_and_Accountability.pdf)
- Gipps, C.; Stobart, G. (2009).  
 Fairness in Assessment. In C. Wyatt-Simth and J. Cumming (Eds). *Educational Assessment in the 21<sup>st</sup> Century: Connecting Theory and Practice*. London/NY: Springer
- Gipps, G. R. (2010).  
 Qualitative Analysis Series. [*Video*]. [Retrieved from [www.YouTube.com](http://www.YouTube.com)]
- Gran, L.; Dodds, J. (1989) (Eds.).  
*The Theoretical and Practical Aspects of Teaching Conference Interpretation*. Udine: Campanotto editore.
- Grbić, N. (2008).  
 Constructing Interpreting quality. In *Interpreting*, 10(2): 232-257. Amsterdam/Philadephia: John Benjamins.

- Harmer, J. (1983),  
*The Practice of English Language Teaching*. (Longman Handbooks for Language Teachers, General Editor: Donn Byrne). NY: Longman, Inc.
- Harvey, L. (2004-12).  
 Analytic Quality Glossary. In *Quality Research International* (An international analytic glossary of issues related to quality in higher education), <http://www.qualityresearchinternational.com/glossary/index.htm#evaluation> (accessed online 2014).
- Havumetsä, N. (2012).  
 The Client Factor: A study of clients' expectations regarding non-literary translators and the quality of non-literary translations. [*Ph.D thesis*]. Faculty of Art of the University of Helsinki. ISBN 978-952-10-8392-1 (PDF).
- Horn, C.; Ramos, M.; Blumer, I.; Madaus, G.; Peter, S.; Carolyn, A. (2000).  
 Cut Scores: Results May Vary. [*NBETPP Monograph*]. 1(1). USA: National Board on Educational Testing and Public Policy; Lynch School of Education (Boston College). Retrieved from: <http://www.bc.edu/research/nbetpp/statements/M1N1.pdf>
- Iaroslavschi, M. (2012).  
 Écouter la cabine muette : petit guide d'évaluation en interprétation simultanée pour les interprètes en herbe. In *Synergies Roumanie*, 7: 177-183. Université Babeş-Bolyai, Cluj-Napoca, Roumanie.
- Inés, P. R. (2009).  
 Evaluating the process and not just the product when using corpora in translator education. In A. Beeby; P. Rodríguez Inés; P. Sánchez-Gijón (Eds), *Corpus use and translating : corpus use for learning to translate and learning corpus use to translate*. Amsterdam/Philadelphia: Benjamins Translation Library.
- The Interpreter Diaries (accessed 2015).  
 Entry test of the University of Westminster. [*E-Brochure*]. Retrived from: <http://theinterpreterdiaries.com/2011/05/25/the-aptitude-test-overcoming-the-first-hurdle/>
- Jones, R. (2002).  
*Conference interpreting explained*. (2nd ed). London & NY: Routledge. (Translation Practices Explained, 1470-966X). (St. Jerome publishing: 1998, 2002). (Accessed online in 2014) <http://www.ewidgetsonline.net/dxreader/Reader.aspx?token=e561e78f8ddc46938d913a60228f4e82&rand=1430212277&buyNowLink=&page=&chapter=>
- Jung, L. A.; Guskey, T. R. (2010).  
 Grading Exceptional Learners. Meeting Students Where They Are. *Educational Leadership*, 67 (5): 31-35. February 2010, Retrieved from: <http://www.ascd.org/publications/educational-leadership/feb10/vol67/num05/Grading-Exceptional-Learners.aspx> (accessed Saturday, September 12, 2015).
- Kalevi, K.; Torop, P. (2003).  
 Biotranslation: Translation between Umwelten. In S. Petrilli (Ed.), *Translation Translation (Approaches to Translation Studies 21)*. Amsterdam and New York, NY. 315-328.
- Kalina, S. (2000).  
 Interpreting Competences as a Basis and Goal for Teaching. In *The Interpreters' Newsletter*, 10: 3-32. Trieste: EUT - Edizioni Università di Trieste.

- Kalina, S. (2002).  
Quality in Interpreting and its prerequisites: A framework for a comprehensive view. In G. Garzone & M. Viezzi (Eds.), *Interpreting in the 21<sup>st</sup> Century: Challenges and Opportunities: Selected papers from the 1<sup>st</sup> Forlì Conference on Interpreting Studies, 9–11 November 2000*. Amsterdam/Philadelphia: Translation Benjamins Library.
- Kalina, S. (2005).  
Quality Assurance for Interpreting Processes. In *Meta : journal des traducteurs / Meta: Translators' Journal*, 50 (2): 768-784.  
<http://id.erudit.org/iderudit/011017ar>
- Kellett Bidoli, C. J. (accessed online in 2013).  
Quality Assessment in Conference Interpreting: An Overview.  
[http://www.openstarts.units.it/dspace/bitstream/10077/7689/1/Kellett\\_Bidoli\\_miscellanea\\_04.pdf](http://www.openstarts.units.it/dspace/bitstream/10077/7689/1/Kellett_Bidoli_miscellanea_04.pdf) (accessed 03/01/2013)
- Klimkowski, Konrad (2015).  
Towards a shared curriculum in translator and interpreter education. Series: Languages in Contact. Vol. 3. Wrocław/ NY: Wyższa Szkoła Filologiczna we Wrocławiu, Polska Akademia Nauk, Oddział we Wrocławiu & International Communicology Institute.
- Kurz, I. (1993).  
Conference Interpretation: Expectations of Different User Groups. The *Interpreters' Newsletter*, (5): 13-21
- Kusimo, P.; Ritter, M. G.; Busick, K.; Ferguson, C.; Trumbull, E.; Solano-Flores, G. (2012).  
Making assessment work for everybody: How to build on students strengths. [Report]. Assessment Laboratory Network of the Regional Educational Laboratories. Project Retrieved from: <http://www.sedl.org/pubs/tl05/assessment-full.pdf>. (accessed 2015).
- Lee, Y.-H. (2005).  
Self-assessment as an Autonomous Learning Tool in an Interpretation Classroom. *Meta: journal des traducteurs / Meta: Translators' Journal*, 50 (4). 2005. Retrieved from URI: <http://id.erudit.org/iderudit/019869ar>
- Lindquist E. F. (1951).  
Preliminary considerations in objective test construction. In E. F. Linguist (Eds). *Educational Measurement*. (pp. 119-159). Washington: American Council on Education. Reprinted In A. W. Ward, H. W. Stoker, M. Murray-Ward (Eds) (1996), *Educational Measurement: Origins, Theories and Explications*. Boston/London: University Press of America.
- Lindsay, M. (2010).  
The Philosophical Underpinnings of Educational Research. *Polyglossia*, Vol. 19, October 2010
- Louis, (2008).  
Les Limites du concept de compétence vu sous l'angle de l'évaluation. *Revue des sciences de l'éducation*. 34(3) : 753-758.
- Liu, Minhua (2009).  
How do experts interpret? Implication from research in interpreting studies and cognitive science. In G., Hansen; A., Chesterman; H., Gerzymisch-Arbogast. *Efforts and Models in Interpreting and Translation Research; A tribute to Daniel Gile*. USA: John Benjamins Publishing Company.
- MA Interpreting (Programme description). [Brochure].  
[Retrieved from : <http://www.surrey.ac.uk/postgraduate/interpreting>]

- MA in Translation & Interpreting (Programme description). [Brochure].  
[Retrieved from: <http://www.westminster.ac.uk/courses/subjects/languages/postgraduate-courses/full-time/p09fptai-ma-translation-and-interpreting>]
- MACINT Handbook (The University of Manchester). [Brochure]. Retrieved from:  
<http://www.llc.manchester.ac.uk/intranet/pg/ma-students/>
- MacMath, S.; Wallace, J.; Chi, X. (2009).  
Curriculum integration: opportunities to maximize assessment as, of, for learning.  
*McGill Journal of Education/Revue des sciences de l'éducation de McGill*, 44 (3): 451-465.
- The Master of Conference Aptitude Test (University of Glendon). [Video]. Retrieved from:  
<http://www.youtube.com/watch?v=4f5UKStRL5Y>
- Membership categories of the ITI (UK). [Brochure].  
<http://www.iti.org.uk/become-a-member/membership-categories>
- Martinez Melis, N.; Hurtado Albir, A. (2001).  
'Assessment in Translation Studies: Resear Needs', in *Meta*, XLVI, 2: 272-283.
- Messick, S. (1992).  
The interplay of evidence and consequences in the validation of performance assessments. [Research Report]. Vol. (1). [Article first published online: 8 AUG 2014]  
USA: Educational Testing Service.
- Messick, S. (1993).  
Foundations of validity: Meaning and consequences in psychological assessment.  
[Research Report]. USA: Educational Testing Service.
- Mikkelson, H. (2013).  
Universities and Interpreter Certification. *The International Journal for Translation & Interpreting Research*. 5(1), DOI: 10.12807/ti.105201.2013.a03.
- Miller, M. (2006).  
Assessment: A literature review. *Research and Information Services Bulletin*. N° 19.  
Glasgo: Scottish Qualifications Authority.
- Mislevy, R. J. and Knowles, K. T. (2002).  
Performance Assessments for Adult Education: Exploring the Measurement Issues.  
(p:42). [REPORT OF A WORKSHOP]. (Committee for the Workshop on Alternatives for Assessing Adult Education and Literacy Programs. Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education.). National Academy of Sciences. USA: NAP.
- Monacelli, C. (2009).  
*Self-preservation in Simultaneous Interpreting: Surviving the Role*. USA & The Netherlands: Benjamins Translation Library.
- Moser, P. (January 1995).  
*Survey on Expectations of Users of Conference Interpretation: Final Report*.  
Translation: Jennifer Mackintosh and Catherine Stenzl. Vienna/Geneva: AIIC.
- Moser-Mercer, B. (1985).  
Screening potential interpreters. *Meta: Translators' Journal*, 30(1): 97-101. Retrived from <http://id.erudit.org/ideerudit/003631ar>
- Moser-Mercer, B. (1996).  
Quality in interpreting: Some methodological issues. *The Interpreters' Newsletter*, 7: 43-55.

- Moser-Mercer, B. (2008).  
Skill acquisition in simultaneous interpreting: a human performance perspective. *The Interpreter and Translator Trainer*. 2(1): 1-28.
- Moser-Mercer, B. (accessed 2015a).  
Developing Expertise in Conference Interpreting, [Video], FTI, University of Geneva. (recordings realised between 2002 and 2004). (accessed online: <http://www.emcinterpreting.org/?q=node/197>)
- Moser-Mercer, B. (accessed 2015b).  
Measuring Quality in Consecutive Interpreting, [Video]. FTI. University of Geneva
- Napier, J. (2012).  
*Editorial: "Putting Interpreter Educators to the TEST: Testing, Ethics, and Technology. International Journal of Interpreter Education*, 4(1): 1-4. Conference of Interpreter Trainers.
- Odendahl, N. V. (2011).  
*Testwise: Understanding Educational assessment*, Volume 1. USA: Rowman and Littlefield Education.
- Oller, J. W., Jr. (1979).  
*Language tests at school*. London: Longman.
- Olsen, W. (accessed online in 2014).  
Triangulation in Social Research: Qualitative and Quantitative Methods Can Really Be Mixed. In M. Holborn and M. Haralambos (Eds.), *Developments in Sociology*. Edinburg: Causeway Press, 2004. [Retrieved from: <http://research.apc.org/images/5/54/Triangulation.pdf>]
- Online EU Training.  
Information Webcast for Interpreters/Translators. [Video]. [Retrieved from: <https://www.youtube.com/watch?v=JLXeN1DMnK8>]
- Osuala, R. .C. Onwuagboke; B. B. C. (2014).  
The Place of Information and Communication Technology in Students' Cognitive Assessment in Imo State Tertiary Institutions. *International Journal of Education and Research*. 2 (11). November 2014. ISSN: 2201-6333 (Print) ISSN: 2201-6740 (Online) www.ijern.com
- Peng, K.-C. (June 2006).  
The Development of Coherence and Quality Performance in Conference Interpreter Training. [Unpublished PhD Thesis]. University of Leeds. School of Modern Languages and Culture. Centre for Translation Studies.
- Phellas, C. N., Bloch, A.; Seale, C. (2012).  
Structured Methods: Interviews, Questionnaires and Observation [Online]. Retrieved from <https://docs.google.com> (accessed 12/01/2013)
- Pippa, S. (2009).  
L'évaluation sommative de l'interprétation simultanée: une réflexion sur les objectifs, les critères et sur les risques d'erreur. In *The Interpreters' Newsletter*, 14: 99-109. Trieste, EUT Edizioni Università di Trieste.
- Pöchhacker, F. (2001).  
Quality Assessment in Conference and Community Interpreting. In *Meta: Translators' Journal*, 46 (2). Retrieved from: <http://id.erudit.org/iderudit/003847ar> (accessed: 13 October 2012)
- Pöchhacker, F. (2002).  
Researching Interpreting Quality: Models and Methods. In G. Garzone & M. Viezzi (Eds), *Interpreting in the 21st Century: Challenges and opportunities. (selected papers*

- from the 1<sup>st</sup> Forli Conference on Interpreting Studies, 9-11 November 2000). Vol. 43. Amsterdam/Philadelphia: John Benjamins B. V.
- Pöchhacker, F. (2004).  
*Introducing Interpreting Studies*. London & NY: Routledge.
- Pöchhacker, F. (2011a).  
Conference Interpreting: Surveying the Profession. In R. Sela-Sheff & M. Shlesinger (Eds), *Identity and Status in the Translational Profession*, Amsterdam/Philadelphia: John Benjamins.
- Pöchhacker, F. (2011b).  
Assessing aptitude for interpreting: The SynClose Test. *Interpreting* 13(1): 106–120. doi 10.1075/intp.13.1.07poc. issn 1384–6647 / e-issn 1569–982X © John Benjamins Publishing Company
- Pöchhacker, F. (2012).  
Interpreting Quality: Global Professional Standards. In W. Ren (ed.), *Interpreting in the Age of Globalization: Proceedings of the 8th National Conference and International Forum on Interpreting* (pp. 305-318). Beijing: Foreign Language Teaching and Research Press.
- Ritchie, J.; Lewis, J. (2003).  
*Qualitative Research Practice: A Guide For Social Science Students and Researchers*. London: SAGE Publications.
- Saldana, J. (2009).  
*The Coding Manual for Qualitative Researchers*. California: SAGE Publications.
- Saunders, M., Lewis, Ph., Thornhill, A. (2007).  
*Research Methods for Business Students*. (7<sup>th</sup> ed.). England: Pearson Education Limited.
- Sawyer, B. D. (2004).  
*Fundamental Aspects of Interpreter Education: curriculum and assessment*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- SCASS/AEAC Glossary of Assessment Terms, (accessed online 2014).  
State Education Agency Directors of Arts Education (SEADAE) on behalf of NCCAS. ([www.nationalartsstandards.org](http://www.nationalartsstandards.org)).
- Schwietzer, J. W.; Ferriera, A. (Eds) (2014).  
The development of Translation Competence: Theories and Methodologies from psycholinguistics and Cognitive Science. UK: Cambridge Scholar Publishing.
- Setton, R.; Liangliang, A. G. (2011).  
Attitudes to role, status and professional identity in interpreters and translators with Chinese in Shanghai and Taipei. In R. Sela-Sheff & M. Shlesinger (Eds), *Identity and Status in the Translational Profession*. Amsterdam/Philadelphia: John Benjamins
- Setton, R. (1998).  
Meaning Assembly in Simultaneous Interpretation. In F. Pöchhacker; M. Schlesinger (Eds). *The Interpreting Studies Reader*. USA/Canada: Routledge.
- Shweda Nicholson, N. (2005).  
Personality characteristics of interpreter trainees: The Myers-Briggs Type Indicator (MBTI). *The Interpreters' Newsletter*, 13: 109-142.
- Stenzl, C. (1983).  
Simultaneous Interpretation – Groundwork towards a Comprehensive Model. [Unpublished MA thesis]. University of London.
- Szuki, A. (1988).  
Aptitudes of Translators and Interpreters. *Meta: Translators' Journal*, 33 (1): 108-114.

- Timarová, Š.; Ungeod-Thomas, H. (2009).  
The predictive validity of admissions tests for conference interpreting courses in Europe: A case study. In C. V. Angelelli & H. E. Jacobson (Eds), *Testing and Assessment in Translation and Interpreting Studies: A call for dialogue between research and practice*. USA & The Netherlands: John Benjamins B. V.
- Tsagari, D; Van Deemter, R. (Eds) (2013).  
Assessment Issues in Language Translation and Interpreting. Vol. 29. (Series: Language Testing and Evaluation, edited by R. Grotjahn & G. Sigott). Frankfurt: Peter Lang Edition GmbH.
- Valencia, S. W. (2002).  
Understanding Assessment: Putting together the puzzle. *Houghton Mifflin Reading*. USA: Houghton Mifflin Company.
- Volante, L. (2006a).  
Principles for Effective Classroom Assessment. *Brock Education*. 15 (2): 134-147.
- Volante, L. (2006b)  
An alternative vision for large-scale assessment in Canada. *Journal of Teaching and Learning*. 4(1): 1-14.
- Wallace, M. L. (November 2012).  
Predictors of Successful Performance on U.S. Consortium Court Interpreter Certification Exams. [*PhD Thesis*]. Directed by C. Glambruno. Department of Translation and Interpreting. Universidad de Alicante. Retrieved from: [http://rua.ua.es/dspace/bitstream/10045/28364/1/tesis\\_melissawallace.pdf](http://rua.ua.es/dspace/bitstream/10045/28364/1/tesis_melissawallace.pdf) (accessed online on 10/08/2014)
- Ward, A. W.; Stoker, H. W.; Murray-Ward, M. (1996) (Eds).  
Educational Measurement: Origins, Theories and Explications – Boston/London: University Press of America.
- Wiggins, G. (1993).  
*Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass.
- Wiggins, G. P. (1993).  
Assessing student performance: Exploring the purpose and limits of testing. The Jossey-Bass Education Series. San Francisco: Jossey-Bass publishers.
- Wiliam, D. (1998)  
Construct-referenced assessment of authentic tasks: alternatives to norms and criteria. [Conference paper]. the 24th Annual Conference of the International Association for Educational Assessment—Testing and Evaluation: Confronting the Challenges of Rapid Social Change, Barbados, May 1998. Retrieved from: [http://www.dylanwiliam.org/Dylan\\_Wiliams\\_website/Papers\\_files/IAEA%2098%20paper.doc](http://www.dylanwiliam.org/Dylan_Wiliams_website/Papers_files/IAEA%2098%20paper.doc) (Saturday, November 28, 2015).
- Wood, R. (1993).  
*Assessment and Testing: A Survey of Research*. Foreign Language Teaching and Research Press. UK: Cambridge University Press.
- Wu, S.-C. (2010),  
Some Reliability Issues of Simultaneous Interpreting Assessment Within Educational Context. In V., Pellatt; K., Griffiths; S.-C., Wu (Eds), *Teaching and Testing Interpreting and Translating*. Switzerland, Bern: PeterLang Ag, International Academic Publishers.

- Zhang, Z.; Burry-Stock, J. A. (2003).  
Classroom Assessment Practices and Teachers' Self-Perceived Assessment Skills. *Applied Measurement In Education*, 16(4), 323–342. Lawrence Erlbaum Associates, Inc.
- Zieky, M.; Perie, M. (2006).  
*A primer on setting cut scores on tests of educational achievement*. USA: ETS.
- Zwischenberger, C. (2009).  
Conference interpreters and their self-representation: A worldwide web-based survey. *Translation and Interpreting Studies*, 4(2): 239–253. DOI. 10.1075/tis.4.2.08zwi ISSN 1932–2798 / e-issn 1876–2700. John Benjamins Publishing Company
- Zwischenberger, C. (2010).  
Quality criteria in simultaneous interpreting: an international vs. a national view. *The Interpreters' Newsletter*, 15: 127-142. Trieste: EUT Edizioni Università di Trieste.
- Zwischenberger, C.; Pöchhacker, F. (Spring 2010)  
Survey on Quality and Role: Conference interpreters' expectations and self-perceptions. *Communicate!*. Retrieved from: <http://www.aiic.net/ViewPage.cfm/article2510.htm>
- Zwischenberger, C.; Behr, M. (Eds) (2015).  
*Interpreting Quality: A Look Around and Ahead*. Berlin: Frank & Timme GmbH.

## Secondary Sources:

- AERA (2000).  
Cautions issued about high-stakes tests. Available [Online Report]: [http://www.educationworld.com/a\\_issues/issues110.shtml](http://www.educationworld.com/a_issues/issues110.shtml).
- Agger-Gupta, N. (2001)  
From 'Making Do' to Established Service, the Development of Health Care Interpreter Services in Canada and the United States of America: A Grounded Theory Study of Health Organization Change and the Growth of a New Profession. [Unpublished PhD dissertation]. The Fielding Graduate Institute.
- Agosti, R. (1995).  
Stress ed interpretazione simultanea: un contributo sperimentale. [Unpublished graduation thesis]. SSLMIT. Università degli Studi di Trieste.
- AIIC (1991).  
Code of Professional Ethics (Code d'Ethique Professionnelle). [Report]. Genève.
- Andres, D. (2000).  
Konsekutivdolmetschen und Notizen. Empirische Untersuchungen mentaler Prozesse bei Anfängern in der Dolmetscherausbildung und professionellen Ausbildung. [Unpublished doctoral thesis], University of Vienna.
- Anderson, B.; Postl, B. (April 2001).  
Using large data sets as a basis for school improvement. [Conference Proceedings]. Paper presented at the Annual Meeting of the American Educational Research Association.
- Arjona, E. (1984).  
Issues in the design of curricula for the professional education of translators and interpreters. In M. L. McIntire (Ed.), *New Dialogues in Interpreter Education. Proceedings of the Fourth National Conference of Interpreter trainers Convention* (pp. 1-35). Silver Spring: RID.

- Babbie, E. (1992).  
The practice of social research. Belmont, California: Wadsworth. Sixth Edition.
- Barik, H.C. (1969).  
A study of simultaneous interpretation. [*University of North Carolina doctoral thesis*].  
Chapel Hill. Ann Arbor Michigan.
- Canadian Teachers Federation. (1999).  
Province-wide assessment programs. [*Online Report*]: <http://www.ctf-fce.ca/e/what/other/assessment/testing-main.htm>.
- Cartellieri, C. (1983).  
The inescapable dilemma. Quality and/or quantity in interpreting. *Babel*, 29(4): 209-213.
- Chernov, G. V. (1985).  
Interpretation Research in the Soviet Union: Results and Prospects. [*Conference Proceedings*]. Proceedings of the Xth World Congress of the FIT, pp. 169-177, H. Bühler (Ed.), Vienna, Braumüller.
- Cheung, A. (sous presse).  
Does accent matter? A comparison study on the impact of accent in simultaneous interpretation performance quality and satisfaction level. [*Conference Proceedings*]. *Actas del I Congreso Internacional sobre Evaluación de la Calidad en Interpretación de Conferencias*. Almuñécar, Granada.
- Collados Aís, Á. (1998).  
*La evaluación de la calidad en interpretación simultánea*. La importancia de la comunicación no verbal. Granada: Comares.
- Cooper C. L.; Davies R.; Tung R. L. (1982).  
Interpreting stress: sources of job stress among conference interpreters. *Multilingua*, 1(2): 97-107.
- Darling-Hammond, L. (2004).  
Standards, accountability, and school reform. *Teachers College Record*, 106(6): 1047-85.
- Darò, V. (1994).  
Non-linguistic factors influencing simultaneous interpretation. In S. Lambert & B. Moser Mercer (Ed.). *Bridging the Gap: Empirical Research in the Teaching of Simultaneous Interpreting* (pp. 249-271). Amsterdam/Philadelphia: John Benjamins,
- Déjean Le Féal, K. (1990).  
Some thoughts on the evaluation of simultaneous interpretation. In D. Bowen & M. Bowen (Eds.), *Interpreting – yesterday, today, and tomorrow*. Binghamton, NY: SUNY, 154-160.
- Doerflinger, O. (sous presse).  
Gestion de la qualité totale. [*Conference Proceedings*]. *Actas del I Congreso Internacional sobre Evaluación de la Calidad en Interpretación de Conferencias*. Almuñécar, Granada.
- Dornic, S. (1978).  
The bilingual's performance: language dominance, stress and individual differences. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 259-270). NY: Plenum Press.
- Edwards, R.; Temple, B.; Alexander, C. (2005).  
Users' experiences of interpreters: The critical role of trust. *Interpreting* 7(1): 77-95.

- Falk, B. (1998).  
Testing the way children learn: Principles for valid literacy assessment. *Language Arts*, 76(1): 57-66.
- The First BABELEA Conference (November, 1999).  
On Community Interpreting, [*Conference Proceedings*]. Vienna.
- Galli, C. (1990).  
Simultaneous interpretation in medical conferences: a case study. In L. Gran & C. Taylor (Eds), *Aspects of Applied and Experimental Research in Conference Interpretation* (pp. 61-82). Università degli di Trieste. Udine: Campanotto.
- Garzone, G. (sous presse).  
Quality and norms in interpretation". [*Conference Proceedings*]. *Actas del I Congreso Internacional sobre Evaluación de la Calidad en Interpretación de Conferencias*. Almuñécar, Granada.
- Gerver, D.; Longley, P.; Long, J.; Lambert, S. (1984).  
Selecting trainee conference interpreters. *Journal of Occupational Psychology*, 57: 17-31.
- Gerver, D. (1971).  
*Simultaneous and consecutive interpretation and human information processing*. Durham: Social Science Research Council.
- Gerver, D. (1976).  
Empirical studies of simultaneous interpretation: a review and a model. In R. W. Brislin (Ed.), *Translation, Applications and Research* (pp. 165-207). NY: Gardner Press Inc.
- Gile, D. (1991).  
A Communication-Oriented Analysis of Quality in Nonliterary Translation and Interpretation. In M. L. Larson (Ed.), *Translation: Theory and Practice. Tension and Interdependence* (pp. 188-200). Binghamton NY: SUNY
- Gile, D. (1984).  
Les noms propres en interprétation simultanée". *Multilingua*, 3(2) : 79-85.
- Gile, D. (1990).  
L'évaluation de la qualité de l'interprétation par les délégués: une étude de cas. *The Interpreters' Newsletter*, 3: 66-71. SSLMIT. Università degli studi di Trieste.
- Gile, Daniel (1995).  
Fidelity Assessment in Consecutive Interpretation: An Experiment. *Target* 7:1, 151-164.
- Gile, D. (1998).  
Internationalization and institutionalization as promoters of interpretation research. In H. Gerzymisch-Arbogast, D. Gile, J. House, & A. Rothkegel (Eds.), *Wege der Übersetzungs- und Dolmetschforschung* (pp. 167-178). Tübingen: Gunter Narr.
- Glaser, R.; Linn, R.; Bohrnstedt, G. (1997).  
*Assessment in transition: Monitoring the nation's educational progress*. Stanford, CA: National Academy of Education (1997).
- Grbić, N. (2006).  
Reconstructing the system of sign language interpreting in Styria. In A. Pym, M. Shlesinger & Z. Jettmarová (Eds.), *Sociocultural aspects of translating and interpreting*. Amsterdam/Philadelphia: John Benjamins.
- Gringiani, A. (1990).  
Reliability of Aptitude Testing: A Preliminary Study. In L. Gran & C. Taylor (Eds), *Aspects of Applied and Experimental Research on Conference Interpretation* (pp. 42-53). Udine: Campanotto editore.

- Glaser, R. (1963).  
Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18: 519–21.
- Gullikson, A. R. (1984).  
Teacher perspectives of their instructional use of tests. *Journal of Educational Research*, 77(4): 244–248.
- Guskey, T. R. (2006).  
Making high school grades meaningful. *Phi Delta Kappan*, 87(9): 670–675.
- Herbert, J. (1952).  
*The interpreter's handbook*. Genève : Gearg.
- Hess, F. M., & Brigham, F. (2000).  
None of the above: The promise and peril of high-stakes testing. *American School Board Journal*, 187(1): 26-29.
- Hönig, Hnas G. (1995).  
*Konstruktives Übersetzen*. Tübingen: Stauffenburg verlag.
- Horguelin (1985).  
*Pratique de la révision*, 2e Ed, Montréal, Linguattech.
- Jung, L. A.; Guskey, T. R. (2007).  
Standards-based grading and reporting: A model for special education. *Teaching Exceptional Children*, 40(2): 48–53.
- Jung, L. A. (2009).  
The challenges of grading and reporting in special education: An inclusive grading model. In T. R. Guskey (Ed.), *Practical solutions for serious problems in standards-based grading* (pp. 27–40). Thousand Oaks, CA: Corwin Press.
- Kahane, E. (2000).  
Thoughts on the Quality of Interpretation. [Online]. Retrieved from <http://www.aicc.net/ViewPage.cfm/page197.htm> (13.05.2000).
- Kalina, S. (1998).  
*Strategische Prozesse beim Dolmetschen. Theoretische Grundlagen, empirische Fallstudien, didktische Konsequenzen*. Tübingen: Narr.
- Kane, M. (2002).  
Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, 21(1): 31-41.
- Keiser, W. (1965).  
Definitions d'aptitude. [Conference Proceedings]. In AIIC, *Colloque sur l'enseignement de l'interprétation*, Paris 18–19 Décembre, 28–37.
- Keiser, W. (1978).  
Selection and training of conference interpreters. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 11-24). NY: Plenum Press.
- Klonowicz, T. (1994).  
Putting one's heart into simultaneous interpreting. In S. Lambert & B. Moser-Mercer (Eds), *Bridging the Gap; Empirical Research in the Teaching of Simultaneous Interpreting* (pp. 213-224). Amsterdam/Philadelphia: John Benjamins.
- Kohn, A. (2000).  
*The case against standardized testing: Raising scores, ruining the schools*. Portsmouth, NH: Heineman.

- Kopczynski, A. (1994).  
Quality in conference interpreting: Some pragmatic problems. In M. Snell Hornby, F. Pöchhacker & K. Kaindl (Eds.), *Translation studies: An interdisciplinary* (pp189–198). Amsterdam/Philadelphia: John Benjamins.
- Kourganoff, V. (1971).  
*La recherche scientifique*, Paris:PUF.
- Kurz I. (1988).  
Conference interpreting: user expectations. [*Conference Proceedings*]. In A.T.A. *Proceedings of the 30<sup>th</sup> Annual Conference*. Medford. NJ: Learned Information Inc.
- Kurz, I. (1989).  
Conference Interpreting - user expectations. In D. L. Hammond (Ed.) (1989). *Coming of age. Proceedings of the 30th Conference of the ATA*. Washington, D.C., Melford, N.J. 143-148.
- Kurz, I.; Pöchhacker, F. (1995).  
Quality in TV Interpreting. *Translatio - Nouvelles de la FIT - FIT Newsletter*, XIV 3-4 : 350-358.
- Kurz, I. (1998).  
Dolmetschleistungen. In M. Snell-Hornby, H. G. Hönic, P. Kussmaul & P. A. Schmitt (Eds.), *Handbuch Translation* (pp. 391-393). Tübingen: Stauffenburg.
- Kurz, I. (2003).  
Quality from the user perspective. In A. Collados Aís, M. M. Fernández Sánchez & D. Gile (Eds.), *La evaluación de la calidad en interpretación : Investigación*. Granada : Comares, (pp. 3-22).
- Kutz, W. (1997).  
Gut für wen? Zur Bewertung von Konsekutivdolmetschleistungen. In E. Fleischmann; W. Kutz & P.A. Schmitt (Eds.), *Translationsdidaktik*. Grundfragen der Übersetzungswissenschaft, Tübingen, Gunter Narr, (pp. 243-254).
- Laver, J. (1980).  
*The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.
- Laver, J. (1991).  
*The Gift of Speech: Papers in the Analysis of Speech and Voice*. Edinburgh: Edinburgh University Press.
- Lambert, S. (1992).  
Aptitude testing for simultaneous interpretation at the University of Ottawa. *The Interpreters' Newsletter*, 4: 25–32.
- Lederer, M. (1975).  
Le colloque sur les examens d'entrée aux Écoles d'Interprètes. *AIIC Bulletin*, 3 (2): 25–28.
- Liu, M. (2001).  
Expertise in Simultaneous Interpreting: A Working Memory Analysis. [*Unpublished dissertation*]. University of Texas at Austin.
- Mack, G.; Cattaruzza, L. (1995).  
User surveys in SI: a means of learning about quality and/or raising some reasonable doubts. In J. Tammola, (Ed.). *Topics in Interpreting Research*. Turku, Centre for Translation and Interpreting, University of Turku, 37-49.
- Marrone, S. (1993).  
Quality: a shared objective. *The Interpreters' Newsletter*, 5: 35-41.

- Mazza, C. (2000).  
*Numbers in simultaneous interpretation*. Tesi di Laurea, SSLMIT, Università di Studi di Bologna, Sede di Forlì.
- McIlvaine, P. H. (1978).  
 Human factors approach to simultaneous interpretation. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 315-321). NY: Plenum Press.
- McNeil, L. (2000).  
 Creating new inequalities: Contradictions of reform. *Phi Delta Kappan*, 81(10): 729-34.
- Meaghan, D. E.; Casas, F. R. (1995).  
 On standardized achievement testing: Response to Freedman and Wilson and a last word. *Interchange*, 26(1): 81-96.
- Meak, Lidia. (1990).  
 Interpretation simultanée et congrès médical: attentes et commentaires. *The Interpreters' Newsletter*, 3: 8-13. SSLMIT. Università Degli Studi Trieste.
- Miller, E.; Tovey, R. (Eds.). (1996).  
*Motivation, achievement, and testing*. Boston, MA: Harvard Education Press.
- Ministry of Justice (2011).  
*The Quality Assurance Test/ALS Interpreter Quality Assurance (QA) Test Candidates' Handbook* October 2011, developed by Middlesex University), (p2-3)
- Moser-Mercer, Barbara (1978).  
 Simultaneous Interpretation: A Theoretical Model and its Practical Application. In D. Gerver & W. H. Sinaiko, *Language, Interpretation and Communication* (pp. 353-368). New York: Plenum Press.
- Mouly, George J. (1978).  
*Educational Research: The Art and Science of Investigation*. Boston: Allyn & Bacon.
- Mousavi, S. A. (2009).  
*An encyclopedic dictionary of language testing*, (4<sup>th</sup> ed.). Tehran: Rahnama Publications.
- Ozolins, U. (2000)  
 Communication Needs and Interpreting in Multilingual Settings: The International Spectrum of Response. In Roberts *et al.* (Eds), (pp. 21–33).
- Pearl, S. (1995).  
 Lacuna, myth and shibboleth in the teaching of simultaneous interpreting. *Perspectives* 3 (2): 161–190.
- Pöchhacker, F. (1995a).  
 Those Who Do...: A Profile of Research(ers) in Interpreting. *Target*, 7(1): 47-64.
- Pöchhacker, F. (1995b).  
 Writings and Research on Interpreting: A Bibliographic Analysis. *The Interpreter's Newsletter*, 6: 17-32.
- Politi, M. (1989).  
 Le signal non-verbal en interprétation simultanée. *The Interpreters' Newsletter*, 2: 6-10. SSLMIT. Università Degli Studi Trieste.
- Popham, W. J. (2001).  
 Teaching to the test. *Educational Leadership*, 58(6): 16-20.
- Poyatos, F. (1993).  
*Paralanguage: A linguistic and interdisciplinary approach to interactive speech and sound*. Amsterdam/Philadelphia: John Benjamins.

- Riccardi, A; (2003).  
The relevance of interpreting strategies for defining quality in simultaneous interpreting. In A. Collados Aís, M. M. Fernández Sánchez & D. Gile (Eds.), *La evaluación de la calidad en interpretación : Investigación*. Granada : Comares, 257-265.
- Russo M.; Rucci M. (1997).  
Verso una classificazione degli errori nella simultanea dallo spagnolo in italiano. In L. Gran & A. Riccardi (Eds), *Nuovi orientamenti negli studi sull'interpretazione* (pp. 179-199). Università Degli Studi Trieste.
- Russo, C. (1989).  
Text Processing Strategies: A Hypothesis to Assess Students' Aptitude for Simultaneous Interpreting. *The Interpreters' Newsletter*, N° 2, (Eds. L. Gran & J; Dodds), SSLM, Univ. of Trieste, Trieste, pp. 57-64
- Schmitt, K. (1995).  
What is Licensure?. In J. C. Impara (Ed), *Licensure Testing: Purposes, Procedures, and Practices* edited (pp. 3–32). Lincoln, NE: Buros Institute of Mental Measurements and Department of Educational Psychology, University of Nebraska-Lincoln.
- Schmitt, P. A. (1998).  
Qualitätsmanagement; In M. Snell-Hornby, H. G. Hönl, P. Kussmaul & P; A. Schmitt (Eds.), *Handbuch Translation* (pp.394-399). Tübingen: Stauffenburh.
- Seleskovitch, D.; Lederer, M. (1989).  
*Pédagogie raisonnée de l'interprétation*. Paris: Didier Erudition.
- Setton, R. (1999).  
*Simultaneous Interpretation: A Cognitive-Pragmatic Analysis*. Amsterdam and Philadelphia: John Benjamins.
- Shlesinger, M. (1994).  
Intonation in the production and perception of simultaneous interpretation. In S. Lambert & B. Moser Mercer (Ed.), *Bridging the Gap. Empirical Research in the Teaching of Simultaneous Interpreting* (pp 225-256). Amsterdam/Philadelphia: John Benjamins,
- Shlesinger, M. (1995).  
Stranger in paradigms: What lies ahead for simultaneous interpreting research?. *Target*, 7(1): 7–28.
- Shlesinger, M. (1997).  
Quality in simultaneous interpreting. In Y. Gambier, D. Gile & C. Taylor (Eds.), *Conference interpreting: Current trends in research*. (pp. 123–131). Amsterdam/Philadelphia: John Benjamins.
- Shweda Nicholson, N. (2005).  
Self-monitoring strategies in simultaneous interpretation. In ITI Proceedings, Conference 5 (pp, 46-51). C. Picken (Ed). London: Aslib.
- Smith, M. L.; Fey, P. (2000).  
Validity and accountability of high-stakes testing. *Journal of Teacher Education*, 51(5): 334-44.
- Snelling, D. (1989).  
A typology of interpretation for teaching purposes in L. Gran & J. Dodds (Eds.), *The Theoretical and Practical Aspects of Teaching Interpretation*. (pp. 141-142). Udine: Campanotto

- Snell-Hornby, M. (1992).  
The professional translator of tomorrow: Language specialist or all-round expert? In C. Dollerup & A. Loddegaard (Eds.), *Teaching Translation and Interpreting: Training, Talent and Experience* (pp. 9–22). Amsterdam and Philadelphia: John Benjamins.
- Snell-Hornby, M. (1998).  
Ausbildungssituation in Europa; In M. Snell-Hornby, H. G; Hönig, P. Kussmaul, & P. A. Schmitt (Eds.), *Handbuch Translation* (pp. 31-33). Tübingen: Stauffenburg.
- Sofr, O. (1976).  
Test of aptitude for interpreting. *L'interprète* 31 (3/4): 3–8.
- Stejskal, J. (2003).  
*International Certification Study*. Alexandria, VA: American Translators Association.
- Stenzl, C. (1989).  
From theory to practice and from practice to theory. In L. Gran & J. Dodds (Eds.), *The Theoretical and Practical Aspects of Teaching Conference Interpretation*, (pp. 266). Udine: Campanotto editore.
- Stiggins, R. (2002).  
Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10): 758-65.
- Stiggins, R. J. (2007).  
An introduction to student-involved assessment for learning (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Strong, M.; Rudser, S. F. (1986/1992).  
The subjective assessment of sign language interpreters. In D. Cokely (Ed.), *Sign language interpreters and interpreting* (pp. 1-14). Burtonsville, MD: Linstock Press.
- Taylor, C. (1990).  
Conference and World Knowledge. In L. Gran & C. Taylor (Eds.), *Aspects of Applied and Experimental Research on Conference Interpretation* (pp. 21-27). Udine: Campanotto editore.
- Thiéry, C. (1990).  
The Sense of Situation in Conference Interpreting. In D. & M. Bowen (Eds), *Interpreting – Yesterday, Today and Tomorrow* (American Translators Association Scholarly Monograph Series IV) (pp. 40-43). Binghamton, New York: SUNY.
- Tryuk, M. (2002).  
Les tests d'aptitude et leur role dans la formation des interprètes de conference. In B. Lewandowska-Tomaszczyk & M. Thelen (Eds.), *Translation and Meaning*, Part 6. Maastricht: UPM, 421–434.
- Tseng, J. (1992).  
Interpreting as an Emerging Profession in Taiwan – a Sociological Model. [*Unpublished MA thesis*]. Fu Jen Catholic University, Taipeh.
- Valencia, R. R.; Villarreal, B. J. (2003).  
Improving students' reading performance via standardsbased reform: A critique. *The Reading Teacher*, 56(7): 612-21.
- Volante, L. (2004).  
Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, 35. [Online]: <http://www.umanitoba.ca/publications/cjeap/articles/volante.html>.

- Vuorikoski, A.-R. (1993).  
 Simultaneous interpretation - User Experience and expectations. In C. Picken (Ed.), *Translation - the vital link. Proceedings XIII FIT World Congress*. Brighton: Institute of Translation and Interpreting, 317-327.
- Wiggins, G. (1996).  
 Honesty and fairness: Toward better grading and reporting. In T. R. Guskey (Ed.), *Communicating student learning: 1996 Yearbook of the Association for Supervision and Curriculum Development* (pp. 141–176). Alexandria, VA: ASCD.
- Wideen, M. R.; O’Shea, T.; Pye, I.; Ivany, G. (1997).  
 High-stakes testing and the teaching of science. *Canadian Journal of Education*, 22(4): 428-44.
- Wiggins, G. (1996).  
 Practicing what we preach in designing authentic assessments. *Educational Leadership*. 54(4). Teaching for Authentic Student Performance, (pp. 18-25). December 1996/January 1997. [Online]. <http://www.ascd.org/publications/educational-leadership/dec96/vol54/num04/Practicing-What-We-Preach-in-Designing-Authentic-Assessments.aspx>
- Wood, R. (1986).  
 The agenda for educational measurement. In D. NUTTALL (Ed), *Assessing Educational Achievement*. London: Falmer Press.
- Zuberga, I. (2001).  
 Variables in quality assessment in interpreting. In A. F. Kelleter (Ed.), *Dolmetschen: Beiträge aus Forschung, Lehre und Praxis* (pp. 279-286). Frankfurt am main (etc.): Peter Lang.

# Appendices

**Table 13: Data Coding of the Pilot Study Questionnaire (for Teachers):**

Theme(s)/Categories	Textual Responses for :	Axial Code	Sub-Code
Theme: Background	Q1: Age range:		
Sub-theme: Age range	TEA1 : /// TEA2: /// TEA3: 45 to 50	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• 40-50</li> </ul>	
• Old			
• Young			
Theme : Background	Q2: Qualifications:		
Sub-theme : Qualifications	TEA1 : /// TEA2 : PhD in translation and interpretation TEA3 : Master degree in translation Ar, Fr, En, Sp Master degree 2 in interpretation Ar, Fr, En	<ul style="list-style-type: none"> <li>• University Graduate Qualification</li> <li>• Professional Qualification</li> </ul>	<ul style="list-style-type: none"> <li>- MA</li> <li>- PhD</li> <li>- MA</li> <li>- Hands-on</li> </ul>
• Academic background			
• Professional background			
Theme: Background	Q3: Work Experience:		
Sub-theme : Experience	TEA1 : /// TEA2 : Interpreter since 1979 TEA3 : From 1988 to 1991 in Algeria Since 1991 to date in Europe with a professional domicile in Brussels	<ul style="list-style-type: none"> <li>• 10 years exp.</li> <li>• More than 10 years</li> <li>• T did not answer the question (TDAQ)</li> </ul>	<ul style="list-style-type: none"> <li>- Local</li> <li>- International</li> <li>- Local</li> <li>- International</li> </ul>
• Long			
• Short			
Theme: Background	Q4: When did you start teaching simultaneous interpreting?		
Sub-theme: SI teaching exp.	TEA1 : 1996 TEA2 : /// TEA3 : 2006	<ul style="list-style-type: none"> <li>• Less than 10 years</li> <li>• More than 10 years</li> <li>• TDAQ</li> </ul>	
• Experienced			
• Unexperinecd			
Theme: Background	Q5: Did you have any training in assessment matters? If yes, would you specify? <input type="checkbox"/> Yes <input type="checkbox"/> No		
Sub-theme: Training in assessment	TEA1 : Yes; in the MA course in Britain TEA2 : Yes. My work with professional interpreters during conferences gave me the opportunity to see first-hand how the assessment is done as well as the qualifications required in an interpreter to be a good one TEA3 : No. When you are a practicing interpreter/ teacher, you assess based on the factual specifications of the market	<ul style="list-style-type: none"> <li>• Academic training</li> <li>• Professional training</li> </ul>	<ul style="list-style-type: none"> <li>- Professional</li> <li>- Research</li> <li>- Hands-on</li> <li>- Observation</li> </ul>
• Trained in assessment			
• Untrained in assessment			
Theme: Aptitude assess.	Q6: What assessment is needed to ascertain if someone has the aptitude to be an interpreter?		
• Criterion-referenced assessment	TEA1: Tests in the same activity (Simultaneous Interpretation)..... TEA2: - Complete mastery of the active language(s) - Presence of mind - The ability to analyze information and spot the different nuances in the utterance - The ability to adapt to changing contexts and situations - Pleasant voice and public-speaking skills - Powers of concentration regardless of the situation - General Knowledge TEA3 : First, we have to make sure the prerequisite are met before you go further in finding out whether or not a candidate has the aptitude or not to be an interpreter.	<ul style="list-style-type: none"> <li>• Assess. of Basic skills</li> <li>• Assess. of Cognitive skills</li> <li>• Assess. of Knowledge</li> <li>• SI Test</li> </ul>	
• Norm-referenced assessment			
• Ipsative assessment			
• Formative assess.			
• Diagnostic assess.			
• Direct assessment			
• Indirect assessment			
Theme: Perception of Interpreting Quality	Q7: What is interpreting quality for you? How would you define it?		

Sub-theme: Definition of Interpreting Quality	TEA1 : It is the ability to respond at the right time in a different language and to the point according to the task TEA2 : // TEA3 : Quality is to be faithful to the original speech. If what was meant to be conveyed is indeed delivered to the listener, what else would you expect.	<ul style="list-style-type: none"> <li>• Abstract definition</li> <li>• Procedural definition</li> <li>• TDAQ</li> </ul>	- Broad Hypothetical perfect performance
Theme: Perception of Interpreting Quality	Q8: How do you assess interpreting quality?		
Sub-theme: Quality Assessment	TEA1 : The same as in (7) above, provided that the interpreter conveys the message and its impact on the target language audiences just as the audiences of L1 text understand it  TEA2 : - Sense and consistency with the original message - Logical cohesion of utterance - Correct grammatical usage - Completeness of interpretation - Pleasant voice  TEA3 : By looking at the reaction of the target group. If listeners interact smoothly in a meeting, then the mission is accomplished. In a class setting, if the teacher gets the meaning and not the words of the original message, that does it. However, even though appropriate lexis is a pre requisite, it can't be enough to render proper interpretation unless it is backed by a proper understanding of the gist of the message.	<ul style="list-style-type: none"> <li>• Product-oriented assess.</li> <li>• Process-oriented assess.</li> </ul>	<ul style="list-style-type: none"> <li>- Theoretical Product</li> <li>- Classroom Product</li> <li>- Market Product</li> <li>- University graduation levels</li> <li>- Professional mastery levels</li> </ul>
Theme: Test & Exam Construction & Design	Q9: What do you consider before taking decisions on the tests specifications?		
Sub-theme: Test specifications	TEA1 : the aim of the test, the applicants and their abilities and what they are expected to do in the would-be job  TEA2 : - The general level of students - Their knowledge of the subject - Their flexibility and ability to adjust to subjects they're not familiar with  TEA3 : Before deciding on specifications, a number of factors need to be taken into account. Observation: students need to be put in a genuine interpretation situation. In real life, interpreters are aware of the subject matters and prepare for it. With this in mind, the teacher should: - Make sure students have the pre requisites in terms of mastery of relevant lexis, knowledge of the topics to be covered at the tests. - Provide for technical as well as general topics. Students perform differently and should be given the opportunity to interpret in both settings	<ul style="list-style-type: none"> <li>• Aim</li> <li>• Factors</li> <li>• Real-job tasks</li> </ul>	<ul style="list-style-type: none"> <li>• Students' level</li> <li>• Students' abilities</li> <li>• Students' Flexibility</li> <li>• Students' knowledge (topic, lexis, terminology,...)</li> <li>- Reference to a genuine professional setting</li> </ul>
Theme: Test & Exam Construction & Design	Q10: What is the construct for simultaneous interpreting?		
Sub-Theme: SI construct(s)	TEA1 : knowledge of two languages and the topic concerned  TEA2 : - To have a sharpness of mind - To keep a cool head	<ul style="list-style-type: none"> <li>• Language mastery</li> <li>• Topic</li> <li>• Personality traits</li> <li>• Invisibility</li> </ul>	

	- To remain as much invisible as possible to the audience TEA3 :////////////////////////////////////	• TDAQ	
<b>Theme: Test &amp; Exam Construction &amp; Design</b>	<b>Q11: What kind of tasks is appropriate to assess interpreting quality as you perceive it?</b>		
<b>Sub-Theme: Tasks/Exercises</b>	TEA1 : Tests in interpreting in the field of knowledge	• SI activity tasks	
<ul style="list-style-type: none"> <li>• According to the Level of Study</li> <li>• According to the level of Performance</li> <li>• According to the Level of Proficiency</li> </ul>	TEA2 : - Assign the trainees with the task of synthesizing in the target language different prompts <ul style="list-style-type: none"> <li>- Test their ability and flexibility when working in an unconventional context</li> <li>- Working in groups (of two) to gauge their entrepreneurial spirit</li> <li>- The ability to do a comprehensive research on the subject they are assigned with</li> </ul> TEA3 : Sight translation builds up to other types of interpretation techniques Listening for the gist Reformulating ideas in the same language Reformulating ideas in the target languages	<ul style="list-style-type: none"> <li>• Preparation for SI</li> <li>• Real-life tasks that are connected to SI</li> </ul>	<ul style="list-style-type: none"> <li>- Non-hierarchical</li> <li>- Hypothetical tasks</li> </ul>
<b>Theme: Test &amp; Exam Construction &amp; Design</b>	<b>Q12: On what basis do you choose the prompts of the test?</b>		
<b>Sub-theme: Prompts' Choice</b>	TEA1 : the aim of tests and the job to be handled by the successful applicant	• Training	
<ul style="list-style-type: none"> <li>• Level of study</li> <li>• Level of Performance</li> <li>• Course content</li> <li>• Target skill or sub-skill</li> <li>• Rate of delivery</li> <li>• Authenticity</li> <li>• Level of proficiency</li> </ul>	TEA2 : - According to the training the students have received so far (practical and theoretical) <ul style="list-style-type: none"> <li>- Subjects with which the students are less comfortable in order to make them get used to tricky unexpected contexts</li> </ul> TEA3 : Prompts are decided upon based on real life work in conferences and seminars. There is no point is accommodating students by slowing the pace or being complacent.	<ul style="list-style-type: none"> <li>• Content</li> <li>• Real-life work</li> </ul>	<ul style="list-style-type: none"> <li>- Aim of the test</li> <li>- Students' Progress</li> <li>- Unexpected subjects</li> <li>- Genuine prompts</li> </ul>
<b>Theme: Test &amp; Exam Construction &amp; Design</b>	<b>Q13: On what basis do you decide on the prompt(s)' length?</b>		
<b>Sub-theme: Prompt's Length</b>	TEA1 : on the basis of aim of test and the would-be job of applicant	• Test Aim	
<ul style="list-style-type: none"> <li>• Level of study</li> <li>• Level of Performance</li> <li>• Level of proficiency</li> <li>• Course content</li> <li>• Target skill</li> <li>• Rate of delivery</li> <li>• Authenticity</li> </ul>	TEA2 : - Based on the trainees abilities and real-life situations <ul style="list-style-type: none"> <li>- Based on the standard time an utterance takes during a conference</li> </ul> TEA3 : See above	<ul style="list-style-type: none"> <li>• Real-life job</li> <li>• Trainees' abilities</li> </ul>	<ul style="list-style-type: none"> <li>- Conferences' standards</li> </ul>
<b>Theme: Test &amp; Exam Construction &amp; Design</b>	<b>Q14: What do you expect from the trainees?</b>		
<b>Sub-theme: Performance Expectations</b>	TEA1 : to be competent in the two languages and the topic concerned (e.g. medicine)	• High-order expectations	
<ul style="list-style-type: none"> <li>• Task's Performance</li> <li>• Learner's Performance</li> <li>• Novice's Performance</li> </ul>	TEA2 : - To have the required terminology <ul style="list-style-type: none"> <li>- To have communicative skills</li> </ul>		<ul style="list-style-type: none"> <li>- Professional performance</li> <li>- Graduate students performance</li> </ul>

<ul style="list-style-type: none"> <li>Professional's Performance</li> </ul>	<ul style="list-style-type: none"> <li>To be accurate</li> <li>To be culturally sensitive</li> <li>To show a measure of confidence while conveying the message</li> </ul> <p>TEA3 : Not to underestimate their obligations. If one wants to be an interpreter, taking things lightly won't make it. Therefore, one is expected to be:</p> <ul style="list-style-type: none"> <li>Accurate in listening and focused</li> <li>Quick in analyzing the message</li> <li>Have enough relevant lexis to render swiftly the received message</li> <li>Control their delivery (voice, pace, body language ...)</li> </ul>	<ul style="list-style-type: none"> <li>Low-order expectations</li> </ul>	<ul style="list-style-type: none"> <li>Mid-term performance</li> <li>Task performance</li> <li>Sub-skill performance</li> </ul>
Theme: Test & Exam Construction & Design	Q15: How do you keep record of their performances?		
Sub-theme: Archiving	TEA1 : marks based on clear criteria should be enough. But extra notes can also be useful		
<ul style="list-style-type: none"> <li>Audio Recording</li> <li>Video</li> </ul>	<p>TEA2 : By recording their performances while they practice</p> <p>TEA3 : If a lab is available, performance should be recorded for further reference. For other types of interpretation, checklists are kept and updated</p>	<ul style="list-style-type: none"> <li>Teacher uses recordings</li> <li>Teacher does not use recordings</li> </ul>	
Theme: Test & Exam Construction & Design	Q16: What do you use for assessment?		
Sub-theme: Assess. Tool	<input type="checkbox"/> Checklists <input type="checkbox"/> Performance list <input type="checkbox"/> Rubric <input type="checkbox"/> Other (Please specify)		
<ul style="list-style-type: none"> <li>Rubrics</li> <li>Checklists</li> <li>Performance lists</li> <li>Other</li> </ul>	<p>TEA1 : Checklists and tests</p> <p>TEA2 : Checklists</p> <p>TEA3 : Checklists, Recordings</p>	<ul style="list-style-type: none"> <li>Assessment with a tool</li> <li>Assessment without a tool</li> </ul>	
Theme: Test & Exam Construction & Design	Q17: Do you consider your assessment summative or formative? Why?		
Sub-theme: Type of assess.	TEA1 : more summative than formative		
<ul style="list-style-type: none"> <li>Summative</li> <li>Formative</li> <li>Both, summative &amp; formative</li> <li>Ipsative</li> </ul>	<p>TEA2 : The assessment is summative as it focuses on the improvement achieved by the student at the end of each unit</p> <p>TEA3 : Assessments should be formative throughout the academic year. Every performance should be graded to enable students to appraise their own learning process. Summative has an extra bonus which to put students under stressful situation, which unfortunately, are the lot of interpreters.</p>	<ul style="list-style-type: none"> <li>Summative</li> <li>Formative</li> <li>Both</li> </ul>	
Theme: Alternative Forms of assess.	Q18: What are the alternative forms of assessment that you use?		
<ul style="list-style-type: none"> <li>Self-assessment</li> <li>Diary keeping</li> <li>Questionnaires</li> <li>Interviews</li> <li>Portfolios</li> </ul>	<p>TEA1 : none</p> <p>TEA2 : - Research papers on specific subjects</p> <ul style="list-style-type: none"> <li>Putting the trainees in unusual circumstances whereby they need to deal with a demanding audience and gauge their ability to successfully deal with such situations.</li> </ul>	<ul style="list-style-type: none"> <li>None</li> <li>Research papers</li> <li>Interpreting in conferences in the Silent Booth</li> <li>Performance in unusual circumstances</li> </ul>	

	TEA3: Students should be given the opportunity to attend conferences and work in a dedicated inactive booth. The teacher should then attend the meeting on a bogus basis only to listen and assess the performance of his students.		
<b>Theme: Test &amp; Exam Construction &amp; Design</b>	<b>Q19: How do you score the trainees' performance?</b>		
<b>Sub-theme: Scoring</b>	TEA1 : in accordance with the message denoted and its reception in comparison with the source text and its impact on its audiences  TEA2 : - Using a 1-10 scale that assesses their ability to work under pressure and render the meaning of the original text in a correct and proper way  TEA3 : By assessing their sub skills and grading each of them. A performance cannot totally be bad. There should be an overall score made up of sub scores related to delivery, content, voice, minor errors, major misinterpretation...	<ul style="list-style-type: none"> <li>• Impression</li> <li>• 1-10 Scale</li> </ul>	<ul style="list-style-type: none"> <li>- Holistic scoring</li> <li>- Sum of sub-scores (Analytic scoring)</li> </ul>
<b>Theme: Test &amp; Exam Construction &amp; Design</b>	<b>Q20: Does your scoring reflect specific criteria? If yes, would you specify?</b>		
<b>Sub-theme: Assessment Criteria</b>	TEA1 : yes; knowledge of language; fluency; factor of time; reception of audiences TEA2 : - Accuracy, grammatical coherence, cohesion, pleasant voice.  TEA3 : Teachers have to literally give a percentage of the message delivered. If the original message contains 5 keys ideas, their the overall grade should be split in 5 parts.	<ul style="list-style-type: none"> <li>• Mastery of languages</li> <li>• Accuracy</li> <li>• Fluency</li> <li>• Consistency</li> <li>• Impact of performance</li> <li>• Time</li> </ul>	
<b>Theme: Test &amp; Exam Construction &amp; Design</b>	<b>Q21: Do you think that your test is authentic? Why?</b>		
<b>Sub-theme: Authenticity</b>	TEA1 : Why not? It depends on the professor & aim of test  TEA2 : I believe my test is authentic as it is adapted to the specificities and shortcomings of each student. In fact, the purpose of the test is to uncover the lacking aspects in the performance of every trainee and work accordingly on bridging them.  TEA3 : I use mostly authentic texts from conferences in which I actually attended and worked as interpreter. So, I can assess bearing in mind the target audience and how your performance would come through.	<ul style="list-style-type: none"> <li>• Yes, authentic</li> </ul>	<ul style="list-style-type: none"> <li>- The professor</li> <li>- Aim of the test</li> <li>- Authentic texts</li> <li>- Adaptation to the level of the students</li> </ul>
<b>Theme: Assessment of Abilities</b>	<b>Q22: Is the training period enough to assess an ability? Why?</b>		
<b>Sub-theme: Construct(s) of abilities</b>	TEA1 : It depends on the period of training and the results obtained  TEA2 : - The theoretical training period can help in the assessment process when paired with a practical training during conferences.  TEA3 : Assessment should be done prior to accepting students in an interpretation course. Unless the minimum criteria are met, no need to go further. If any major shortcomings such as: <ul style="list-style-type: none"> <li>- Faulty grammar, poor lexis</li> <li>- Inability to self-control and remain focused</li> </ul> Students should then be made aware of their defects and be given extra time	<ul style="list-style-type: none"> <li>• Not clear</li> <li>• Conditional</li> <li>• Teachers do not refer to the training period for every pair before every exam</li> </ul>	

	before enrolling in an interpretation course.		
<b>Theme: Assessment of Abilities</b>	<b>Q23: Do you consider your test an interim or a final test? Why?</b>		
<b>Sub-theme: Exam type</b>	TEA1 : Never final, except for a definite purpose defined by the relevant institution		
<ul style="list-style-type: none"> <li>• Interim exam</li> <li>• Final exam</li> </ul>	<p>TEA2 : - My test is an interim as it prepares the trainee for real-life situations whereby they are expected to render the message regardless of how good their knowledge of subject is. The final test is the conference where they have to be on their own and work their way out successfully.</p> <p>TEA3 : Both summative and formative are needed.</p>	<ul style="list-style-type: none"> <li>• Interim, never final</li> <li>• Summative and formative</li> </ul>	
<b>Theme: Assessment of Abilities</b>	<b>Q24: What are the other criteria of professional interpreting that you don't assess? Why?</b>		
<b>Sub-theme: Other Criteria of Assessment</b>	TEA1 : If they are criteria, then they are assessed		
<ul style="list-style-type: none"> <li>• Criteria-referenced assessment</li> <li>• Criterion-referenced assessment</li> </ul>	<p>TEA2 : - The behavior of the trainee in the booth and their interaction with their colleagues during conferences.</p> <ul style="list-style-type: none"> <li>- Calm and regular delivery</li> <li>- The ability of the trainee to avoid embarrassing situations especially when dealing with jokes and cultural references.</li> </ul> <p>TEA3 : - Code of conducts that interpreters should abide by.</p> <ul style="list-style-type: none"> <li>- Stamina</li> <li>- Understanding various accents</li> </ul>	<ul style="list-style-type: none"> <li>• Booth's behaviour</li> <li>• Stress</li> <li>• Code of conduct</li> <li>• Interaction</li> <li>• Ability to deal with jokes &amp; cultural references</li> <li>• Dealing with accents</li> <li>• Stamina</li> </ul>	
<b>Theme: Assessment of Achievement Performance</b>	<b>Q25: How do you decide if someone has the competence to qualify from an interpreter education program and go out into the real world as a practitioner?</b>		
<b>Sub-theme: Exit Profile</b>	TEA1 : as stated above, tests could be enough provided that they focus on the criteria mentioned in (23) above		
<ul style="list-style-type: none"> <li>• Exit profile</li> <li>• Achievement tests</li> <li>• Certification</li> </ul>	<p>TEA2 : - Conferences are the best way to judge a trainee's progress. If he/she manages to keep calm, faithfully render the message, and meet the expectations of the audience then he/she is ready to go to the real world and work as a practitioner.</p> <p>TEA3 : If a student succeeds in summative and formative tests, he should qualify to become a junior interpreter provided he is mentored properly by a senior interpreter</p>	<ul style="list-style-type: none"> <li>• Tests</li> <li>• Performance in Conferences</li> <li>• Summative tests</li> <li>• Formative tests</li> <li>• Mentoring by a senior interpreter</li> </ul>	
<b>Theme: Certification</b>	<b>Q26: How do you verify whether a practitioner deserves certification or accreditation in his or her country of practice?</b>		
<ul style="list-style-type: none"> <li>• University qualifications</li> <li>• Associations' membership</li> <li>• Simple endorsements</li> <li>• Proficiency tests</li> </ul>	<p>TEA1 : again, the results of tests should reveal this aspect of the interpreter</p> <p>TEA2 : - By checking the requirements of interpreting in that country and making sure the qualifications of the trainee are up to date with them.</p> <ul style="list-style-type: none"> <li>- By setting comprehensive standards that apply to different countries and contexts.</li> </ul> <p>TEA3 : Clients are the final target group. So, conferees are satisfied, then one can consider himself as certified interpreter.</p> <p>Peer to peer evaluation should be key to vet each other's performance.</p> <p>Be accepted in international professional associations</p>	<ul style="list-style-type: none"> <li>• Test results</li> <li>• Interpreting requirements</li> <li>• Qualifications</li> <li>• Standards</li> <li>• Clients' satisfaction</li> <li>• Peers' opinion</li> <li>• International Associations' membership</li> </ul>	

**Table 14: Data Coding of the First-year Students' Responses (Semester 1):**

Theme(s)/Categories	Response Text for :	Axial Code	Sub-Code
Theme : Background	Q1: What is your gender?		
Sub-theme : Gender			
<ul style="list-style-type: none"> <li>Male</li> <li>Female</li> </ul>	STU1 : Female ; STU2 : female ; STU3 : Female ; STU4 : female ; STU5 : female STU6 : female ; STU7 : Female ; STU8 : Female ; STU9 : Femal. ; STU10 : Male STU11: Female ; STU12: Female ; STU13: MALE	<ul style="list-style-type: none"> <li>Male</li> <li>Female</li> </ul>	
Sub-theme: Age range	Q2: What is your age?		
<ul style="list-style-type: none"> <li>Young learner</li> <li>Old learner</li> </ul>	STU1 : 26 ; STU2 : 33 ; STU3 : 24 ; STU4 : 27 years old ; STU5 : 21 ; STU6 : 22 STU7 : 22 years old ; STU8 : 21 ; STU9 : 31 ; STU10 : 38 ; STU11: 35 ; STU12: 20 years ; STU13: 50	<ul style="list-style-type: none"> <li>20-30</li> <li>30-40</li> <li>40-50</li> </ul>	
Sub-theme : Origin	Q3 : Where are you from?		
<ul style="list-style-type: none"> <li>Algeria</li> <li>Arab Country</li> <li>Foreign Country</li> </ul>	STU1 : SETIF ; STU2 : Palestine ; STU3 : Blida ; STU4 : Msila, Algeria ; STU5 : Algeria ; STU6 : Algeria ; STU7 : Algeria ; STU8 : Algeria ; STU9 : I am from Yemen ; STU10 : Yemen ; STU11: Algeria ; STU12: Algeria ; STU13: ALGERIA	<ul style="list-style-type: none"> <li>Country of Origin</li> <li>City of Origin</li> </ul>	<ul style="list-style-type: none"> <li>- Algeria</li> <li>- Yemen</li> <li>- Palestine</li> <li>- North of Algeria</li> <li>- South of Algeria</li> </ul>
Sub-theme : Qualifications	Q4 : What is your highest degree? In what field? <input type="checkbox"/> BA <input type="checkbox"/> MA <input type="checkbox"/> Other		
<ul style="list-style-type: none"> <li>Min. Requirement</li> <li>More than the Min. Requirements</li> </ul>	STU1 : I have a licence in Management, a licence in English language (TEFL), and journalism, Italien language culture STU2 : BA / Translation STU3 : MA / English Master STU4 : BA / Translation STU5 : BA translation Arabic – English - French STU6 : BA /Translation AR-FR-EN STU7 : BA in English STU8 : BA / English linguistics STU9 : BA in Teaching English STU10 : BA, MA / English language STU11: BA / Translation and Interpreting degree STU12: BA / English "Linguistics" STU13: Other: ELECTRONIC ENGINEER	<ul style="list-style-type: none"> <li>Translation</li> <li>Translation &amp; Interpreting</li> <li>TEFL</li> <li>English Ling.</li> <li>Other degree</li> <li>Did not mention the field</li> <li>DAQ</li> </ul>	<ul style="list-style-type: none"> <li>- MA English (2)</li> </ul>
Sub-theme : Mastery of languages	Q5 : What is/are your ability(ies) in the following languages on a scale from 1 to 5? *Level 5: Excellent *Level 4: Very good *Level 3: Good *Level 2: Average *Level 1: Weak - Arabic (.....) – French (.....) – English (.....) - Other: .....(.....) ; .....(.....)		
<ul style="list-style-type: none"> <li>Medium level</li> <li>Proficient bilingual</li> <li>Proficient tilingual</li> <li>Polyglot</li> </ul>	STU1 : Arabic (3) – French (4) – English (3) – Other: Economy (4) ; Italien (2) STU2 : Arabic (5) – French (4) – English (2) STU3 : Arabic (04) – French (03) – English (03) – Other: Spanish (01) STU4 : Arabic (4) –French (4) – English (3) STU5 : Arabic (5) – French (2) – English (3) STU6 : Arabic (5) – French (5) – English (3)	<ul style="list-style-type: none"> <li>The general level corresponds to the entrance level</li> <li>STU needs remedial courses</li> <li>STU is weak in one of the languages</li> <li>STU masters other languages</li> </ul>	

	<p>STU7 : Arabic (3) – French (3) – English (3) – Other: Spanish (average)  STU8 : Arabic (5) – French (2) – English (5)  STU9 : ///  STU10 : Arabic (Excellent) – French (Weak) – English (V. Good)  STU11: Arabic (4) – French (4) – English (3) – Other: Spanish (2)  STU12: Arabic (3) – French (3) – English (4) – Other: Spanish  STU13: Arabic (2) – French (4) – English (3)</p>	<ul style="list-style-type: none"> <li>• DAQ</li> </ul>	
Theme: Perceptions about the Admission Test	<p>Q6 : How did you find the Admission Test (Entrance Test /The Aptitude Test)? Tick one box only.  <input type="checkbox"/> Very easy   <input type="checkbox"/> Easy   <input type="checkbox"/> A little bit difficult   <input type="checkbox"/> Difficult   <input type="checkbox"/> Very difficult</p>		
Sub-theme: Easiness of the Admission Test	<p>Say why: .....</p>		
<ul style="list-style-type: none"> <li>• Very Easy</li> <li>• Easy</li> <li>• A little Difficult</li> <li>• Difficult</li> <li>• Very Difficult</li> </ul>	<p>STU1 : Very easy. To give chance to the student or any student to have the opportunity to study if he is volunteer.  STU2 : A little bit difficult  STU3 : Very easy. The test covered some usual questions that's why it could be considered more than easy.  STU4 : Very easy. The questions were very clear, direct questions.  STU5 : Easy. It was acceptable and anyone who has good or acceptable level in Eng can find it easy. ; STU6 : /// ; STU7 : Easy ; ///  STU8 : I didn't take the test because I came late  STU9 : /// ; STU10 : A little bit difficult. There was no preparation.  STU11: Easy ; /// ; STU12: I did not take the test because I came late.  STU13: Easy ; ///</p>	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> <li>• AJ</li>   <li>• Reasons for the easiness of the test</li> </ul>	<ul style="list-style-type: none"> <li>- Admitted without being tested</li> <li>- STU finds the test very easy</li> <li>- STU finds the test easy</li> <li>- STU finds the test a little difficult</li>   <li>- Testers provide opportunities to the learners</li> <li>- Usual, direct questions</li> <li>- Easiness of the language</li> </ul>
Sub-theme: Duration of the Admission Test	<p>Q7 : How long did the test last ?  <input type="checkbox"/> 15 mn   <input type="checkbox"/> 30mn   <input type="checkbox"/> 45mn   <input type="checkbox"/> More :.....</p>		
<ul style="list-style-type: none"> <li>• 15mn</li> <li>• 30mn</li> <li>• 45mn</li> <li>• More than 45mn</li> </ul>	<p>STU1 : 15 mn ; STU2 : 30 mn ; STU3 : 15 mn ; STU4 : 15 mn ; STU5 : 15 mn  STU6 : /// ; STU7 : /// ; STU8 : /// ; STU9 : ///  STU10 : /// ; STU11: 15 mn ; STU12: /// ; STU13: 15 mn</p>	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• AJ</li> </ul>	
Sub-theme: Satisfaction with the duration of the Admission Test	<p>Q8 : Was this duration sufficient to you?  <input type="checkbox"/> Yes   <input type="checkbox"/> No</p>		
<ul style="list-style-type: none"> <li>• Yes, the duration was sufficient</li> <li>• No, it was not sufficient</li> </ul>	<p>Say why:.....  STU1 : Yes. It was oral very clear questions that evaluate the culture of the student more than his ability in that languages performances.  STU2 : Yes. I preferred if they give us marks not only giving the level.  STU3 : Yes, because the questions were not difficult and less than 15 mn could be sufficient also  STU4 : Yes. Because the questions were very simple, there was no need to a longer duration.  STU5 : Yes ; /// ; STU6 : /// ; STU7 : Yes ; ///  STU8 : /// ; STU9 : /// ; STU10 : ///  STU11: No. I would like to talk about translation courses more deeply. And also to talk more about what I want to learn from ISAT.  STU12: /// ; STU13: Yes , ///</p>	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> <li>• AJ (Yes)</li>   <li>• AJ (No)</li> </ul>	<ul style="list-style-type: none"> <li>- The Qs evaluate culture</li> <li>- The Qs were not difficult</li> </ul>
Sub-theme: Information before sitting for the Admission Test	<p>Q9: Did you have any idea about the test before taking it?</p>		
<ul style="list-style-type: none"> <li>• Yes, I had an idea</li> <li>• No, I did not have any idea</li> </ul>	<p>STU1 : No idea but if I'e asked someone studied before here is the same test that they repeat it. ; STU2 : No ; STU3 : No ; STU4 : No, I haven't  STU5 : No, I just heard that it will be oral.</p>	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> <li>• AJ</li> </ul>	

	STU6 : //////////////// ; STU7 : No ; STU8 : I didn't have any idea about the test STU9 : // ; St10 : // ; STU11 : No, I did not ; STU12 : No idea ; STU13 : No		- I heard it will be oral
Sub-theme: Sharing of Assessment Criteria	Q10 : Were you informed about the criteria for a good performance? <input type="checkbox"/> Yes <input type="checkbox"/> No		
<ul style="list-style-type: none"> <li>Informed Sts about the criteria</li> <li>Uninformed Sts about the criteria</li> </ul>	STU1 : Yes ; St2 : No ; STU3 : No STU4 : //////////////// ; STU5 : Yes ; STU6 : No STU7 : //////////////// ; STU8 : //////////////// ; STU9 : //////////////// STU10 : //////////////// ; STU11 : No ; STU12 : //////////////// ; STU13 : No	<ul style="list-style-type: none"> <li>DAQ</li> <li>AJ</li> </ul>	<ul style="list-style-type: none"> <li>No, I was not informed</li> <li>Yes, I was informed</li> </ul>
Sub-theme: Lists of Assessment Criteria	Q11 : If your answer to Q10 is yes, would you list these criteria?		
<ul style="list-style-type: none"> <li>Basic criteria</li> <li>Advanced-level criteria</li> </ul>	STU1 : Yes. The criteria of a good performance is the acquisition of a language if me may docteur that language must* STU2 : //////////////// ; STU3 : //////////////// ; STU4 : //////////////// STU5 : that I must speak fluently with self confidence. STU6 : //////////////// ; STU7 : //////////////// ; STU8 : //////////////// ; STU9 : //////////////// STU10 : //////////////// ; STU11 : //////////////// ; STU12 : //////////////// ; STU13 : ////////////////	<ul style="list-style-type: none"> <li>DAQ</li> <li>AJ</li> </ul>	<ul style="list-style-type: none"> <li>Lge Acquisition</li> <li>Fluency</li> <li>Self-confidence</li> </ul>
Sub-theme: Preparedness for the Admission Test	Q12 : Did you prepare for the test? <input type="checkbox"/> Yes <input type="checkbox"/> No If yes, what did you do?.....		
<ul style="list-style-type: none"> <li>Prepared Sts for the test</li> <li>Unprepared Sts for the test</li> </ul>	STU1 : Yes. I revised all thing in English what I had at university but I was very Shemel* for arabic. STU2 : No ; STU3 : No ; STU4 : No ; STU5 : No ; STU6 : No STU7 : // ; STU8 : // ; STU9 : // ; STU10 : // ; STU11 : No STU12 : //////////////// STU13 : Yes. Reading – Grammatical rules	<ul style="list-style-type: none"> <li>DAQ</li> <li>ADJ</li> <li>AJ</li> </ul>	<ul style="list-style-type: none"> <li>Revision of University lectures</li> <li>Reading</li> <li>Grammar</li> </ul>
Sub-theme: Tasks of the Admission Test	Q13 : What tasks/exercises were involved in this test? Please, describe them accurately. <input type="checkbox"/> Listening:..... <input type="checkbox"/> Reading:..... <input type="checkbox"/> Speaking:..... <input type="checkbox"/> Writing:..... <input type="checkbox"/> Vocabulary: ..... <input type="checkbox"/> Grammar:..... <input type="checkbox"/> Pronunciation: ..... <input type="checkbox"/> Translating:..... <input type="checkbox"/> Interpreting:..... <input type="checkbox"/> Interview..... <input type="checkbox"/> Other (Please specify): .....		
<ul style="list-style-type: none"> <li>Listening</li> <li>Reading</li> <li>Speaking</li> <li>Writing</li> <li>Vocabulary</li> <li>Grammar</li> <li>Pronunciation</li> <li>Translating</li> <li>Interpreting</li> <li>Interview</li> <li>Other tasks</li> </ul>	STU1 : Listening: when teachers were asking me by using high style of language. Speaking: They were controlling my confidence in my sens and language Vocabulary: They asked me if I know a lot of that languages I know. Pronunciation: They insist that I used English (British American) Translating: The teachers were asking me a same question in different languages. Interview: Yes it was like an interview with 4 teachers. Other: They insist on my culture background even history, communication STU2 : Listening, Reading, Speaking, Writing, Vocabulary, Grammar STU3 : Listening: I needed to listen to the teacher's questions to answer. Speaking: I was required to speak using three languages. Vocabulary: I had to use my vocabulary to express my ideas.	<ul style="list-style-type: none"> <li>DAQ</li> <li>ADJ</li> <li>AJ</li> </ul>	<ul style="list-style-type: none"> <li>STUs interpreting the purpose of the interview</li> <li>Interview</li> </ul>

	<p>Pronunciation: I had to pronounce correctly to prove myself.  Interview: I needed to communicate with different teachers differently.  STU4 : Interview ; STU5 : Speaking: I was asked many questions concern languages  Translating: I was asked to translate my name  STU6 : ///; STU7 : //////////////; STU8 : //////////////; STU9 : //////////////; STU10 : //////////////  STU11: Interview ; STU12: //////////////  STU13: Listening, Speaking, Pronunciation, Interview</p>		
Sub-theme : Instructions of the Admission Test	Q14 : What directives/instructions/questions were given by the tester(s) during the test?		
<ul style="list-style-type: none"> <li>Directives</li> <li>Instructions</li> <li>Questions</li> </ul>	<p>STU1 : Arabic: Present yourself, if you know all this languages did you travel to those countries.  French: as you studied in Farhat Abbas University who is he?  English: How do you find English as a language?  Culture: What do you know about FIFO and LIFO  STU2 : NA  STU3 : What's your name? Where are you from? Why did you choose this institut?  What do you do in your life and where did you study before?  STU4 : There was questions about our degree, about our choice, and what we are willing to do with this Master.  STU5 : - Why did you choose our institute?  -What are translation obstacles?  STU6 : //////////////; STU7 : //////////////; STU8 : //////////////; STU9 : //////////////; STU10 : //////////////  STU11: A personal question first – name – working experience and why have we chose to study at ISAT.  STU12: //////////////  STU13: NA (Not applicable). We were not given any directives or instructions.</p>	<ul style="list-style-type: none"> <li>DAQ</li> <li>ADJ</li> <li>AJ</li> </ul>	<p>-NA</p> <ul style="list-style-type: none"> <li>- Introducing</li> <li>- Reasons for choosing the HAIT</li> <li>- Introducing</li> <li>- Qs on Local Culture</li> <li>- Qs on Business knowledge</li> <li>- Qs about translation-related knowledge</li> <li>- Qs about future projects</li> </ul>
Sub-theme: Languages used in the Admission Test	Q15 : In what language was the test conducted?		
<ul style="list-style-type: none"> <li>Arabic only</li> <li>French only</li> <li>English only</li> <li>Arabic &amp; French</li> <li>Arabic &amp; English</li> <li>More than two languages</li> </ul>	<p>STU1 : In three languages (Arabic – French – English)  STU2 : English ; STU3 : the test was conducted in English – French and Arabic.  STU4 : (3) three languages.  STU5 : the 3 languages. Arabic – French - English  STU6 : //////////////; STU7 : //////////////; STU8 : //////////////; STU9 : //////////////; STU10 : //////////////  STU11: 3 languages – Arabic – French - English  STU12: //////////////; STU13: Arabic, French, &amp; English interviews</p>	<ul style="list-style-type: none"> <li>DAQ</li> <li>ADJ</li> <li>AJ</li> </ul>	
Sub-theme: Sharing of Assessment Criteria	Q16 : Did the tester(s) share their assessment criteria with you before you answered the test? <input type="checkbox"/> Yes <input type="checkbox"/> No		
<ul style="list-style-type: none"> <li>Yes, the tester(s) shared their assessment criteria with me</li> <li>No, the tester(s) did not share their assessment criteria</li> </ul>	<p>STU1 : No ; STU2 : No ; STU3 : No ; STU4 : No ; STU5 : No  STU6 : // ; STU7 : //////////////; STU8 : //////////////; STU9 : //////////////; STU10 : //////////////  STU11: No ; STU12: ////////////// STU13: No</p>	<ul style="list-style-type: none"> <li>DAQ</li> <li>ADJ</li> </ul>	
Sub-theme : Lists of Assessment Criteria	Q17 : If your answer to Q16 is yes, would you list these criteria?		
<ul style="list-style-type: none"> <li>Beginner level criteria</li> <li>Advanced level criteria</li> </ul>	<p>STU1 : //////////////; STU2 : //////////////; STU3 : //////////////  STU4 : //////////////; STU5 : //////////////; STU6 : //////////////  STU7 : //////////////; STU8 : //////////////; STU9 : //////////////  STU10 : //////////////; STU11: //////////////; STU12: //////////////  STU13: //////////////</p>	<ul style="list-style-type: none"> <li>DAQ</li> </ul>	<ul style="list-style-type: none"> <li>- Refused to justify</li> <li>- Did not have a view or an idea</li> </ul>



Theme : Intentions of specialization	<b>Q23 : What do you intend to study in the second semester?</b> <input type="checkbox"/> Written Translation <input type="checkbox"/> Interpreting <input type="checkbox"/> Translation Technology		
<ul style="list-style-type: none"> <li>• Written translation</li> <li>• Interpreting</li> <li>• Translation Technology</li> </ul>	STU1 : Interpreting ; STU2 : Written Translation ; STU3 : Interpreting STU4 : Written Translation ; STU5 : Interpreting ; STU6 : Interpreting STU7 : /////////////// ; STU8 : Written Translation ; STU9 : Written Translation STU10 : Interpreting ; STU11 : Interpreting ; STU12 : Written Translation STU13 : Written Translation	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>- Interpreting</li> <li>- Written Translation</li> </ul>
Theme : Students' Opinion	<b>Q24 : Would you please add any comments or remarks that you think might be useful to this research? Thank you.</b>		
<ul style="list-style-type: none"> <li>• Comments</li> <li>• Remarks</li> <li>• Suggestions</li> </ul>	STU1 : I think that that kind of test might be very easy or they omit it as in developed countries. STU2 : Yes this research is very important to ameliorate the conditions of the test. Levels of students are very different. STU3 : I wish a good luck for you dear researcher STU4 : the test was very short, there was 03 teachers for each language, I want to add a Remarque in French [l'évaluation n'était pas objective]. STU5 : the test should be more serious and formal. STU6 : // STU7 : I didn't take the test / It's better to have a test STU8 : // STU9 : No, I think it's enough. However, I hope to have some practice in order to make us good in translation. STU10 : Certain measures for testing language fluency & accuracy here to be taken before admitting a student in different divisions of the translation. There has to be a test to interpreter: Fluency in spoken English or French and writing accuracy test to be taken in consideration in written translation. STU11 : - It is better to make -writing - listening and speaking tests. - Evaluate us more in translation. STU12 : // STU13 : I have been submitted to an oral exam only. A written examination would have been welcome.	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>- Omission of the test</li> <li>- Heterogeneity of STUs level</li> <li>- Wishes</li> <li>- Objectivity</li> <li>- Formality</li> <li>- Compulsoriness of the test</li> <li>- Practice</li> <li>- Test type and tasks</li> </ul>

**Table 15: Data Coding of the First-year Students' Responses (Semester 2):**

Theme(s)/Categories	Response Text for :	Axial Code	Sub-Code
Theme : Background	Q1: What is your gender?		
Sub-theme: Gender	STU14 : femal STU15 : female STU16 : Female	• Female	
• Male • Female			
Sub-theme : Age range	Q2: What is your age?		
• Young learner • Old learner	STU14 : 23 STU15 : 24 STU16 : 25	• 20-30	
Sub-theme : Origin	Q3 : Where are you from?		
• Algeria • Arab Country • Foreign Country	STU14 : Algeria /Constine STU15 : Algiers STU16 : Algiers	• Country of Origin • City of Origin	- Algeria - Capital of Algeria - East of Algeria
Sub-theme : Qualifications	Q4 : What is your highest degree? In what field? <input type="checkbox"/> BA <input type="checkbox"/> MA <input type="checkbox"/> Other		
• Min. Requirement • More than the Min. Requirements	STU14: BA/Translation STU15 : BA in translation STU16 : BA ; //////////////////////////////////////	• St did not specify the field	
Sub-theme : Mastery of Languages	Q5 : What is/are your ability(ies) in the following languages on a scale from 1 to 5? *Level 5: Excellent   *Level 4: Very good   *Level 3: Good *Level 2: Average   *Level 1: Weak - Arabic (.....) – French (.....) – English (.....) - Other: .....(.....) ; .....(.....)		
• Medium level • Proficient bilingual • Proficient tilingual • Polyglot	STU14 : Arabic (good) – French (Very good) – English (.../....) STU15 : Arabic (3) – French (3) – English (3) STU16 : Arabic (3) – French (3) – English (3)	• STU did not specify the level of lge mastery • STU needs remedial courses	
Sub-theme: Scoring in the Admission Test	Q6 : What were your score and rank among those who passed the Admission Test (i.e. the Entrance Test) of The Higher Arab Institute of Translation (HAIT)? Score:..... Rank:.....		
• Score • Rank	STU14 : Score: ..../..... ; Rank: ...../ .....	• No score • No rank	
STU15 : no score ; no rank STU16 : No score ; No rank			
Theme: Average Score in SI	Q7 : What was your average score in the first semester (SI)?		
• Under average • Above average	STU14: ////////////////////////////////////// STU15 ://////////////////////////////////// STU16 : No score	• DAQ • No score	STU did not understand the Q
Theme : Subjective objectives of training	Q8 : Why did you choose Interpreting?		
• Research purposes • Professional purposes	STU14 : the aim was to get a higher level STU15 : because it's among the best works and also it's a well payed job STU16 : I prefer interpreting more than translating.	• Getting a higher level • A good job • I like it	
Theme : Perceptions about the	Q9 : How did you find the Interpreter Aptitude Test (IAT) of the HAIT? Tick one box only.		



<ul style="list-style-type: none"> <li>• Interpretinf daily life dialogues</li> <li>• Other</li> </ul>			
Sub-theme: Satisfaction with the specific training period	Q16 : If your answer to Q14 is yes, was the course/training period sufficient? <input type="checkbox"/> Yes <input type="checkbox"/> No Say why: .....		
<ul style="list-style-type: none"> <li>• Yes, the trainig period was sufficient</li> <li>• No, It was not.</li> </ul>	STUI4 : ////////////////////////////////////// STUI5 : ////////////////////////////////////// STUI6 : //////////////////////////////////////	<ul style="list-style-type: none"> <li>• DAQ</li> </ul>	<ul style="list-style-type: none"> <li>- Sts did not have a specific training period</li> </ul>
Theme : IAT Construction & Design	Q17 : How long did the whole Interpreter Aptitude Test (IAT) last? <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> More .....		
Sub-theme : Duration of the IAT			
<ul style="list-style-type: none"> <li>• 10mn</li> <li>• 15mn</li> <li>• 30mn</li> <li>• 45mn</li> <li>• More than 45mn</li> </ul>	STUI4 : 10mn STUI5 : 10mn STUI6 : 10mn	<ul style="list-style-type: none"> <li>• 10 mn for all the Sts</li> </ul>	
Sub-theme: Satisfaction with the duration of the IAT	Q18 : Was this duration sufficient to you? <input type="checkbox"/> Yes <input type="checkbox"/> No Say why:.....		
<ul style="list-style-type: none"> <li>• Yes, the duration of the IAT was sufficient</li> <li>• No, It was not</li> </ul>	STUI4 : Yes, I judged so. STUI5 : No, because i might better in another topic STUI6 : Yes, because we were biginner	<ul style="list-style-type: none"> <li>• Yes, sufficient</li> <li>• No, not sufficient</li> </ul>	<ul style="list-style-type: none"> <li>- Unclear reasons</li> <li>- Beginner</li> <li>- I might do better</li> </ul>
Sub-theme : Tasks of the IAT	Q19 : What tasks/exercises were involved in the Interpreter Aptitude Test (IAT)? Would you please tick the appropriate box and describe the tasks accurately. <input type="checkbox"/> Shadowing:..... <input type="checkbox"/> Community Interpreting:..... <input type="checkbox"/> Simultaneous Interpreting:..... <input type="checkbox"/> Consecutive Interpreting:..... <input type="checkbox"/> Sight translation/interpreting:..... <input type="checkbox"/> Other:.....		
<ul style="list-style-type: none"> <li>• Shadowing</li> <li>• Community Interpreting</li> <li>• SI</li> <li>• CI</li> <li>• STI</li> <li>• Other</li> </ul>	STUI4 : Simultaneous Interpreting: It was a political speech STUI5 : Simultaneous Interpreting STUI6 : Simultaneous Interpreting ; Consecutive Interpreting ; Sight translation/interpreting	<ul style="list-style-type: none"> <li>• Tasks for the French Group</li> <li>• Tasks for the English Group</li> </ul>	<ul style="list-style-type: none"> <li>- SI of a Political Content</li> <li>- CI</li> <li>- Sight Transl.</li> <li>- SI</li> </ul>
Sub-theme : Combinations of Interpreting in the IAT	Q20 : What were the combinations of interpreting? <input type="checkbox"/> A into B <input type="checkbox"/> A into C <input type="checkbox"/> B into A <input type="checkbox"/> C into A <input type="checkbox"/> B into C <input type="checkbox"/> C into B		
<ul style="list-style-type: none"> <li>• A into B</li> <li>• A into C</li> <li>• B into A</li> <li>• C into A</li> <li>• B into C</li> <li>• C into B</li> </ul>	STUI4 : A into B ; B into A (French) STUI5 : B into A STUI6 : A into B ; B into A (English)	<ul style="list-style-type: none"> <li>• AC ; CA (Stl)</li> <li>• BA</li> <li>• AB;BA</li> </ul>	<ul style="list-style-type: none"> <li>- Francophone</li> <li>- Anglophones</li> </ul>
Sub-theme : The Length of the prompt(s) in the IAT	Q21 : In the task(s) of “Interpreting”, what was/were the duration(s) of the prompt(s) for the Interpreter Aptitude Test? Prompt l: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:.....		

	Prompt 2: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... Prompt 3: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... More prompts? .....		
<ul style="list-style-type: none"> <li>Prompt 1</li> <li>Prompt 2</li> <li>Prompt 3</li> <li>More prompts</li> </ul>	STU14 : Prompt 1: 5mn ; 10 mn STU15 : Prompt 1: 10mn STU16 : Prompt 1: 5mn ; 10mn	<ul style="list-style-type: none"> <li>5mn, 10mn</li> <li>10 mn</li> </ul>	
Sub-theme : Number of takes	Q22 : Did you answer in one take? <input type="checkbox"/> Yes <input type="checkbox"/> No, in .....		
<ul style="list-style-type: none"> <li>Yes, in one take</li> <li>No, in ...</li> </ul>	STU14 : Yes STU15 : Yes STU16 : Yes	<ul style="list-style-type: none"> <li>Sts answered in one take</li> </ul>	
Sub-theme: Negotiation of the prompts' length	Q23 : Did you negotiate the prompts' length (duration) with the teacher? <input type="checkbox"/> Yes <input type="checkbox"/> No		
<ul style="list-style-type: none"> <li>Yes, we negotiated with teacher</li> <li>No, we did not negotiate with the teacher</li> </ul>	STU14 : No STU15 : No STU16 : No	<ul style="list-style-type: none"> <li>No negotiation</li> </ul>	
Sub-theme : Tasks of the IAT	Q24 : What other tasks/exercises were involved in the Interpreter Aptitude Test (IAT)? Would you please tick the appropriate box and describe them accurately. <input type="checkbox"/> Listening Comprehension..... <input type="checkbox"/> Reading Comprehension..... <input type="checkbox"/> Speaking:..... <input type="checkbox"/> Writing: <input type="checkbox"/> Re-writing <input type="checkbox"/> Other: ..... <input type="checkbox"/> Vocabulary: <input type="checkbox"/> Word Recognition <input type="checkbox"/> Synonyms <input type="checkbox"/> Other:..... <input type="checkbox"/> Grammar:..... <input type="checkbox"/> Pronunciation: ..... <input type="checkbox"/> Translating:..... <input type="checkbox"/> Interview..... <input type="checkbox"/> Summarization skills:..... <input type="checkbox"/> Cloze test:..... <input type="checkbox"/> Memory test:..... <input type="checkbox"/> Nufferno test: (Speed-Stress)..... <input type="checkbox"/> Other (Please specify): .....		
<ul style="list-style-type: none"> <li>Listening</li> <li>Reading</li> <li>Speaking</li> <li>Writing</li> <li>Vocabulary</li> <li>Grammar</li> <li>Pronunciation</li> <li>Translating</li> <li>Interview</li> <li>Summerization</li> <li>Cloze test</li> <li>Memory test</li> <li>Nufferno test</li> <li>Other tasks</li> </ul>	STU14 : /////////////////////////////////////// STU15 : Speaking ; Translating ; Memory test STU16 : Speaking ; Translating ; Memory test	<ul style="list-style-type: none"> <li>DAQ</li> <li>Sts tried to explain the activities involved in the 'task' to answer the Q</li> </ul>	<ul style="list-style-type: none"> <li>Speaking, Translating, Memory test</li> </ul>
Sub-theme : Instructions of the IAT	Q25 : What directives/instructions/questions were given by the tester(s) during the Interpreter Aptitude Test (IAT)?		
<ul style="list-style-type: none"> <li>Directives</li> <li>Instructions</li> <li>Questions</li> </ul>	STU14 : to be calm STU15 : Weather we were following the current events. Do a lot of training. STU16 : Reading newspaper – following the current events – training more	<ul style="list-style-type: none"> <li>Instructions</li> <li>Questions</li> </ul>	<ul style="list-style-type: none"> <li>Calmness</li> <li>Events &amp; News knowledge</li> <li>Newspaper reading</li> <li>More training</li> </ul>
Sub-theme : Languages used in the IAT	Q26 : In what language was the Interpreter Aptitude Test conducted?		
<ul style="list-style-type: none"> <li>Arabic only</li> <li>Arabic &amp; French</li> <li>Arabic &amp; English</li> <li>More than two languages</li> </ul>	STU14 : French for francophones STU15 : Arabic STU16 : English into Arabic	<ul style="list-style-type: none"> <li>French</li> <li>Arabic</li> <li>English, Arabic</li> </ul>	<ul style="list-style-type: none"> <li>St3 confuses between the language of the test and that of the prompts</li> </ul>

Sub-theme: Sharing of Assessment Criteria	Q27 : Did the tester(s) share their assessment criteria with you before you answered the Interpreter Aptitude Test? <input type="checkbox"/> Yes <input type="checkbox"/> No		
<ul style="list-style-type: none"> <li>• Yes, the tester(s) shared their assessment criteria</li> <li>• No, they did not share their assess. criteria</li> </ul>	STU14 : Yes STU15 : No STU16 : No	<ul style="list-style-type: none"> <li>• Yes, there was sharing</li> <li>• No, there was no sharing</li> </ul>	
Sub-theme : List of assessment Criteria	Q28 : If your answer to Q27 is yes, would you list these criteria?		
<ul style="list-style-type: none"> <li>• Basic criteria</li> <li>• Advanced level criteria</li> </ul>	STU14 : to have a perfect comprehension, to have a vast culture STU15 : /////////////////////////////////// STU16 : //////////////////////////////////	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>- Perfect Comprehens.</li> <li>- Vast Culture</li> </ul>
Sub-theme: Meaningfulness of Assessment Criteria	Q29 : Are these criteria meaningful/important for you? <input type="checkbox"/> Yes <input type="checkbox"/> No Say in what way:.....		
<ul style="list-style-type: none"> <li>• Yes, the criteria are meaningful</li> <li>• No, they are not meaningful</li> </ul>	STU14: Yes, because that what important for an interpreter STU15 : Yes, in order to be evaluated STU16 : Yes. This criteria help us to be a good interpreter	<ul style="list-style-type: none"> <li>• The criteria are meaningful</li> </ul>	<ul style="list-style-type: none"> <li>- Evaluation</li> <li>- Profile</li> </ul>
Sub-theme: Difficult part in the IAT	Q30 : What part of the test was the most difficult to answer? Why?		
<ul style="list-style-type: none"> <li>• Interpreting from A into B</li> </ul>	STU14 : there was no questions STU15 : It was easy to answer STU16 : The 1 <sup>st</sup> part where we started the interpretation	<ul style="list-style-type: none"> <li>• The beginning</li> <li>• None</li> </ul>	<ul style="list-style-type: none"> <li>- No questions asked about the interpreting task e.g. Questions about the content</li> </ul>
Sub-theme: Easiest part in the IAT	Q31 : What part of the test was the least difficult to answer? Why?		
<ul style="list-style-type: none"> <li>• Interpreting from B into A</li> </ul>	STU14 : there was no questions STU15 : almost all STU16 : After being familiar with the speech, we could render it	<ul style="list-style-type: none"> <li>• All the parts</li> <li>• The end</li> </ul>	<ul style="list-style-type: none"> <li>• No questions asked about the interpreting task</li> </ul>
Sub-theme : Scoring in the IAT	Q32 : What were your score and your rank among those who passed the Interpreter Aptitude Test (IAT)? Score:..... Rank:.....		
<ul style="list-style-type: none"> <li>• Score</li> <li>• Rank</li> </ul>	STU14 : /////////////////////////////////// STU15 : No score ; No rank STU16 : No score ; No rank	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>- No score</li> <li>- No rank</li> </ul>
Theme: Satisfaction with the IAT results	Q33 : Were you satisfied with the results you obtained? <input type="checkbox"/> Yes <input type="checkbox"/> No Say why: .....		
<ul style="list-style-type: none"> <li>• Satisfied learners</li> <li>• Unsatisfied learners</li> </ul>	STU14 : /////////////////////////////////// STU15 : No results STU16 : //////////////////////////////////	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>- Sts understand that results mean marks and ranks</li> <li>- No results were displayed</li> </ul>
Theme : Students' Opinion	Q34 : Would you please add any comments or remarks that you think might be useful to this research? Thank you.		
<ul style="list-style-type: none"> <li>• Comments</li> <li>• Remarks</li> <li>• Suggestions</li> </ul>	STU14 : No remarks STU15 : As far as I'm concerned, it was a wise questionnaire on which you have delt with the most important points. STU16 : You have focused on the essentiels points, and I think you should add some questions about the consecutive interpretation because it is the basis of simultaneous one. Thank you.	<ul style="list-style-type: none"> <li>• No remarks</li> <li>• Appraisal</li> <li>• Suggestion</li> </ul>	<ul style="list-style-type: none"> <li>- To add questions on CI</li> </ul>

**Table 16: Data Coding of the Second-year Interpreter Students' Responses (Semester 3 + Semester 4):**

Theme(s)/ Categories	Response Text for :	Axial Code	Sub-Code
Theme : Background	Q1: What is your gender?		
Sub-theme: Gender			
<ul style="list-style-type: none"> <li>Male</li> <li>Female</li> </ul>	STUI7 : female STUI8 : Female STUI9 : Male		
Sub-theme: Age range	Q2: What is your age?		
<ul style="list-style-type: none"> <li>Young learner</li> <li>Old learner</li> </ul>	STUI7 : 24 STUI8 : 24 years old STUI9 : 24 years old.	<ul style="list-style-type: none"> <li>20-30</li> </ul>	- 24 (3)
Sub-theme : Origin	Q3 : Where are you from?		
<ul style="list-style-type: none"> <li>Algeria</li> <li>Arab Country</li> <li>Foreign Country</li> </ul>	STUI7 : Algiers STUI8 : Algeria STUI9 : Algeria.	<ul style="list-style-type: none"> <li>Country of Origin</li> <li>City of Origin</li> </ul>	<ul style="list-style-type: none"> <li>- Algeria</li> <li>- Algiers</li> </ul>
Sub-theme : Qualifications	Q4 : What is your highest degree? In what field? -BA - MA - Other		
<ul style="list-style-type: none"> <li>Min. Requirement</li> <li>More than the Min.</li> </ul>	STUI7 : MA/Interpretation* STUI8 : MA, ////////////////* STUI9 : MA in translation and conference interpreting .	<ul style="list-style-type: none"> <li>MA Interpreting</li> <li>MA Transl. &amp; Interp.</li> <li>MA</li> </ul>	- Did not mention the field
Sub-theme: Mastery of languages	Q5 : What is/are your ability(ies) in the following languages on a scale from 1 to 5? *Level 5: Excellent *Level 4: Very good *Level 3: Good *Level 2: Average *Level 1: weak - Arabic (.....) - French (.....) - English (.....) Other: .....(.....) ; .....(.....)		
<ul style="list-style-type: none"> <li>Medium level</li> <li>Proficient bilingual</li> <li>Proficient trilingual</li> <li>Polyglot</li> </ul>	STUI7 : Arabic (3) – French (4) – English (3) STUI8 : Arabic (good) – French (good) – English (good) St19 : Arabic (5) – French( 4) – English (4)	<ul style="list-style-type: none"> <li>The general level corresponds to the entrance level</li> <li>St needs remedial courses</li> </ul>	
Sub-theme: Working languages	Q6 : Is English your B or C language? <input type="checkbox"/> B <input type="checkbox"/> C		
<ul style="list-style-type: none"> <li>English is my B</li> <li>English is my C</li> </ul>	STUI7 : C STUI8 : B STUI9 : C	<ul style="list-style-type: none"> <li>English as B (1)</li> <li>English as C (2)</li> </ul>	<ul style="list-style-type: none"> <li>- Anglophone group</li> <li>- Francophone group</li> </ul>
Theme : Mark in SI in (S2)	Q7 : What was your mark in Simultaneous Interpreting (SI) Exam (From Arabic/A into English) in the second semester (S2) of the 1 <sup>st</sup> year?		
<ul style="list-style-type: none"> <li>Under average</li> <li>Above average</li> </ul>	STU1 :13 from English to Arabic STU2 :14 STU3 :15.5	<ul style="list-style-type: none"> <li>St specified the pair</li> <li>St did not specify the pair</li> </ul>	- CA
Theme : Perceptions about SI Exam in (S3)	Q8 : How did you find the Simultaneous Interpreting (SI) Exam (From A into English) of the third semester (S3) (2 <sup>nd</sup> Year)? <input type="checkbox"/> Very Easy <input type="checkbox"/> Easy <input type="checkbox"/> A little bit Difficult <input type="checkbox"/> Difficult <input type="checkbox"/> Very Difficult		
Sub-theme: Easiness of SI Exam in S3	Say why:.....		
<ul style="list-style-type: none"> <li>Very Easy</li> </ul>	STUI7 : A little bit difficult. The topics were more intricate, and because the	<ul style="list-style-type: none"> <li>Intricate topics</li> </ul>	

<ul style="list-style-type: none"> <li>• Easy</li> <li>• A little Difficult</li> <li>• Difficult</li> <li>• Very Difficult</li> </ul>	<p>vocabulary in the topics were more difficult.  STUI8 : A little bit difficult, couldn't keep pace with the speaker  STUI9 : Easy. Because I have a good ability of understanding .</p>	<ul style="list-style-type: none"> <li>• Difficult vocabulary</li> <li>• High rate of Delivery</li> <li>• Comprehension ability</li> </ul>	
<b>Theme: Perceptions about SI Exam in (S4)</b> <b>Q9 : How did you find the SI exam (From A into English) of the fourth semester (S4) (2<sup>nd</sup> Year)?</b> <input type="checkbox"/> Very Easy <input type="checkbox"/> Easy <input type="checkbox"/> A little bit Difficult <input type="checkbox"/> Difficult <input type="checkbox"/> Very Difficult Say why: .....			
<b>Sub-theme: Easiness of SI Exam in S4</b> <ul style="list-style-type: none"> <li>• Very Easy</li> <li>• Easy</li> <li>• A little Difficult</li> <li>• Difficult</li> <li>• Very Difficult</li> </ul>	STUI7 : A little bit difficult. The same reason mentioned in the previous question. STUI8 : Easy. We earned more vocab in different topics and fields mainly in economy and politics. STUI9 : A little bit difficult. The topic was a bit difficult	<ul style="list-style-type: none"> <li>• Intricate topics</li> <li>• Difficult vocabulary</li> <li>• Familiarity with vocabulary</li> </ul>	
<b>Theme: Difference between First-year level &amp; Second-year level Exams</b> <b>Q10 : Were the S3 and S4 exams in SI (from A into English) (the 2<sup>nd</sup> year) different from the S2 exam in SI (from A into English) (1<sup>st</sup> year)?</b> <input type="checkbox"/> Yes <input type="checkbox"/> No Say in what way: .....			
<ul style="list-style-type: none"> <li>• Yes, S3 and S4 exams were different from S2 exam in SI</li> <li>• No, they were not.</li> </ul>	STUI7 : Yes, we were accustomed more with the module. STUI8 : Yes, we get more used to interpret simultaneously, so the exam sounds easier. STUI9 : Yes, mainly in everything, especially in my command of languages which was clearly better due to intensive courses.	<ul style="list-style-type: none"> <li>• Familiarity with the module</li> <li>• Improved command of language</li> </ul>	
<b>Sub-theme: S3 Exam &amp; Course Content</b> <b>Q11 : Did the S3 exam reflect the content of the course? (i.e. Were the tasks/exercises questions related to the content of the course?)</b> <input type="checkbox"/> Yes <input type="checkbox"/> No Say in what way: .....			
<ul style="list-style-type: none"> <li>• Yes, the S3 exam reflected the course content</li> <li>• No, It did not.</li> </ul>	STUI7 : Yes, the debated theme: Economy, policy - law STUI8 : It depends STUI9 : Yes, same topics.	<ul style="list-style-type: none"> <li>• Undecided</li> <li>• Familiarity with the themes or topics</li> </ul>	
<b>Sub-theme: S4 Exam &amp; Course Content</b> <b>Q12 : Did the S4 exam reflect the content of the course?</b> <input type="checkbox"/> Yes <input type="checkbox"/> No Say in what way: .....			
<ul style="list-style-type: none"> <li>• Yes, the S4 exam reflected the course content.</li> <li>• No, It did not.</li> </ul>	STUI7 : Yes. The debated themes were the same in addition to some technology topic STUI8 : Yes, kind of. STUI9 : Yes, exam summed up what we had in class.	<ul style="list-style-type: none"> <li>• Familiarity with the themes or topics</li> <li>• Summary</li> </ul>	
<b>Sub-theme : Difference between S3 &amp; S4 Exams</b> <b>Q13 : What difference do you make between the S3 and S4 exams of Simultaneous Interpreting (SI) (from A into English)?</b>			
<ul style="list-style-type: none"> <li>• There is a difference.</li> <li>• There is no difference</li> </ul>	STUI7 : no difference. STUI8 : There is no difference mentioned; except our performance which is obviouslt much better in S4. STUI9 : S4 was easier.	<ul style="list-style-type: none"> <li>• ADJ</li> <li>• AJ</li> </ul>	- Performance better in S4 - S4 easier
<b>Sub-theme: Perceptions about the skills &amp; Sub-skills</b> <b>Q14 : What are the skills or sub-skills assessed in the S3 and S4 exams (SI from A into English)?</b> S3 exam: ..... S4 exam: .....			
<ul style="list-style-type: none"> <li>• The skills assessed in S3 exam</li> <li>• The skills assessed in S4 exam</li> </ul>	STUI7 : S3 exam: language and comprehension skills. S4 exam: At this level the grasp and reformulation without missing ideas. STUI8 : S3 exam: language comprehension S4 exam: being fast and brief STUI9 : S3 exam: language skills (more vocabulary and a better ability to	<ul style="list-style-type: none"> <li>• S3 Skills:</li> </ul>	- Language skills - Comprehension - Vocabulary

	understand) S4 exam: language skills (more vocabulary and a better ability to understand)	<ul style="list-style-type: none"> <li>S4 skills:</li> </ul>	<ul style="list-style-type: none"> <li>Grasp</li> <li>Reformulation</li> <li>Speed</li> <li>Brevity</li> <li>Vocabulary</li> <li>Comprehension</li> </ul>
Sub-theme : Assessment Criteria in the 2nd-year level	Q15 : According to you, what are the criteria of a good performance at the second year level? Say why:.....		
<ul style="list-style-type: none"> <li>1st-year level criteria</li> <li>2<sup>nd</sup>-year level criteria</li> </ul>	STUI7 : At this level, the student must be able to understand quickly the text and interpret it without missing ideas. STUI8 : Having a rich vocab and a good knowledge of the topic. Understanding what the speaker is speaking about is the key of giving a good performance. STUI9 : Speed and fluency, because it's important for a GOOD performance.	<ul style="list-style-type: none"> <li>The skills:</li> </ul>	<ul style="list-style-type: none"> <li>Quick comprehension</li> <li>Interpret the message</li> <li>Rich Vocabulary</li> <li>Topic-related knowledge</li> <li>Speed</li> <li>Fluency</li> </ul>
Sube-theme: Sharing of Assessment Criteria in S3 & S4 Exams	Q16 : Did the teachers share their assessment criteria with you before the exams? S3 exam: <input type="checkbox"/> Yes <input type="checkbox"/> No S4 exam: <input type="checkbox"/> Yes <input type="checkbox"/> No If your answer is yes, would you list them? .....		
<ul style="list-style-type: none"> <li>Yes, the teacher(s) shared their assessment criteria</li> <li>No, the teacher(s) did not.</li> </ul>	STUI7 : No ; No ; ////////////////////////////////////// STUI8 : No ; No ; ////////////////////////////////////// STUI9 : //////////////////////////////////////	<ul style="list-style-type: none"> <li>DAQ</li> <li>No, TEAs did not share the assessment criteria</li> </ul>	-
Sub-theme: Meaningfulness of Assessment Criteria	Q17 : Are these criteria meaningful/important for you? <input type="checkbox"/> Yes <input type="checkbox"/> No Would you say in what way?.....		
<ul style="list-style-type: none"> <li>Yes, the criteria are meaningful</li> <li>No, they are not meaningful</li> </ul>	STUI7 : ////////////////////////////////////// STUI8 : ////////////////////////////////////// STUI9 : //////////////////////////////////////	<ul style="list-style-type: none"> <li>DAQ</li> </ul>	
Sub-theme : Tasks of S3 Exam in SI	Q18 : What were the tasks/exercises of the S3 exam in SI (from A into English)? Please give a <u>detailed</u> description.		
<ul style="list-style-type: none"> <li>S3 exam various tasks which correspond to the level of study &amp; performance</li> </ul>	STUI7 : Topics about international organizations, policy, Economy STUI8 : ////////////////////////////////////// STUI9 : Listening to different records, trying to interpret them simultaneously then listening to our records and correct the mistakes using dictionaries, internet or asking the teacher.	<ul style="list-style-type: none"> <li>DAQ</li> <li>AJ</li> <li>St does not differentiate between topic and task</li> <li>Listening, SI, Listening to recordings, Correcting mistakes</li> </ul>	- Topic: Intl Org., Policy, Economy
Sub-theme : Tasks of S4 Exam in SI	Q19 : What were the tasks/exercises of the S4 exam in SI (from A into English)? Please give a <u>detailed</u> description.		
<ul style="list-style-type: none"> <li>S4 exam tasks which correspond to the level of study &amp; performance</li> </ul>	STUI7 : In Obama's speech about attacks. STUI8 : ////////////////////////////////////// STUI9 : Interpreting simultaneously in the booths.	<ul style="list-style-type: none"> <li>DAQ</li> <li>St does not differentiate between topic and task</li> <li>St does not differentiate between setting and task</li> </ul>	<ul style="list-style-type: none"> <li>Topic: Obama's speech</li> <li>SI in the booth</li> </ul>
Sub-theme : Duration of S3 Exam in SI	Q20 : How long did the S3 exam in Simultaneous Interpreting (SI) (from A into English) last? <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more?.....		
<ul style="list-style-type: none"> <li>5mn</li> <li>10mn</li> <li>15mn</li> <li>30mn</li> </ul>	STUI7 : ////////////////////////////////////// STUI8 : 10mn STUI9 : 10mn	<ul style="list-style-type: none"> <li>DAQ</li> <li>10mn</li> </ul>	<ul style="list-style-type: none"> <li>A into B</li> <li>C into A</li> </ul>

<ul style="list-style-type: none"> <li>45mn</li> <li>More than 45mn</li> </ul>			
Sub-theme : Prompts' length for S3 Exam in SI	Q21 : What was the length/duration of the prompt(s) for the S3 exam? Prompt 1: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... Prompt 2: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... More?.....		
<ul style="list-style-type: none"> <li>Prompt 1</li> <li>Prompt 2</li> <li>More prompts</li> </ul>	STUI7 : Prompt 1: 30 mn ; Prompt 2: 45mn STUI8 : Prompt 1: 5mn ; Prompt 2: 10mn STUI9 : Prompt 1: 10mn ; Prompt 2: 10 mn	<ul style="list-style-type: none"> <li>Prompt 1: 5mn-30mn</li> <li>Prompt 2: 10mn-45mn</li> </ul>	- STUs not sure about the meaning of the term - Take Vs Prompt
Sub-theme : Number of performance takes	Q22 : Did you interpret in one take? <input type="checkbox"/> Yes <input type="checkbox"/> No,.....		
<ul style="list-style-type: none"> <li>Yes, in one take</li> <li>No, in...</li> </ul>	STUI7 : Yes STUI8 : Yes STUI9 : Yes		
Sub-theme: Prompts' Negotiation	Q23 : Did you negotiate the prompts' length/duration with the teacher? <input type="checkbox"/> Yes <input type="checkbox"/> No		
<ul style="list-style-type: none"> <li>Yes, we negotiated with the teacher</li> <li>No, we did not</li> </ul>	STUI7 : ////////////////////////////////////// STUI8 : No STUI9 : Sometimes	<ul style="list-style-type: none"> <li>DAQ</li> <li>No</li> <li>Sometimes</li> </ul>	
Sub-theme : Duration of S4 Exam in SI	Q24 : How long did the S4 exam in SI (from A into English) last? <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more?...		
<ul style="list-style-type: none"> <li>5mn</li> <li>10mn</li> <li>15mn</li> <li>30mn</li> <li>45mn</li> <li>More than 45mn</li> </ul>	STUI7 : 30mn STUI8 : 10mn STUI9 : From 10mn to 15mn	<ul style="list-style-type: none"> <li>10mn-15mn</li> </ul>	
Sub-theme : Prompts' length for S4 Exam in SI	Q25 : What was the length/duration of the prompt(s) for the S4 exam? Prompt 1: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... Prompt 2: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... More?.....		
<ul style="list-style-type: none"> <li>Prompt 1</li> <li>Prompt 2</li> <li>More prompts</li> </ul>	STUI7 : Prompt 1: 30 mn ; Prompt 2: 45 mn. STUI8 : Prompt 1: 10mn ; Prompt 2: 10mn STUI9 : Prompt 1: 15mn ; Prompt 2: 15mn	<ul style="list-style-type: none"> <li>Prompt 1: 15-30mn</li> <li>Prompt 2: 15-45mn</li> <li>Prompt 1: 10 ; 10 (AB)</li> </ul>	- STUs from the same group gave different answers
Sub-theme : Number of takes in S4 Exam	Q26 : Did you interpret in one take? <input type="checkbox"/> Yes <input type="checkbox"/> No,.....		
<ul style="list-style-type: none"> <li>Yes, in one take</li> <li>No, in ...</li> </ul>	STUI7 : Yes. STUI8 : Yes STUI9 : Yes	<ul style="list-style-type: none"> <li>Yes, in one take</li> </ul>	
Sub-theme: Prompts' Negotiation	Q27 : Did you negotiate the prompts' length /duration with the teacher? <input type="checkbox"/> Yes <input type="checkbox"/> No		
<ul style="list-style-type: none"> <li>Yes, we negotiated with the teacher</li> </ul>	STUI7 : No STUI8 : No	<ul style="list-style-type: none"> <li>No negotiation</li> </ul>	

<ul style="list-style-type: none"> <li>No, we did not negotiate with the teacher</li> </ul>	STUI9 : No		
Sub-theme : Scoring of S3 & S4 Exams	Q28 : What were your S3 and S4 marks in SI (from A into English)? S3 mark: ..... S4 mark: .....		
<ul style="list-style-type: none"> <li>S3 mark</li> <li>S4 mark</li> </ul>	STUI7 : 13; 14 (from Eng into Ar) STUI8 : 12 ; 14 STUI9 : 13 ; 16	<ul style="list-style-type: none"> <li>Progress in S4</li> </ul>	
Theme: Feedback from teachers	Q29 : Did you have feedback from your teachers after the exams? If you did, would you give an example of this feedback.		
<ul style="list-style-type: none"> <li>Yes, we/I had feedback</li> <li>No, we/I did not.</li> </ul>	STUI7 : After every Exam, our teacher give us a feedback of the exam and remarke of our work. STUI8 : Yes; some remarks about the use of some terms STUI9 : Yes, about the terms' use and the speed .	<ul style="list-style-type: none"> <li>Terminology</li> <li>Speed</li> </ul>	
Theme : Satisfaction with the results	Q30 : Were you satisfied with your results in the S3 and S4 exams? Did they reflect your performance?		
<ul style="list-style-type: none"> <li>Yes, I was satisfied</li> <li>No, I was not</li> </ul>	STUI7 : In the S4, I was satisfied cause I've noticed that my performance was better than S3. STUI8 : Yes STUI9 : More satisfied of the S4 result .	<ul style="list-style-type: none"> <li>ADJ</li> <li>AJ</li> </ul>	<ul style="list-style-type: none"> <li>- Performance</li> <li>- Result</li> </ul>
Theme: Perceptions about courses	Q31 : To what extent have the SI courses (from A into English) of S3 and S4 prepared you for the S3 and S4 exams? S3 course:..... S4 course:.....		
<ul style="list-style-type: none"> <li>The S3 course was a good preparation for the exam</li> <li>The S4 course was a good preparation for the exam</li> <li>The S3 course was not good enough.</li> <li>The S4 course was not good enough.</li> </ul>	STUI7 : More able to understand quickly the text. STUI8 : S3 course: paved the way to a better performance S4 course: training STUI9 : S3 course: it paved the way to a better performance. S4 course: too helpful indeed.	<ul style="list-style-type: none"> <li>Training</li> <li>Helpful</li> <li>Increased text understanding</li> <li>Improved performance</li> </ul>	
Theme: Opinion on Exams	Q32 : Do you think that the exams reflect the real-life job characteristics? Why?		
<ul style="list-style-type: none"> <li>Yes, they reflect the real-life job characteristics</li> <li>No, they do not reflect the real-job.</li> </ul>	STUI7 : No, I don't. The real life of the interpreter cause in real situation, the SI is more diffcult and more stressful. STUI8 : Yes; since we work in a real conditions (in booth – with a limited duration and trying different pace and speakers) STUI9 : Because we tried the real-life job conditions and it was so similar of the ones of exams.	<ul style="list-style-type: none"> <li>The real job is more difficult</li> <li>The real job is more stressful</li> <li>They are similar to the real job characteristics</li> </ul>	
Theme: Comments & Suggestions	Q33 : Would you please add any comments or remarks that you think might be useful for this research? Thank you.		
<ul style="list-style-type: none"> <li>Comments</li> <li>Remarks</li> <li>Suggestions</li> </ul>	STUI7 : Our formation as futur interpreter able us to have a global idea about interpreting in real situation. STUI8 : ////////////////////////////////////// STUI9 : No	<ul style="list-style-type: none"> <li>The training enables the STUs to have an idea about the real-job contexts.</li> </ul>	

**Table 17: Data Coding of the Graduate Students' Responses:**

Theme(s)/ Categories	Textual Response for :	Axial Code	Sub-Code
Theme : Background	Q1: What is your gender?		
Sub-theme : Gender			
<ul style="list-style-type: none"> <li>Male</li> <li>Female</li> </ul>	GSTU1: Male ; GSTU2: Female ; GSTU3: Female ; GSTU4: Female ; GSTU5: Male ; GSTU6: FEMALE GSTU7: female ; GSTU8: female	<ul style="list-style-type: none"> <li>Male</li> <li>Female</li> </ul>	
Sub-theme: Age range	Q2: What is your age?		
<ul style="list-style-type: none"> <li>Young learner</li> <li>Old learner</li> </ul>	GSTU1: 40 ; GSTU2: 30 years old ; GSTU3: 32 ; GSTU4: 28 y. old ; GSTU5: 32 GSTU6: 28 ; GSTU7: 26 ; GSTU8: 42	<ul style="list-style-type: none"> <li>20-30</li> <li>30-40</li> <li>40-50</li> </ul>	
Sub-theme: origins	Q3 : Where are you from?		
<ul style="list-style-type: none"> <li>Algeria</li> <li>Arab Country</li> <li>Foreign Country</li> </ul>	GSTU1: Algiers ; GSTU2: Algeria ; GSTU3: Mila, Algeria GSTU4: Algiers, Algeria. GSTU5: Tunisia GSTU6: ALGIERS- ALGERIA GSTU7: Algeria GSTU8: Egypt	<ul style="list-style-type: none"> <li>Country of origins</li> <li>City of origins</li> </ul>	<ul style="list-style-type: none"> <li>- Algeria</li> <li>- Tunisia</li> <li>- Egypt</li> <li>- Algiers</li> <li>- Mila</li> </ul>
Sub-theme: Qualifications	Q4 : What was your highest degree when you enrolled in the Higher Arab Institute of Translation (HAIT)? In what field? <input type="checkbox"/> BA <input type="checkbox"/> MA <input type="checkbox"/> Other		
<ul style="list-style-type: none"> <li>Min. Requirement</li> <li>More than the Min. Requirements</li> </ul>	GSTU1: MA GSTU2: bachelor of arts in translation GSTU3: BA GSTU4: Ba, Translation and Interpreting. GSTU5: BA GSTU6: BA GSTU7: BA ; Translation GSTU8: //////////////////////////////////////	<ul style="list-style-type: none"> <li>MA</li> <li>BA</li> <li>BA Transl.</li> <li>BA Transl. &amp; Interpreting</li> <li>Field, not mentioned</li> <li>DAQ</li> </ul>	
Sub-theme: Mastery of Languages	Q5 : What is/are your ability(ies) in the following languages on a scale from 1 to 5? *Level 5: Excellent *Level 4: Very good *Level 3: Good *Level 2: Average *Level 1: weak <ul style="list-style-type: none"> <li>Arabic (.....) – French (.....) – English (.....)</li> <li>Other: .....(.....) ; .....(.....)</li> </ul>		
<ul style="list-style-type: none"> <li>Medium level</li> <li>Proficient bilingual</li> <li>Proficient tilingual</li> <li>Polyglot</li> </ul>	GSTU1: Arabic Good – French Good – English Good GSTU2: Arabic (...5.....) – French (.....4.....) – English (...3...) GSTU3: Arabic (5) – French (3) – English (5) ; German (1) GSTU4: Arabic (5) – French (4) – English (3) GSTU5: Arabic (average) – French (good.) – English (very good) GSTU6: Arabic (05) – French (05) – English (04) GSTU7: Arabic (5) – French (3) – English (4) ; Spanish (Beginner) GSTU8: Arabic (4) – French (2) – English (4)	<ul style="list-style-type: none"> <li>The general level corresponds to the exit level</li> <li>Graduate still needs remedial courses</li> </ul>	
Theme: Graduation	Q6 : When did you graduate from the HAIT?		
<ul style="list-style-type: none"> <li>Old Graduate</li> <li>Fresh Graduate</li> </ul>	GSTU1: Dec 2007 ; GSTU2: DECEMBER 2007 ; GSTU3: December 2007 ; GSTU4: November 2008 GSTU5: 2009 ; GSTU6: 2010 ; GSTU7: 2013 ; GSTU8: 2008	<ul style="list-style-type: none"> <li>2007</li> <li>2008</li> </ul>	

		<ul style="list-style-type: none"> <li>• 2009</li> <li>• 2010</li> <li>• 2013</li> </ul>	
Theme: Working languages	Q7 : Is English your B or C language? <input type="checkbox"/> B <input type="checkbox"/> C		
<ul style="list-style-type: none"> <li>• Anglophones' group</li> <li>• Francophones' group</li> </ul>	GSTU1: B; GSTU2: C; GSTU3: B; GSTU4: B; GSTU5: B; GSTU6: C; GSTU7: B GSTU8: /////	<ul style="list-style-type: none"> <li>• Not mentioned</li> <li>• DAQ</li> </ul>	
Sub-theme: Jobs	Q8 : What do you do now for a living?		
<ul style="list-style-type: none"> <li>• Translator</li> <li>• Interpreter</li> <li>• Both</li> <li>• Other</li> </ul>	GSTU1: I am a consultant and have my own office of translation. GSTU2: SWORN TRANSLATOR ; GSTU3: English teacher ; GSTU4: Sworn translator GSTU5: I work as a journalist and freelance translator. GSTU6: WORKING FOR AN OIL AND GAS COMPANY GSTU7: Freelance translator ; GSTU8: working in translation	<ul style="list-style-type: none"> <li>• Consultant</li> <li>• Teacher</li> <li>• Journalist</li> <li>• Administrator</li> </ul>	
Theme: Scoring in the Admission Test	Q9 : What was your score and rank among those who passed the Admission Test (the interview) of the HAIT? Score:..... Rank:.....		
<ul style="list-style-type: none"> <li>• Good score and rank</li> <li>• Bad score and rank</li> </ul>	GSTU1: Everybody was accepted without sitting for admission test. However, we were given a test to assess our level, but no mark was given. GSTU2: I WASN'T INVITED TO DO SUCH A TEST GSTU3: do not know GSTU4: No test passed. Enrollment on the basis of a scholarship granted by the Ministry of Higher Education and Scientific Research to the top class GSTU5: I did not take the Admission Test. I just mailed my application and got admitted to the MA programme. GSTU6: N/A; N/A ; GSTU7: Rank:1 ; GSTU8: /////	<ul style="list-style-type: none"> <li>• Admitted without being tested</li> <li>• Admitted with a test</li> <li>• DAQ</li> <li>• NA</li> </ul>	<ul style="list-style-type: none"> <li>• Do not know</li> <li>• Not invited</li> <li>• Scholarship</li> <li>• Application</li> <li>• Rank 1</li> </ul>
Sub-theme: Perceptions about the Admission Test	Q10 : How did you find the Admission Test (the interview)? <input type="checkbox"/> Very Easy <input type="checkbox"/> Easy <input type="checkbox"/> A little Difficult <input type="checkbox"/> Difficult <input type="checkbox"/> Very Difficult Say why:.....		
<ul style="list-style-type: none"> <li>• Very Easy</li> <li>• Easy</li> <li>• A little Difficult</li> <li>• Difficult</li> <li>• Very Difficult</li> </ul>	GSTU1: Easy ; ///// ; GSTU2: ///// ; GSTU3: no test GSTU4: ///// ; GSTU5: ///// GSTU6: Easy ; ///// GSTU7: Easy ; because it was about our ambitions, so it was easy talking about yourself GSTU8: /////	<ul style="list-style-type: none"> <li>• No test</li> <li>• DAQ</li> <li>• ADJ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>• Easy</li> <li>• Self-centered questions</li> </ul>
Theme: Scoring in the Interpreter Aptitude Test (IAT)	Q11 : What was your score and rank among those who passed the Interpreter Aptitude Test of the HAIT (Beginning of S2)? Score:..... Rank:.....		
<ul style="list-style-type: none"> <li>• Good score and rank</li> <li>• Bad score and rank</li> </ul>	GSTU1: Everybody was accepted without sitting for admission test. GSTU2: Score:16 ; Rank: very honorable GSTU3: no test , ///// GSTU4: No test passed. ///// GSTU5: ///// GSTU6: N/A ; N/A GSTU7: Score:13 ; Rank:1 GSTU8: /////	<ul style="list-style-type: none"> <li>• Admitted without a test</li> <li>• Admitted with a test</li> </ul>	
Sub-theme: Easiness of the IAT	Q12 : How did you find the Interpreter Aptitude Test? <input type="checkbox"/> Very Easy <input type="checkbox"/> Easy <input type="checkbox"/> A little Difficult <input type="checkbox"/> Difficult <input type="checkbox"/> Very Difficult Say why:.....		

<ul style="list-style-type: none"> <li>• Very Easy</li> <li>• Easy</li> <li>• A little difficult</li> <li>• Difficult</li> <li>• Very Difficult</li> </ul>	GSTU1: Please read the previous answer GSTU2: A little Difficult ; a little difficult due to stress of the exam but we were adapted and well- trained in interpreting to such a test GSTU3: no test GSTU4: /////////////// ; GSTU5: /////////////// ; GSTU6: Easy ; /////////////// ; GSTU7: A little Difficult ; because I had no idea what interpretation meant, I studied translation at the university. GSTU8: ///////////////	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>• Stress</li> <li>• I studied translation, I wasn't prepared for interpreting</li> </ul>
Theme: Perceptions about Exams' Specifications Sub-theme: Mark in S2 Exam in SI	Q13 : What was your mark in Simultaneous Interpreting (SI) Exam (From Arabic/A into English) in the second semester (S2) of the 1 <sup>st</sup> year?		
<ul style="list-style-type: none"> <li>• Numerical mark</li> <li>• Nonnumerical</li> </ul>	GSTU1: The simultaneous interpreting was part of the second year curriculum. GSTU2: 13 /20 GSTU3: 13 ; GSTU4: 15 ; GSTU5: 13 GSTU6: 16 ; GSTU7: 16 GSTU8: ///////////////	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• Unexpected answer</li> <li>• 13 ; 15 ; 16</li> </ul>	<ul style="list-style-type: none"> <li>• No course</li> </ul>
Sub-theme: Easiness of S3 Exam in SI	Q14 : How did you find the Simultaneous Interpreting (SI) Exam (From A into English) of the third semester (S3) (2 <sup>nd</sup> Year)? <input type="checkbox"/> Very Easy <input type="checkbox"/> Easy <input type="checkbox"/> A little Difficult <input type="checkbox"/> Difficult <input type="checkbox"/> Very Difficult Say why: .....		
<ul style="list-style-type: none"> <li>• Very Easy</li> <li>• Easy</li> <li>• A little Difficult</li> <li>• Difficult</li> <li>• Very Difficult</li> </ul>	GSTU1: A little Difficult ; As a beginner the exam was not easy. GSTU2: Very Easy ; we were lucky to have a very good and talented teacher, an ex-Un interpreter: Faleh ABDALLAH WHO GAVE US A GOOD BASIS IN THAT MATTER GSTU3: Easy ; I was fluent in English and had no pb in Arabic. The test was around issues covered in class. GSTU4: Easy ; /////////////// GSTU5: Very challenging ; /////////////// GSTU6: Easy ; /////////////// GSTU7: A little Difficult ; /////////////// GSTU8: ///////////////	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>• Easy</li> <li>• Very challenging</li> <li>• A little difficult</li> <li>• Training</li> <li>• Bilingualism</li> <li>• Content</li> </ul>
Sub-theme: Easiness of S4 Exam in SI	Q15 : How did you find the SI exam (From A into English) of the fourth semester (S4) (2 <sup>nd</sup> Year)? <input type="checkbox"/> Very Easy <input type="checkbox"/> Easy <input type="checkbox"/> A little Difficult <input type="checkbox"/> Difficult <input type="checkbox"/> Very Difficult Say why: .....		
<ul style="list-style-type: none"> <li>• Very Easy</li> <li>• Easy</li> <li>• A little Difficult</li> <li>• Difficult</li> <li>• Very Difficult</li> </ul>	GSTU1: A little Difficult ; Any exam has its own degree of difficulty to test the student ability to cope with it. GSTU2: Easy ; we participated in a lot of conferences GSTU3: Easy ; I was fluent in English and had no pb in Arabic. The test was around issues covered in class. GSTU4: Easy ; /////////////// GSTU5: Challenging too ; /////////////// GSTU6: Easy ; /////////////// GSTU7: Easy ; /////////////// GSTU8: ///////////////	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>• Every exam is difficult</li> <li>• Conferences preparation made it easy</li> <li>• Bilingualism is enough</li> </ul>
Sub-theme: Differences between Exams	Q16 : Were the S3 and S4 exams in SI (from A into English) (the 2 <sup>nd</sup> year) different from the S2 exam in SI (from A into English) (1 <sup>st</sup> year)? <input type="checkbox"/> Yes <input type="checkbox"/> No Say in what way: .....		
<ul style="list-style-type: none"> <li>• Yes</li> <li>• No</li> </ul>	GSTU1: Yes ; Say in what way: We acquired some technics that helped us to get acquainted with the module. GSTU2: Yes ; we had been invited to interpret the whole speeches not just a part of them which were spoken in a fast manner as well. ; GSTU3: Yes ; more complex in style, lexis and theme and longer. GSTU4: Yes ; for the first year the exam was consisting of simultaneous interpreting of a documentary whereas the 2 <sup>nd</sup> year exams were speeches. ; GSTU5: Not really. They were pretty much the same.	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> <li>• AJ</li> </ul>	<ul style="list-style-type: none"> <li>• Training</li> <li>• Rate of delivery</li> <li>• More complex</li> </ul>

	GSTU6: Yes, THE SPEED AND THE NATURE OF THE TEXT ARE DIFFERENT GSTU7: Yes, they were more difficult and varied (different fields) ; GSTU8: //		<ul style="list-style-type: none"> <li>Text type</li> <li>Longer</li> </ul>
Sub-theme: S3 Exam & Course Content	Q17 : Did the S3 exam reflect the content of the course? (i.e. Were the tasks/exercises questions related to the content of the course?) <input type="checkbox"/> Yes <input type="checkbox"/> No Say in what way: .....		
<ul style="list-style-type: none"> <li>Yes, S3 exam reflected the course content</li> <li>No, S3 exam did not reflect the course content</li> </ul>	GSTU1: Yes ; We were tested according the courses given. GSTU2: Yes ; it related to political actuality ; GSTU3: Yes ; // GSTU4: Yes ; texts already done in class GSTU5: Actually, there was no course. We had a lot of practice each day and exams were completely different from we studied during course sessions. ; GSTU6: No, THE EXAM WAS TOTALLY DIFFERENT GSTU7: Yes, the exam was one of the speeches we took during class GSTU8: //	<ul style="list-style-type: none"> <li>DAQ</li> <li>ADJ</li> <li>AJ</li> </ul>	<ul style="list-style-type: none"> <li>Yes</li> <li>The same texts</li> <li>The same text type</li> <li>Different content</li> </ul>
Sub-theme: S4 Exam & Course Content	Q18 : Did the S4 exam reflect the content of the course? <input type="checkbox"/> Yes <input type="checkbox"/> No Say in what way: .....		
<ul style="list-style-type: none"> <li>Yes, S4 exam reflected the course content</li> <li>No, S4 exam did not reflect the course content</li> </ul>	GSTU1: Yes ; We were tested according the courses given GSTU2: No ; it was varied between technical, political and scientific subjects GSTU3: Yes ; /////////////// ; GSTU4: Yes, The subjects were previously tackled GSTU5: Same like the previous question ; GSTU6: No, THE SUBJECTS WERE DIFFERENT GSTU7: Yes, /////////////// ; GSTU8: //	<ul style="list-style-type: none"> <li>DAQ</li> <li>ADJ</li> <li>AJ</li> </ul>	<ul style="list-style-type: none"> <li>Yes</li> <li>Varied subjects/themes</li> <li>Same subjects/themes</li> </ul>
Sub-theme: Difference between S3 & S4	Q19 : What difference do you make between the S3 and S4 exams of Simultaneous Interpreting (SI) (from A into English)?		
<ul style="list-style-type: none"> <li>There is no difference</li> <li>They are different</li> </ul>	GSTU1: I think they were the same. GSTU2: The degree of difficulty GSTU3: S3 exams more organized more focused more relevant to content fairer than those of S4 GSTU4: Exams duration (lesser in S3 and S4) and texts type (more difficult) GSTU5: // GSTU6: THE RYTHME, THE FLOW, THE SPEED, AND THE VOCABULARY WERE MORE A LESS DIFFICULT IN S4 GSTU7: They are much alike GSTU8: //	<ul style="list-style-type: none"> <li>Not different</li> <li>Different</li> <li>DAQ</li> </ul>	<ul style="list-style-type: none"> <li>Degree of difficulty</li> <li>Duration</li> <li>Text types</li> <li>Rate</li> <li>Vocabulary</li> </ul>
Sub-theme: The assessed skills in the Tests & Exams	Q20 : What were the skills or sub-skills assessed in the Admission Test, the Interpreter Aptitude Test, S2, S3 and S4 exams (SI from A into English)? The Admission Test:..... The Interpreter Aptitude Test:..... S2 exam:..... S3 exam: ..... S4 exam: .....		
<ul style="list-style-type: none"> <li>Unaware about the skills</li> <li>Aware about the skills</li> </ul>	GSTU1: The Admission Test: The mastery of the target and the source languages. The Interpreter Aptitude Test: We had not such a test S2 exam: We had not such a test. S3 exam: We had not such a test. S4 exam: We had not such a test. GSTU2: The Admission Test: be quick in finding the appropriate words and meaning as well The Interpreter Aptitude Test: to be quiet calm and give the right meaning S2 exam: to keep a moderate tone of voice S3 exam: fluidity and rapidity S4 exam: further to all the here above answers, be able to interpret on and on during 30 TO 35mn GSTU3: The Admission Test:...../... The Interpreter Aptitude Test:...../...	<ul style="list-style-type: none"> <li>Skills match the skills assessed by teachers</li> <li>Skills do not match the skills assessed by teachers</li> </ul>	



	<p>The Interpreter Aptitude Test: <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>S2 exam: <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>S3 exam: <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>S4 exam: <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>If your answer is yes, would you list them? .....</p>		
<ul style="list-style-type: none"> <li>• Yes, the teachers shared their assessment criteria</li> <li>• No, they did not.</li> </ul>	<p>GSTU1: No ; //</p> <p>GSTU2: The Admission Test: <input type="checkbox"/> No The Interpreter Aptitude Test: <input type="checkbox"/> Yes S2 exam: <input type="checkbox"/> No S3 exam: <input type="checkbox"/> Yes S4 exam: <input type="checkbox"/> Yes ;</p> <p>Ms. Lynne FRANJIE Mr. Hassen Hamze and Mr. FALEH Abdallah</p> <p>GSTU3: No ; //</p> <p>GSTU4: /// , S4 exam: <input type="checkbox"/> Yes ; Those mentioned in Q20</p> <p>GSTU5: //</p> <p>GSTU6: No ; //</p> <p>GSTU7: The Admission Test: No The Interpreter Aptitude Test: Yes S2 exam: No S3 exam: Yes S4 exam: No</p> <p>Listen carefully Understand the speech Make meaningful sentences</p> <p>GSTU8: //</p>	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> </ul> <ul style="list-style-type: none"> <li>• Yes</li> <li>• No</li> </ul>	- No
Sub-theme: Meaningfulness of the criteria	<p>Q23 : Are these criteria meaningful/important for you? <input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>Would you say in what way?.....</p>		
<ul style="list-style-type: none"> <li>• Yes, the criteria are meaningful</li> <li>• No, the criteria are not meaningful</li> </ul>	<p>GSTU1: //</p> <p>GSTU2: Yes ; To be conscious of the task we 're asked to do and how to act</p> <p>GSTU3: Yes ; You bear them in mind and try to meet them while preparation for and performance in the exam so as to both get a good mark (succeed) and graduate with a good level</p> <p>GSTU4: Yes ; //</p> <p>GSTU5: //</p> <p>GSTU6: No ; //</p> <p>GSTU7: Yes, Would you say in what way? They helped me to be better and understand what I was doing, because when you concentrate on giving a complete meaningful sentences, after understanding the speech, you will be able to improve your interpretation</p> <p>GSTU8: //</p>	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• ADJ</li> <li>• Yes, meaningful</li> </ul>	<p>- No</p> <ul style="list-style-type: none"> <li>• Consciousness/Repetition</li> <li>• Performance/Mark</li> <li>• Completeness of sentence/ Meaning</li> </ul>
Sub-theme: Tasks of S3 exam in SI	<p>Q24 : What were the tasks/exercises of the S3 exam in SI (from A into English)? Please give a <u>detailed</u> description.</p>		
<ul style="list-style-type: none"> <li>• Tasks according to level of study</li> <li>• Tasks according to level of performance</li> <li>• Tasks according to level of proficiency</li> </ul>	<p>GSTU1: Sight translation and speech interpreting.</p> <p>GSTU2: First, do the up sight translation of UN speeches, conventions and then interpret them consecutively to be able to hear the whole speech and interpret it rightly</p> <p>Gst3: Listen to an address delivered by a Nigerian President at the General Assembly of the UN and interpret it simultaneously into Arabic...The students felt unable to interpret into English (though it was their specialization) therefore they asked the teacher not to have an exam Ar/Eng until the fourth semester and the teacher accepted because he saw their very weak level</p> <p>GSTU4: Simultaneous interpreting</p> <p>GSTU5: //</p> <p>GSTU6: DIRECT INTERPRETATION OF OFFICIAL SPEECHES, TV SHOWS, AND DOCUMENTARIES.</p>	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• Learner confuses the different modes</li> </ul> <ul style="list-style-type: none"> <li>• Did not do an exam</li> <li>• Did not give details</li> <li>• SI from B into A</li> <li>• Broad description</li> </ul>	<p>- Speech interpreting</p> <p>- Sight transl.</p> <p>- SI</p> <p>- SI of speeches, TV</p>

	GSTU7: // GSTU8: //		shows, documentaries
Sub-theme: Tasks of S4 exam in SI	Q25 : What were the tasks/exercises of the S4 exam in SI (from A into English)? Please give a <u>detailed</u> description.		
<ul style="list-style-type: none"> <li>• Tasks according to level of study</li> <li>• Tasks according to level of performance</li> <li>• Tasks according to level of proficiency</li> </ul>	GSTU1: Sight translation and speech interpreting ; /////////////// GSTU2: Interpret whatever kind of a speech or a text never seen before GSTU3: Listen to a passage (MAY BE RECORDED FROM Aljazeera) and interpret it simultaneously into ENGLISH GSTU4: Simultaneous interpreting ; // GSTU5: // GSTU6: DIRECT INTERPRETATION OF OFFICIAL SPEECHES, TV SHOWS, AND DOCUMENTARIES. ; /////////////// GSTU7: // GSTU8: //	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• Learner confuses the different modes</li> <li>• Did not do an exam</li> <li>• Did not give details</li> <li>• SI from B into A</li> <li>• Broad description</li> </ul>	<ul style="list-style-type: none"> <li>- Speech interpreting</li> <li>- Sight transl.</li> <li>- SI of speeches, TV shows, documentaries</li> </ul>
Sub-theme: Duration of S3 Exam in SI	Q26 : How long did the S3 exam in Simultaneous Interpreting (SI) (from A into English) last? <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more?.....		
<ul style="list-style-type: none"> <li>• 5mn</li> <li>• 10mn</li> <li>• 15mn</li> <li>• 30mn</li> <li>• 45mn</li> <li>• More than 45mn</li> </ul>	GSTU1: 5mn GSTU2: 15mn ; I'm not sure sorry..... GSTU3: 5mn GSTU4: 15 mn GSTU5: I don't really remember. It has been more than 5 years from now . Sorry. GSTU6: 15 mn ; GSTU7: 05 mn ; GSTU8: ///////////////	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• Learner does not remember</li> <li>• 5mn-15mn</li> </ul>	
Sub-theme: Prompts' length in S3 Exam in SI	Q27 : What was the length/duration of the prompt(s) for the S3 exam? Prompt 1: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... Prompt 2: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... More?.....		
<ul style="list-style-type: none"> <li>• 5mn</li> <li>• 10mn</li> <li>• 15mn</li> <li>• 30mn</li> <li>• 45mn</li> <li>• More than 45mn</li> </ul>	GSTU1: Prompt 1: 5mn ; Prompt 2: 5mn GSTU2: Prompt 1: 30mn ; Prompt 2: 30mn GSTU3: // GSTU4: /////////////// GSTU5: // GSTU6: Prompt 1: 15mn ; Prompt 2: 15mn GSTU7: /////////////// ; GSTU8: //	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• Prompt(s) 1</li> <li>• Prompt(s) 2</li> </ul>	<ul style="list-style-type: none"> <li>- 5mn</li> <li>- 15mn</li> <li>- 30m</li> <li>- 5m</li> <li>- 15mn</li> <li>- 30mn</li> </ul>
Sub-theme : Number of Takes	Q28 : Did you interpret in one take? <input type="checkbox"/> Yes <input type="checkbox"/> No,.....		
<ul style="list-style-type: none"> <li>• Yes, in one take</li> <li>• No,...</li> </ul>	GSTU1: Yes ; GSTU2: Yes ; GSTU3: Yes ; GSTU4: Yes ; GSTU5: I guess so ; GSTU6: Yes GSTU7: /////////////// ; GSTU8: ///////////////	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• One take</li> </ul>	
Sub-theme: Prompts' Negotiation	Q29 : Did you negotiate the prompts' length/duration with the teacher? <input type="checkbox"/> Yes <input type="checkbox"/> No		
<ul style="list-style-type: none"> <li>• Yes, we negotiated with the teacher</li> <li>• No, we did not negotiate with the teacher</li> </ul>	GSTU1: No ; GSTU2: No ; GSTU3: No GSTU4: /////////////// ; GSTU5: I don't think so. Timing is challenge number in interpreting. GSTU6: No ; GSTU7: Yes GSTU8: ///////////////	<ul style="list-style-type: none"> <li>• DAQ</li> <li>• No negociation</li> <li>• Yes (1)</li> </ul>	
Sub-theme: Duration of S4 Exam in SI	Q30 : How long did the S4 exam in SI (from A into English) last? <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more?.....		
<ul style="list-style-type: none"> <li>• 5mn</li> <li>• 10mn</li> <li>• 15mn</li> <li>• 30mn</li> </ul>	GSTU1: 5mn GSTU2: 45mn GSTU3: 10mn GSTU4: 30 mn	<ul style="list-style-type: none"> <li>• DAQ</li> </ul>	

<ul style="list-style-type: none"> <li>45mn</li> <li>More than 45mn</li> </ul>	GSTU5: /// GSTU6: 30 mn ; GSTU7: 05 mn ; GSTU8: ///		
Sub-theme: Prompts' length of S4 Exam in SI	Q31 : What was the length/duration of the prompt(s) for the S4 exam? Prompt 1: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... Prompt 2: <input type="checkbox"/> 5mn <input type="checkbox"/> 10mn <input type="checkbox"/> 15mn <input type="checkbox"/> 30mn <input type="checkbox"/> 45mn <input type="checkbox"/> more:..... More?.....		
<ul style="list-style-type: none"> <li>5mn</li> <li>10mn</li> <li>15mn</li> <li>30mn</li> <li>45mn</li> <li>More than 45mn</li> </ul>	GSTU1: Prompt 1: <input type="checkbox"/> 5mn ; Prompt 2: <input type="checkbox"/> 5mn GSTU2: Prompt 1: <input type="checkbox"/> 15mn ; Prompt 2: <input type="checkbox"/> 30mn GSTU3: /// GSTU4: /// GSTU5: /// GSTU6: Prompt 1: 30mn ; Prompt 2: 30mn GSTU7: /////////////// ; GSTU8: ///	<ul style="list-style-type: none"> <li>DAQ</li> <li>Prompts 1</li> </ul>	- 5mn - 15 mn - 30mn
Sub-theme : Number of Takes in S4 Exam	Q32 : Did you interpret in one take? <input type="checkbox"/> Yes <input type="checkbox"/> No,.....		
<ul style="list-style-type: none"> <li>Yes, in one take</li> <li>No, in more than one take</li> </ul>	GSTU1: Yes ; GSTU2: Yes ; GSTU3: Yes ; GSTU4: Yes GSTU5: /// ; GSTU6: Yes ; GSTU7: /////////////// ; GSTU8: ///////////////	<ul style="list-style-type: none"> <li>DAQ</li> </ul>	
Sub-theme: Negotiation of Prompts' Length	Q33 : Did you negotiate the prompts' length /duration with the teacher? <input type="checkbox"/> Yes <input type="checkbox"/> No		
<ul style="list-style-type: none"> <li>Yes, we negotiated with the teacher</li> <li>No, we did not</li> </ul>	GSTU1: No ; GSTU2: No ; GSTU3: No ; GSTU4: No GSTU5: /// GSTU6: No ; GSTU7: No ; GSTU8: ///	<ul style="list-style-type: none"> <li>DAQ</li> <li>No negotiation</li> </ul>	
Sub-theme: Scoring in S3 & S4 Exams in SI	Q34 : What were your S3 and S4 marks in SI (from A into English)? S3 mark: ..... S4 mark: .....		
<ul style="list-style-type: none"> <li>S3 Mark in SI</li> <li>S4 Mark in SI</li> </ul>	GSTU1: S3 mark: 15 ; S4 mark: 11 ; GSTU2: S3 mark:16 ; S4 mark: 14 GSTU3: S3 mark: .....15 ; S4 mark: .....15 GSTU4: S3 mark: 15,5 ; S4 mark: 13 GSTU5: /////////////// ; GSTU6: S3 mark: 17 ; S4 mark: 17 GSTU7: S3 mark: 16 ; S4 mark: 15 ; GSTU8: ///////////////	<ul style="list-style-type: none"> <li>DAQ</li> <li>Decrease in marks, decline in performance</li> <li>Same Mark</li> </ul>	- 04 Sts - 02 Sts
<ul style="list-style-type: none"> <li>Yes, we did have some feedback.</li> <li>No, we did not have any feedback.</li> </ul>	Q35 : Did you have feedback from your teachers after the exams? If you did, would you give an example of this feedback. GSTU1: No, At all. GSTU2: Ms Lynne FRANJIE SAID THAT I GAVE THE RIGHT MEANING, TRIED TO FOLLOW UP THE SPEAKER . she also gave marks about the tone voice and coherence of my text GSTU3: You should perform in a louder voice you're your voice should be more rigorous GSTU4: Sometimes. Terminology corrections. GSTU5: Most teachers were really kind. We received a lot of encouragement. GSTU6: No ; /////////////// GSTU7: /////////////// ; GSTU8: ///////////////	<ul style="list-style-type: none"> <li>DAQ</li> <li>ADJ</li> <li>A&amp;J</li> </ul>	<ul style="list-style-type: none"> <li>Appraisal</li> <li>Encouragement</li> <li>Shadowing</li> <li>Coherence</li> <li>Voice control</li> <li>Terminology</li> </ul>
<ul style="list-style-type: none"> <li>Yes, I was satisfied</li> <li>No, I was not</li> </ul>	Q36 : Were you satisfied with your results in the S3 and S4 exams? Did they reflect your performance? GSTU1: No ; No ; GSTU2: Yes, I was GSTU3: S3 SATISFIED WITH HE PERFORMANCE BUT NOT WITH THE MARK...S4 not satisfied with performance because I was very ill and had tremendous problems GSTU4: Actually yes ; GSTU5: Yes. But I always believed that I could do better. GSTU6: Yes ; ++++++ GSTU7: Not really, sometimes we feel like some teachers give us more credit than we deserve	<ul style="list-style-type: none"> <li>DAQ</li> <li>Yes, but...</li> <li>No, because</li> </ul>	

	GSTU8: *****		
Theme: Exams & Course Content	Q37 : To what extent have the SI courses (from A into English) of S3 and S4 prepared you for the S3 and S4 exams? S3 course:..... S4 course:.....		
<ul style="list-style-type: none"> <li>S3 SI course content was a good preparation for the exam</li> <li>It was not</li> <li>S4 SI course content was a good preparation for the exam</li> <li>It was not</li> </ul>	GSTU1: //////////////// ; GSTU2: S3 course: it was an initiation ..... / S4 course: adaptation GSTU3: S3 course: approximately no preparation ; S4 course: just a bit GSTU4: S3 course: terminology and subjects already seen during courses S4 course: terminology and subjects already seen during courses GSTU5: //////////////// GSTU6: S3 course: CONTROL THE ACCESSORIES (VOICE FOR EXAMPLE) S4 course: READY FOR THE PROFESSIONAL LIFE GSTU7: S3 course: a lot, it gave me the reflex and the ability to be a better interpreter S4 course: / GSTU8: ////////////////	<ul style="list-style-type: none"> <li>Satisfied</li> <li>Unsatisfied</li> </ul>	
Theme: Authenticity	Q38 : Do you think that the exams reflect the real-life job characteristics? Why?		
<ul style="list-style-type: none"> <li>Yes</li> <li>No</li> </ul>	GSTU1: An Exam of 5 minutes have nothing to do with a whole day consecutive/simultaneous interpreting. GSTU2: Although our exam was a little difficult because the speaker was reading in a very fast way, we all passed it successfully so, I can say that the real-life job wouldn't be such difficult GSTU3: Yes... you are put in the same situation GSTU4: Indeed they reflect but real life is easier than exams GSTU5: I gained thorough knowledge about interpreting and translation skills, which have greatly helped me in my current job. GSTU6: NOT THE EXAM ITSELF. BUT THE COURSES ARE GOOD EXERCICES AS THE STUDENT TACKLES ALMOST ALL SUBJECTS WITH NDIFFEREND SPEED. GSTU7: No, because we don't express the same stress, when you interpret for you (during lessons) or even for getting a better mark, is not like when you know that there are people who are listening and depend entirely on you in order to understand what the speaker is saying. GSTU8: ////////////////	<ul style="list-style-type: none"> <li>DAQ</li> <li>Yes, but</li> <li>No, because</li> <li>The course is good exercise</li> </ul>	<ul style="list-style-type: none"> <li>- Real life is easier</li> <li>- Exam duration not enough</li> <li>- Stress is different</li> </ul>
Theme: Opinion	Q39 : Would you please add any comments or remarks that you think might be useful for this research? Thank you.		
<ul style="list-style-type: none"> <li>Comments</li> <li>Remarks</li> <li>Suggestions</li> </ul>	GSTU1:***** GSTU2: The field of interpreting has specialized sub-fields so it would be helpful to be specialized in the area you're good at than interpreting only, for instance: medical interpreting, court interpreting, etc. GSTU3: I did not feel that teachers used to base their evaluations on specific criteria and they should, I think. If there are any criteria, they should be displayed to the students. Feedback was needful but was missing most of the time ... Good Luck! GSTU4: No comments to add GSTU5: The research looks really good all in all though I suggest you delete some very detailed questions, especially for those who graduated from a long time. Thank you. I wish you all the best. GSTU6: THE HAIT EXPERIENCE ENABLES ME TO BE A COMPLETE INTERPRETER. I GOT A DIFFERENT KNOWLEDGE FROM THE UNIVERSITY. GSTU7: I think that in order to be a better translator or an interpreter, we need to practice with experienced people in the field, listen and read a lot. GSTU8: ////////////////	<ul style="list-style-type: none"> <li>DAQ</li> <li>Suggestions :</li> <li>Remarks :</li> </ul>	<ul style="list-style-type: none"> <li>- The learner suggests specialization</li> <li>- Displaying the criteria of evaluation</li> <li>- Giving feedback</li> <li>- More practice in the field</li> <li>-The questions were very detailed</li> </ul>