

2012

Assessment and Evaluation in Labs for Public Service Interpreting Training

Carmen Valero Garcés

University of Alcalá, carmen.valero@uah.es

Denis Socarrás-Estrada

University of Alcalá, deniswcu@yahoo.es

Follow this and additional works at: <https://tigerprints.clemson.edu/ijie>



Part of the [Education Commons](#), and the [Sign Languages Commons](#)

Recommended Citation

Garcés, Carmen Valero and Socarrás-Estrada, Denis (2012) "Assessment and Evaluation in Labs for Public Service Interpreting Training," *International Journal of Interpreter Education*: Vol. 4 : Iss. 2 , Article 3.

Available at: <https://tigerprints.clemson.edu/ijie/vol4/iss2/3>

This Research Article is brought to you for free and open access by TigerPrints. It has been accepted for inclusion in *International Journal of Interpreter Education* by an authorized editor of TigerPrints. For more information, please contact kokeefe@clemson.edu.

Assessment and Evaluation in Labs for Public Service Interpreting Training

Carmen Valero-Garcés¹

University of Alcalá

Denis Socarrás-Estrada

University of Alcalá

Abstract

This study reports on the development and application of two bilingual interpreting tests given to master's students during three academic years (2009–2012) at the University of Alcalá, Madrid, Spain. Its main objective is to compare trainees' test performance at two different points in time. The study analyzes the degree of accuracy and the speed of response, considering the variables of mother tongue, gender, age, and undergraduate education. Our customized tests drew upon two aptitude tests developed by Pöchhacker (2009) and Russo (2009) and combine oral-aural exercises with tasks requiring listening skills, expressional fluency, and public-service-setting terminology. The tests are administered in a 24-seat multimedia lab, which allows recording students' performance for further evaluation. The results show the validity of the tests (Baker, 1989) to measure the students' aptitudes before and after training, and thus the tests prove to be useful tools to predict professional performance as well.

Key Words: interpreting education, competences, aptitude test, public service, assessment, evaluation

¹ Correspondence to: carmen.valero@uah.es

Assessment and Evaluation in Labs for Public Services Interpreting Training

1. Introduction

The University of Alcalá (UAH), Madrid, Spain, has been training interpreters and translators for public service since 2001 in several language combinations. It has been a member of the European Master's in Translation (EMT) network since 2009. Its main training program is the MA in Intercultural Communication, Public Service Interpreting and Translation (PSIT; 60 European Credit Transfer System [ECTS] credits). The MA is structured in five modules: Interlinguistic and Intercultural Communication, Translation and Interpreting in Health Care Settings, Translation and Interpreting in Legal and Administrative Settings, Internship in public and private institutions; and a Master's Degree/Research project (see <http://www2.uah.es/traduccion>).

As members of the FITISPos-UAH² and FITISPos-E-Learning research groups, the authors of this study are interested in creating a repository of useful learning materials for students, and in the design and application of different assessment tools to evaluate students' skills. In this article, we report on and analyze the development and partial results of two bilingual interpreting tests given at UAH during three academic years. The tests are compulsory for two required courses in the MA program: Interpreting in Healthcare Settings (5 ECTS) and Interpreting in Legal and Administrative Settings (8 ECTS).

In our analysis, we focus specifically on assessing students' performance, taking as the initial reference point an aptitude test in health care settings the students take at the beginning of the on-site classes (health care is the first module in the program). The final reference point is an aptitude test in health care, legal, and administrative settings, which students take after some 325 hours of theoretical and practical specialized training in these settings; with the second test we assess students' competence acquisition and their readiness to enter the practicum module in real institutions. Our intention in this study was to measure the students' aptitudes before and after specialized training by comparing the results of the first test to those of the second test, which has a higher degree of difficulty.

We can describe our first *aptitude test* as a standardized test designed to measure our students' abilities (verbal comprehension, reasoning, and expressional fluency) to develop skills and acquire specific knowledge in health care settings. Our second test can be considered both an aptitude test and an *achievement test*: a standardized test designed to assess aptitude and knowledge in interpreting gained through education and training, as well as measure the students' abilities to develop skills and acquire specific knowledge in legal and administrative settings.

² FITISPos is the Spanish acronym for Training and Research in Public Service Translation and Interpreting. FITISPos group website: <http://www2.uah.es/traduccion>

Assessment and Evaluation in Labs for Interpreter Training

2. Literature Review

In developing our aptitude test, we considered previous approaches to testing. Russo (2009) concluded that the interpreting training field can benefit from aptitude tests, arguing that interpreting-related cognitive skills and verbal fluency can be both measured and predictive. Pöchhacker (2009) found that the aptitude test he designed was effective in measuring aural-oral language use proficiency and basic interpreting-related subskills. We followed Russo and Pöchhacker's recommendations (presented at the Symposium on Aptitude for Interpreting held in Antwerp, Belgium, in May 2009) in creating an effective screening tool for our own program and useful reference points to aid our future professional interpreters in their learning process.

Developing an assessment instrument requires analyzing four main factors: the competences and skills to be assessed, the scales and grading method to be applied, the reliability and validity of the test, and the different types of exercises to be used in the test. Following, we look at each of these factors.

Competences and skills. In addition to mother-tongue competence, a professional interpreter should possess an array of other competences and skills. Although there is not yet an established standard set of parameters to measure a candidate's skills, according to Pöchhacker (2004), there is consensus regarding the nature and extent of the abilities to be demonstrated on entry into a training program.

In this respect, Schaeffner (2000) states that the process of translation involves at least the following specific competences, which we consider basic, necessary, and relevant also for interpreters:

- Linguistic competence of the languages concerned,
- Cultural competence,
- Domain- or subject-specific competence,
- (Re)search competence, and
- Transfer competence. (p. 146)

Pöchhacker (2004) includes the following competences in the profile of professional interpreters (here referring specifically to conference interpreting): general knowledge, cognitive skills (analysis, attention, and memory), and personality traits (stress tolerance and intellectual curiosity). When he discusses dialogue interpreting (the most common type used in PSIT), he also includes note-taking, whispered simultaneous interpreting, intercultural communication, turn-taking, and role performance.

Although all of the aforementioned competences should be developed in students, we designed our tests to measure only linguistic, domain, and transfer competences—those noted by Schaeffner—and cognitive skills, stress tolerance and note-taking competences—those noted by Pöchhacker. The first module of our program trains and evaluates students in cultural competence; (re)search competence is trained and evaluated in the subsequent modules

Scales and grading. We used an ordinal scale suggested by Sawyer (2004, p. 105): *high pass* = 75 and higher, *pass* = 75–50, *borderline fail* = 50–25, and *fail* = 25 and lower. Grading is done by giving the first two exercises 12 marks each, the third exercise 15 marks, the fourth exercise 20 marks, the fifth exercise 16 marks, and the seventh exercise five marks.

Validity. We agree with Sawyer's argument (2004) that a central concern of testing should be the need to conduct reliability and validity studies and to foster greater awareness of the role of professional judgment in assessment practices. Internal validity is ensured through content validity, based on the extent to which our tests reflect the public service interpreting domain.

Types of exercises and the grading systems to be used in the test are explained in detail below in Section 4.

3. Method

The aptitude test measures aptitude prior to entering a training course; based on the test results, trainers adapt the content of the syllabus to the students' individual characteristics and needs. The

Assessment and Evaluation in Labs for Interpreter Training

achievement test measures students' acquired skills and competences after completion of the third and last module of the course and determines placement.

Both tests are given in a multimedia lab in which trainers can assess and record 24 students at a time and easily collect results to be evaluated later. The lab setting allows us to assess a higher number of candidates in a shorter period of time, but because the recorded test cannot be paused or stopped, we cannot assess turn-taking and role performance skills.

The structure of the assessment instrument is similar in both tests: a vocabulary exercise, a synonyms and antonyms exercise, a comprehension and summary exercise, a short consecutive interpreting exercise, a cloze exercise, a sight translation exercise, and a short interview. We keep this structure for both tests to help students concentrate and focus better, so that they can recall primed vocabulary and anticipate accurate subsequent expressions. The exercises are ordered in a logical language development cycle, with the degree of difficulty increasing progressively from Exercise 1 to Exercise 4. Exercise 5 would probably be better placed as number 4 because it deals with anticipation, a useful skill for interpreting. But a cloze exercise is also helpful to prime a sight translation, because it activates the mental capacity to structure grammar patterns correctly. It also allows decreasing the degree of difficulty between the two interpreting-specific exercises, Exercise 4 and Exercise 6.

The basic approach of the instruments is criterion-referenced, measuring performance against a known criterion. "It is a more meaningful approach; given the need for the interpreter to perform adequately in all situations, they should be judged against a scale of absolute criteria" (Arjona, 1984, cited in Sawyer 2004, p. 115). The result the instruments generate is a qualitative description of performance and a numerical score based on the objectives. The reporting mechanisms are feedback for the candidates, the instructors, and the department.

The results are given to all the interpreting trainers so that they can use them to inform the creation of their syllabi. After the first test, students receive their grades and comments individually, although the general results of the group are analyzed in class so that everyone can take measures to improve their performance. Students are grouped into threes according to their results, so that they can help one another. The results of the second test are given to students together with the final grades of the interpreting course, right before they start the practicum.

As mentioned, the tests were mandatory for all students in the academic years 2010–2011 and 2011–2012; they were optional in the first academic year they were given, hence a bigger difference in the number of test-takers in that year.

The authors of this study were the tests designers. One author teaches a practical interpreting course in the Spanish–English language pair; this author is also one of the two test administrators and one of the two test graders in this language pair.

As a general fact, approximately 30% of the students who enter the MA course have an undergraduate degree in Translation and Interpreting (T&I); approximately 45% of the students have a degree in Language and Literature; and approximately 25% of the students have undergraduate degrees not directly related to languages.

The competences and subskills the tests were designed to assess are shown in Table 1.

Table 1: Competences and subskills assessed by each exercise

Exercise	Vocabulary	Synonyms & Antonyms	Verbal Reasoning	Consecutive Interpretation	Oral Cloze	Sight Translation	Interview

Assessment and Evaluation in Labs for Interpreter Training

Competence	Linguistic, Domain, Transfer	Linguistic, Domain, Transfer	Linguistic, Domain	Linguistic, Domain, Transfer	Linguistic, Domain	Linguistic, Domain, Transfer	Linguistic
Subskill	Cognitive, Stress	Cognitive, Stress	Cognitive, Stress	Cognitive, Note-taking, Stress	Cognitive Stress	Cognitive, Stress	Cognitive Stress
Indicator	Accuracy, Speed	Accuracy, Speed	Accuracy, Speed	Addition, Speed, Self-correction, Omission, Hesitation	Anticipation, Speed	Addition, Hesitation, Self-correction, Omission, Speed	Coherence, Cohesion

4. Instruments

4.1 Aptitude test in health care settings

The tests were translated and adapted to the seven language-pair combinations in which the MA is offered (Spanish and Arabic, Chinese, English, French, Polish, Romanian, and Russian); for the purposes of this article, we use the Spanish–English language combination test as an example.

The aptitude test consists of seven exercises of different types designed to measure different competences, as explained below. For each exercise, we provide a short description focused on aspects such as objectives, scoring, specific elements, and required skills.

Before every exercise, the recording includes a short fragment of classical music. The objective is twofold: to help the students relax, so that they forget previous thoughts and thus connect with the new topic; and to help as a stressor to increase their level of anxiety. Initially the music lasted 8–10 seconds and included selections from different composers, with different degrees of intensity. As the music served more the second goal than the first, we reduced the time to 5 seconds and played music from just one instrument and composer. (Five seconds is the exact time available to answer all of the exercises except for Exercise 4.

Exercise 1 (Table 2) is a vocabulary exercise in which students must provide a word or a short phrase to interpret the terms. The terms belong to different syntactic categories. The exercise is composed of 12 Spanish words and 12 English words that are read/heard in alternating languages. Half of the 24 words are in common use and the other half are more specific, thus increasing the level of difficulty. The audio is recorded by native speakers of Spanish and English. For scoring, a half point is assigned to every word, which is then discounted for functional inaccuracies.

Table 2: Vocabulary exercise

Question	Answer	Question	Answer
1. Cicatriz	1.	1. Next of kin	1.

Exercise 2 (Table 3) is a synonyms and antonyms exercise composed of 12 Spanish words and 12 English words. Half of the 24 words are in common use and the other half are more specific, increasing the level of difficulty. The synonyms are all read/heard in alternating languages. Candidates are required to provide a synonym (similar or identical) for every term. Then the procedure is repeated with the antonyms. The terms belong to different syntactic categories.

The task here is not to interpret the terms, but to provide synonyms or antonyms. The exercise is challenging because it follows an exercise based on interpreting into the second language; because the previous exercise lasts for a considerable period of time, candidates tend to continue to interpret rather

Assessment and Evaluation in Labs for Interpreter Training

than provide a synonym to the term. Tests graders consider only a synonym or an antonym as a valid answer. Direct interpretations are marked as wrong answers.

For scoring, a half point is given to every word, which is then discounted for functional inaccuracies.

Table 3: Synonyms and antonyms exercise

Question	Answer	Question	Answer
Synonym		Antonym	
1. Misapprehend	1.	1. Estima	1.

Exercise 3 is a verbal reasoning and summary exercise consisting of 111 English words about a common health topic, cardiovascular disease. The first task is a comprehension question with four sentences and three options (*true, false, or not enough information*). Students are prompted when to start talking by means of tones. The degree of difficulty increases in this exercise with the appearance of figures and more specific medical terminology.

Students then must summarize the listening exercise in 60 seconds, which measures comprehension, memory, vocabulary, and fluency. By giving such a short period of time for this task, we are adding pressure, which causes anxiety. Test administrators observed and reported signs that could indicate students' stress levels increasing when no tone was heard at the end of the sentence. Students who do not perform well under pressure tend to rush, make more mistakes, and not finish the phrases appropriately, or they freeze and say nothing. Only a few students summarized at least the main ideas of the audio, despite being also influenced by the lack of tone, their classmates' discontented expressions, and loudly expressed errors.

For scoring, two points are assigned to every sentence and seven points to the summary. Points are taken off for grammar mistakes, lack of coherence, incompleteness, and fluency breakdown. The tasks are the following:

Say whether the following statements are true, false or whether there is not enough information. You can speak after the beep.

Make a summary of the verbal text you just listened to in sixty seconds. You can speak after the beep.

Further increasing the level of difficulty, Exercise 4 is a short consecutive interpreting exercise, the first one for which interpreting skills are really required, with 313 words to be interpreted. The audio is a doctor-patient interview related to an arm fracture, of which we reproduce a fragment below:

Patient – Will it hurt, doctor?

Doctor – No, estará bajo el efecto de la anestesia y no sentirá absolutamente nada.

Patient – Should I know anything about the anaesthetic before taking it?

Students are instructed to interpret the dialogue using the consecutive mode and may take notes. They are prompted when to start talking by tones. Two elements of the exercise increase the degree of difficulty: The nonnative speaker hesitates frequently when using confusing medical terms, and there is a short fragment deleted almost at the end of one of the doctor's explanations. This functions as another stressor and makes most of the students lose the information given after the deletion.

To score the exercise, the trainer takes into account omissions, editions, additions, speed of response, and accurate renditions, as well as rhythm, intonation, and the appropriate use of the specific terminology. Twenty points are assigned to this exercise.

Assessment and Evaluation in Labs for Interpreter Training

Exercise 5 is a cloze exercise divided into two texts. The first one is a repetition of the English text used in Exercise 3, consisting of a total of 111 words with eight gaps. The second part is a Spanish text about a topic related to the previous one (cardiovascular disease), with 113 words and eight gaps. All the deletions in both texts belong to different syntactic categories.

The degree of difficulty increases in this exercise due to the presence of gaps to be anticipated and also to the use of more specific medical terminology. The short time slot allowed for students to provide an answer also adds stress. To perform appropriately in this kind of exercise, students need a good memory, concentration, imagination, and speed of response. To score the exercise the trainer gives a point to every correct answer. Below are two short fragments in both languages:

- *English: Similarly, high-pressure, stressful work, even where it does not involve physical activity, should also be (avoided).*

- *Spanish: Es la causa más importante por la cual las personas sufren (ataques cardiacos).*

Exercise 6 is a sight translation of two texts. The first one, 195 words, in English, relates to the measurement of blood sugar levels. It is sent by the trainer to the students' computer screen right after they hear the piece of music and the title of the exercise—students do not need to operate the computers; they only read and interpret when they feel ready.

The degree of difficulty increases here due to the constant use of acronyms. Various methods have been used for this exercise in different academic years and language pairs, such as “launching” the text to a large screen at the end of the lab, making it difficult to see, or giving the students a low-quality printed copy of the text and suggesting they can either write on it or underline some useful necessary phrases. Writing distracts students' attention and takes more time to go through the whole text. The instruction to start sight translating is given once the first student shows signs of having finished reading. All of the methods have proven to increase students' stress levels.

Once all the students have finished the translation, the next text appears on their screens. The second text, 209 words, in Spanish, is about informed consent. The degree of difficulty increases due to the Spanish writing style, which uses many subordinate clauses, and due to its length; it is a long text to deal with after more than 20 minutes of interpreting activity.

Twenty points are assigned to this exercise, 10 to each text. An example is given below:

1. *Read the following texts and translate them at sight.*

(...) Blood will be obtained by sticking the finger with a fine-point needle.

Finally, Exercise 7 is a short interview consisting—initially—of 14 questions. Six are closed questions; seven are questions requiring specific information; and the last one is an open-ended question to obtain as much information as possible about the students' feelings and impressions on the test. This exercise is used as a qualitative research method to collect ethnographic data as well as feedback on the students' own performance and on the test. The number and type of questions have been adapted and changed every year to obtain different information.

The questions go from simple to a higher degree of complexity, from asking students to state their personal data to having them provide information that shows their language competence and intercultural capacity. The questions go from professional details to personal expectations and end by asking about the test. This exercise helps to, on the one hand, reduce stress, because students feel they control the situation when speaking about themselves or analyzing the test. On the other hand, it maintains a certain level of anxiety when students have to assess their own performance.

Five points are assigned to this exercise to evaluate the students' expression of the response rather than its content.

1. *Please, tell us all your impressions about this test?*

Assessment and Evaluation in Labs for Interpreter Training

4.2. Achievement test in legal-administrative settings

The structure of the achievement test given at the end of the program is similar to the initial aptitude test. It differs mainly in its content, which is specific to the legal-administrative setting. (See Valero-Garcés & Socarrás, 2011, for a deeper analysis of this test).

Another characteristic that differentiates this test from the previous one is the response-time limit for students. It was initially the same (5 seconds), but was changed to only 3 seconds. At this point students have more vocabulary related to different settings, so they are given less time to react. They are expected to cope better with stress and to recall the studied vocabulary in a very short period of time.

Exercise 1 (see Table 4) is a vocabulary exercise composed of 20 Spanish words and 20 English words related to the legal setting, which are read/heard in alternating languages. Of these words, 50% are in common use and the other 50% are more specific, thus increasing the level of difficulty.

Table 4: Vocabulary exercise, legal-administrative setting

Question	Answer	Question	Answer
1. To Acquit	1.	1. Catastro	1.

Exercise 2 (see Table 5) is a synonyms and antonyms exercise composed of 16 Spanish words and 14 English words related to the legal setting. Of these, 50% are in common use and the other 50% are more specific, so increasing the level of difficulty. The synonyms are all read/heard first in alternating languages. Candidates are required to provide a synonym (similar or identical) for every term. The degree of difficulty is increased as some Spanish words are given consecutively, which breaks the normal pattern and causes confusion. Students tend to fail even when most of the words are not difficult terms. The antonyms are all read in alternating languages.

For scoring, a half point is assigned to every word, then discounted for functional inaccuracies.

Table 5: Synonyms and antonyms exercise, legal-administrative setting

Question	Answer	Question	Answer
Synonym		Antonym	
1. Refutar	1.	1. Revocar	1.

Exercise 3 is a verbal reasoning exercise consisting of 251 English words about a topic common to (im)migrants, Spain's Residence Permit Application. The first task is a comprehension question with four sentences and three options (*true, false, or not enough information*). The second task is for students to summarize the listening exercise in 60 seconds, which measures comprehension, vocabulary, and speaking fluency. The degree of difficulty increases in this exercise with the appearance of dates, acts, and laws, not common linguistic borrowings, and informal language examples. Tones prompt candidates when to start talking.

For scoring, two points are assigned to every sentence and seven points to the summary. Points are taken off for grammar mistakes, lack of coherence, incompleteness, and fluency breakdown.

Say whether the following statements are true, false or there is not enough information. You can speak after the beep.

Make a summary of the listening in sixty seconds. You can speak after the beep.

Assessment and Evaluation in Labs for Interpreter Training

Exercise 4 is a short consecutive interpreting exercise with an introduction of 80 words and 680 words to be interpreted. The text is a Direct Examination of the proceedings of a trial. Students are instructed to interpret the dialogue using the consecutive mode and may take notes. They are prompted to start talking by means of tones. The degree of difficulty increases in this exercise for two reasons: the introduction has many proper names and figures and students start taking notes to interpret, but it is not necessary. By the time they realize this, the questioning has already started and students are surprised. Some students take their headsets off; others hit the table in disbelief.

To score the exercise, the trainer takes into account omissions, editions, additions, speed of response, and accurate renditions, as well as rhythm, intonation, and the appropriate use of the specific terminology. Twenty marks are assigned to this exercise.

Q. Y para que quede claro ¿en el Centro de Maltrato Doméstico, usted trabaja no solo con las víctimas de la violencia doméstica, sino que creo que ha testificado con maltratadores también?

A. Yes. I co-facilitate that group.

Exercise 5 is a cloze exercise of 313 words about a topic related to the listening used in Exercise 3. The exercise is divided into two parts. The first one is a listening exercise in English (about working permits in Spain) consisting of a total of 153 words and 10 gaps. The deletions belong to different syntactic categories: one verb, eight nouns, and one adverb.

The second part is a listening exercise in Spanish about a topic related to the previous ones (ways to work in Spain) with 158 words and 13 gaps. The deletions belong to different syntactic categories: one verb, 10 nouns, two adverbs, and one gap that does not need completion but is meant to check on students' concentration and also works as a stressor. The degree of difficulty increases here due to the use of dates, phrases in the second language, and the presence of gaps to be anticipated. The short time slot allowed for the subjects to supply an answer adds stress as well as difficulty. The second listening exercise is also more difficult because it deals with specific vocabulary related to governmental guidelines and regulations.

To perform appropriately in this kind of exercise, candidates need a good memory, concentration, imagination, and speed of response. The use of specific terminology in the texts also contributes to the increase in the degree of difficulty. To score the exercise, the trainer gives a point to every correct answer. A fragment of the text can be seen below:

Should you lose your employment and have contributed to the Spanish social security system whilst working, you will also be entitled to unemployment (benefit).

El contingente podrá establecer un número de visados para búsqueda de (empleo) dirigidos a hijos o nietos de español de origen.

Exercise 6 is a sight translation exercise with 349 words in the main text and 49 words in a supporting footnote. This text is related to the texts used in the previous exercises. The text appears on the student's screen right after they hear the piece of music and the title of the exercise. The degree of difficulty increases here due to the constant use of acronyms and formal language—the text is part of the documentation used by an official institution—in contrast with the previous texts that were less formal; in addition, it is a long text after more than 40 minutes of interpreting activity. Twenty points are assigned to this exercise, as follows:

Read the following text and translate it at sight. You can take some minutes to read it, and then start sight translating when you feel ready.

Exercise 7 is a short interview consisting of 14 questions asked in English. Six are closed questions, seven are questions requiring specific information, and the last one is an open question to obtain as much information as possible about the students' feelings and impressions on the test. It is used as a qualitative

Assessment and Evaluation in Labs for Interpreter Training

research method to collect ethnographic data as well as feedback on the test. Five points are assigned to this exercise. See an example below:

Do you consider this test an important part of the Continuing Professional Development program?

5. Results and Discussion

In total, 74 students have taken the test for health care settings and 63 students have taken the test for legal settings (see Table 6). Women represent a big majority, 86.7%. The average age was 25–26 in the 3 years. In the academic year 2009–2010, 10 students decided not to take the legal test because it was optional. The two tests were mandatory in the following years. Students come from six to eight different countries every year and they have very varied undergraduate studies (9–14).

Table 6: Aptitude tests application to MA students in Spanish–English groups

A. Year	Test Health	Test Legal	Women	Men	Nationality	B. Studies
2009-2010	23	13	20	3	6	11
2010-2011	27	28	25	3	8	14
2011-2012	24	22	20	4	6	9
Total	74	63	65	10		

5.1 Comparing Academic Years 2010–2011 and 2011–2012

Comparing the average marks of the two groups of students who took the tests in the last two academic years reveals similar results (see Table 7): The performance of both groups improved after receiving training.

Table 7: Aptitude and achievement tests: Spanish–English groups' final marks

Group Mean Marks (100)		
A. Year	Health Test	Legal Test
2009-2010	-	-
2010-2011	50,03	56,75
2011-2012	48,07	56,21
Total	49,05	56,48

We selected a random sample of nine students from the last two academic years to compare their results in both tests. They represent the two genders, various nationalities, and different undergraduate studies. As we can see in Table 8, representing the academic year 2010–2011, seven students improved their results in the second test. One student maintained his/her results, and one student obtained a lower result in the second test.

Table 8: Aptitude tests sample comparison Spanish–English 2010–2011.

Assessment and Evaluation in Labs for Interpreter Training

Group 2010-2011			
Nationality	U. Studies	Health Test	Legal Test
Esp	En Teacher	73	79.2
Esp	T&I	72.5	77.5
Esp	T&I	70	70
EU	Sp-En Filology	54	59.6
Esp	En Filology	53.5	50.2
Ita	En Filology	53	56.5
Esp	Turism	40	41.6
P.Ri	Fr Filology	39	52.6
Esp	En Filology	38.2	53.4

During the academic year 2011–2012, the results were similar (see Table 9); eight students improved their results whereas only one obtained a lower result in the second tests.

Table 9: Aptitude tests sample comparison, Spanish–English, 2011–2012

Group 2011-12		
Nationality	Health Test	Legal Test
Spa	69,5	78
Spa	62,5	69,5
Spa	60,5	67,5
Pol	48	56,5
Spa	46,5	53,5
Ven	45,5	53,5
Fra	54,5	45,5
US	42,5	45
Can	43	44,5

5.2 Comparing Students' Performance During Academic Year 2010–2011

To illustrate the analysis we carry out to assess every student's performance evolution, we have selected a random sample of nine students from the academic year 2010–2011. We show their results in both tests: first the aptitude or health care test and then the achievement or legal test.

For the analysis, we group the exercises according to the main skills they assess. Thus Exercises 1 and 2 are grouped together as language-related exercises, Exercises 3 and 5 are grouped as oral comprehension exercises, and Exercises 4 and 6 are grouped as interpreting-related exercises.

We compare speed of response with accuracy in Exercise 1 (vocabulary) and Exercise 2 (synonyms and antonyms; see Table 10). It can be noticed in the second and third columns that students are faster and more accurate when interpreting Exercise 1.

Table 10: Aptitude test sample comparison: Speed of response and accuracy in language exercises, Spanish–English, 2010–2011

Assessment and Evaluation in Labs for Interpreter Training

Exercises 1 & 2 Language				
	Exercise 1		Exercise 2	
Health Test	Speed (Sec)	Accuracy %	Speed (Sec)	Accuracy %
73	1.6-1.4	91.7	2.5-2.3	75.0-83.3
72.5	1.4-1.2	83.3	2.3-2.1	75.0-83.3
70	1.5-1.1	83.3	2.5-2.1	66.6-83.3
54	1.4-1.8	91.6-75	3-2.1	41.7-62.5
53.5	1.5-1.7	83.3	3.1-2.2	62.5
53	1.9-1.8	83.3-75	3.1-2.5	41.7
40	1.6-1.3	91.7-79.2	3.0-3.3	66.7-25
39	1.7-1.3	91.7-79.2	3.0-3.8	62.5-25
38.2	1.7-1.4	83.3-79.2	3.1-3.7	58.3-33.3

As shown in Table 11, we can see that students are also faster and more accurate when interpreting Exercise 1 of the legal test. The reaction time is now higher than in the health care test, however. We consider it might be due to different reasons: Now students have a larger vocabulary, so the recall/retrieval processes take longer; the exercise includes terminology from three different settings, which affects the priming effect for they do not have cuing keywords; and the exercise has many more words. Also the degree of accuracy is lower than in the health care test. We argue that the specific terminology of legal settings is more complex and it poses a bigger challenge for students who have only studied it for a couple of months

Table 11: Achievement test sample comparison: Speed of response and accuracy in language exercise, Spanish-English, 2010-2011

Exercises 1 & 2				
	Exercise 1		Exercise 2	
Legal Test	Speed (Sec)	Accuracy %	Speed (Sec)	Accuracy %
79.2	2.1	59.7	2.4-2.3	33.6-40
77.5	1.9	57.5	2.0-2.1	25-40
70	2.0	55.8	2.3-2.1	25-33.6
59.6	2.3	55.8	2.5-2.0	40
50.2	2.5	50.2	3.1-2.9	25
56.5	2.3	55.8	2.4-2.3	33.6
41.6	2.2	50.2	2.5-2.7	25
52.6	2.0	55.8	2.3-2.4	25
53.4	2.3	57.5	2.1-2.4	33.6

In oral comprehension exercises, students show the highest improvement in performance in the health care test; they are both faster and more accurate in Exercise 5 (see Table 11).

Assessment and Evaluation in Labs for Interpreter Training

Table 12: Aptitude test sample comparison: Speed of response and accuracy in oral comprehension exercises, Spanish-English, 2010-2011

Exercises 3 & 5 Oral Comprehension				
	Exercise 3		Exercise 5	
Health test	Speed (Sec)	Accuracy %	Speed (Sec)	Accuracy %
73	1.7	50	0.7-1.4	100-87.5
72.5	1.7	50	0.5-1.2	100-87.5
70	1.6	50	0.6-1.3	100-87.5
54	1.4	75	1.4-1.7	87.5
53.5	1.5	50	1.4-1.1	87.5
53	1.7	50	1.4-1.6	87.5
40	1.4	25	1.6-1.5	68.7-87.5
39	1.0	25	0.9	68.7-87.5
38.2	1.3	25	1.5-1.4	68.7-87.5

Table 13 shows that students have developed their comprehension skills by lowering their reaction times and increasing their degree of accuracy in Exercises 3 and 5 of the legal test.

Table 13: Achievement test sample comparison: Speed of response and accuracy in oral comprehension exercises, Spanish-English, 2010-2011

Exercises 3 & 5 Oral Comprehension				
	Exercise 3		Exercise 5	
Legal test	Speed (Sec)	Accuracy %	Speed (Sec)	Accuracy %
79.2	1.2	75	1.3-0.9	87.5-100
77.5	0.8	75	1.2-0.7	100-87.5
70	0.7	50	1.1-0.6	100-83.5
59.6	0.8	50	1.0-1.3	100-85.5
50.2	1.3	50	1.4-1.1	87.5-62.5
56.5	1.0	50	1.2-1.3	87.5

Assessment and Evaluation in Labs for Interpreter Training

41.6	1.3	25	1.4-1.3	62.5
52.6	1.1	50	1.1-1.0	87.5
53.4	0.9	75	1.1-1.0	87.5

To analyze the results of the exercises directly related to interpreting skills and competences, we measure students' main errors such as additions, omissions, self-corrections, and hesitations (see Table 12). Additions are not included because they are not significant. Self-correction is seen as an error only when the correction is wrong or it is a repetition of a nearby student's expression. Students do not add too much information in Exercise 4, but they omit a great deal of the content. They also hesitate a lot, which shows their lack of interpreting strategies, information recall, and note-taking skills. In Exercise 6 they hesitate and correct themselves even more but omit much less information.

Table 14: Aptitude test sample comparison: Interpreting skills exercises, Spanish–English, 2010–2011

Exercises 4 & 6 Interpreting skills								
Health Test	Exercise 4				Exercise 6			
	Additions	Omissions	Selfcorrectio	Hesitation	Additions	Omissions	Selfcorrectio	Hesitation
73	1	5	-	6	-	-	10	10-12
72.5	2	5	-	5	-	-	8	7-9
70	2	6	-	5	1	-	9	8-9
54	2	8	-	9	2	-	14	15-16
53.5	3	7	1	7	1	-	16	16-13
53	3	8	-	8	3	-	15	16-16
40	1	19	1	8	3	4	16	15-13
39	-	17	3	7	1	2	14	13-14
38.25	-	20	3	5	2	5	17	17-18

After receiving training, students are already capable of correcting themselves in Exercise 4 (see Table 15). It means they are monitoring themselves and are thus aware of their interpreting process. They still add and omit much information, likely because they still have deficient note-taking skills.

Table 15: Achievement test sample comparison: Interpreting skills exercises, Spanish–English, 2010–2011

Exercises 4 & 6 Interpreting skills								
Legal Test	Exercise 4				Exercise 6			
	Additions	Omissions	Selfcorrectio	Hesitation	Additions	Omissions	Selfcorrectio	Hesitation
79.2	2	5	6	11	1	1	10	11
77.5	1	6	5	7	-	1	7	9
70	-	7	5	7	-	2	6	8
59.6	1	7	5	8	-	-	12	14
50.2	3	14	4	13	3	4	9	19
56.5	2	9	6	9	1	2	8	13
41.6	3	17	1	9	2	4	19	18
52.6	2	12	3	8	1	3	18	15
53.4	1	9	4	9	1	1	16	14

Assessment and Evaluation in Labs for Interpreter Training

5.3 Correlation: Accuracy, Speed of Response, and Personality Trait (Stress Tolerance)

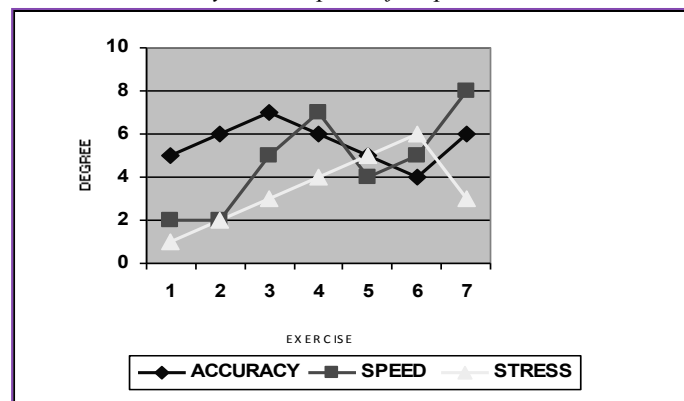
Students' performances in both tests show similar curves in the correlation of accuracy, speed of response, and stress tolerance (see Table 16), although they improved their accuracy and speed of response while their stress decreased considerably. We analyze each aspect separately across the same test.

Accuracy: Students achieved medium accuracy in Exercise 1, which gradually increased until Exercise 3. Accuracy then gradually decreased in the next three exercises, mainly due to the fact that those are the interpreting-related exercises. In Exercise 7, students regained some control, and their expression is more coherent and fluent.

Stress: We found that students arrive at the lab for both tests in a typical nervous state, and little by little their stress level increases. Their body language changes, and we observe that students make more changes or corrections as they are working through the test as well as hesitate more frequently. Students' renditions tend to be less accurate 20 minutes into the tests.

Speed of response: Students began the tests with very low speeds of response, but speed increased with the rhythm marked by the "speakers" and with the students' realization that they knew many of the words/phrases. Students had similar speeds of response compared with their own performance in Exercises 1 and 2. In Exercise 3, their speed of response increased. In Exercise 4 their speed of response gradually increased, which might be due to their getting used to the vocabulary. In Exercise 5 their speed of response was much slower, although in the second part they did better. In Exercise 6 their speed of rendering was low at the beginning, but it increased toward the middle of the text, and it ended at a medium level. In Exercise 7 their speed of response increased considerably.

Table 16: Accuracy versus speed of response versus stress level



In general, most of the 75 students who took the tests in the 3 academic years showed a good level of knowledge both in the general language and in the specific terminology. Students with a degree in Translation and Interpreting (T&I) showed a higher level of knowledge than the others; they also were more skilled in dealing with the specialized exercises (interpreting and sight translation) and showed better short-term memory skills.

Students with a degree in T&I tried harder to render more complete messages and were therefore more coherent and comprehensible. They were also more faithful to the original message, amending their own errors when necessary.

Students with a degree in Language and Literature performed better than those who had degrees in non-language-related fields. Women performed better than men, although male students showed a higher stress tolerance. Few differences were found regarding age because the majority of students were of similar age, but younger students performed better than older ones. Spanish speakers performed better than non-Spanish speakers.

Assessment and Evaluation in Labs for Interpreter Training

Students performed slightly better in the first two exercises in their second language than in their mother tongue. During the rest of the test, students performed better from the B language into their mother tongue. Few students had any experience, but those with some experience in the profession performed better than those who had none.

Examination of qualitative feedback: Most of the Spanish students had not lived more than a year in an English-speaking country, which may have affected their language proficiency and speed of response. Those students who lived in English-speaking countries were more proficient and reacted faster. All the students expressed their desire to be in the profession for a long time, and to become professional interpreters and translators in the public service sector. They considered the aptitude test to be a very difficult but useful test and said they would retake it if offered.

6. Conclusions

Following a review of the literature on interpreter training and assessment, we proposed and examined the four main features of an aptitude test: the competences and skills to be assessed, the scales and grading method to be applied, the validity of the test, and the different kind of exercises to be used in the test. The professional interpreter-translators we need in our multicultural societies today would ideally possess a combination of all of these competences and skills. Therefore, these criteria should be the core of our training and assessment in higher education institutions, although most of the competences and skills may well be learned and obtained through intensive practice during subsequent training courses.

We have described the two assessment instruments we designed and applied to master's students. The objective was first to determine how prepared the students were to start a postgraduate training course, then to prove the effect of that training on the students' competence acquisition. The first test focused on health care settings; the second focused on legal-administration settings.

Based on the results of the aptitude (achievement) test in health care and legal settings explained in the Findings section of this study, and after comparing them with the results of two academic years, we can confirm that the battery of exercises designed to measure the ability of our students to develop skills and acquire knowledge is a fair predictive instrument. Students with a high level of performance in the first aptitude test obtained a remarkably better result in their achievement test; other students either improved a little or maintained their performance level.

In showing that most of students improved their performance from one test to the other, our data support the hypothesis that our students had abilities to develop skills and that training positively affected their acquisition of those skills.

6.1 Limitations

In closing, we acknowledge that these two assessment instruments need to be further tested for reliability and validity. Also, because we gave the tests in a lab we were able to use its technology to assess a higher number of candidates in a shorter period of time, but the assessment instruments are mainly based on lexicon and not on face-to-face performance.

Assessment and Evaluation in Labs for Interpreter Training

References

- Arjona, E. (1984). Testing and evaluation. In M. L. McIntire (Ed.), *New dialogs in interpreter education: Proceedings of the Fourth National Conference of Interpreter Trainers Convention* (pp. 111–138). Silver Spring, MD: RID.
- Baker, D. (1989). *Language testing: A critical survey and practical guide*. London, England: Edward Arnold.
- Pöchhacker, F. (2004). *Introducing interpreting studies*. London, England: Routledge.
- Pöchhacker, F. (2009, May). *Testing aptitude for interpreting: The SynCloze test*. Presentation at the Symposium on Aptitude for Interpreting: Towards Reliable Admission Testing. Lessius University College, Antwerp, Belgium.
- Russo, M. (2009, May). *Aptitude testing over the years*. Presentation at the Symposium on Aptitude for Interpreting: Towards Reliable Admission Testing. Lessius University College, Antwerp, Belgium.
- Sawyer, D. B. (2004). *Fundamental aspects of interpreter education*. Philadelphia, PA: John Benjamins.
- Schaeffner, C. (2000). Running before walking? Designing a translation programme at undergraduate level. In C. Schäffner & B. Adab (Eds.), *Developing translation competence* (pp. xvi, 244). Amsterdam, the Netherlands: John Benjamins.
- Schjoldager, A. (1996). Assessment of simultaneous interpreting. In C. Dollerup & V. Appel (Eds.), *Teaching translation and interpreting 3: New Horizons* (pp. 187–195). Amsterdam, the Netherlands: John Benjamins.
- Socarrás-Estrada, D., Valero-Garcés, C., & Vitalaru, B. (2012). Aptitude testing design and implementation in public service interpreting training. *Interpreting and Translation Studies*, 16(1), 271–300.
- Valero Garcés, C., & Socarrás, D. (2011). Development and assessment of an aptitude test for interpreters. Using labs for interpreting training. In T. Suau et al. (Eds.), *Interdisciplinarity and languages: Current issues in research, teaching, professional applications and ICT*. Vienna, Austria: Peter Lang.