

SELF-ASSESSMENT IN TRANSLATOR TRAINING

*Bryan J. Robinson, Clara I. López Rodríguez and María I. Tercedor Sánchez,
University of Granada, Spain
robinson@ugr.es*

Abstract

Self-assessment is a process that enables students to assess their own performance and greatly facilitates their acquisition of increased responsibility for their learning and performance. The advent of e-learning in the Spanish university system has provided opportunities to reorient translator training and accept new challenges that enhance the quality of undergraduate studies.

The present study demonstrates that the benefits of this new mode of instruction are especially appropriate to translator training. We describe the results of centring e-learning course design on self- and peer-assessment combined with tutor-moderation through the application of a rating scale of criterion-referenced descriptors. Our data demonstrate that the quantitative outcomes of learning remain constant. Furthermore, we include a range of qualitative data which indicate that students have a positive perception of the new learning and assessment processes, suggesting that these are responsible for increasing their awareness of important aspects of the translation process.

Key words: Spain: translator training; e-learning; self-assessment; peer-assessment; tutor moderation; criterion-referenced descriptors; tertiary education.

Introduction¹

Bridging the gap between e-learning and telework

Professional translating is prototypical telework (Castells 1996, Olvera Lobo et al. 2005 and In press): translators work in or for agencies for clients all over the globe; they may seldom have personal contact with the authors or users of the texts they create and may, indeed, have little contact with colleagues involved in a particular project. Consequently, the Information and Communications Technology skills required must form an integral part of undergraduate translator training in order to prepare students adequately for their future. To this end, it seems more than logical that university translation course modules should be taught in an online environment that models as closely as possible the realities of the professional world (Olvera Lobo et al. 2005, Alcina Caudet 2002).

Furthermore, because translators are increasingly called upon to work with as wide a range of linguistic combinations as possible in order to guarantee their employability, and since the traditional barriers on directionality ('Thou shalt only translate into thy mother tongue') have been trampled into the dust, the ability to self-assess translation quality is vital. Translators must be able to assess the quality of their work with the highest possible degree of objectivity. In this context, Martínez Melis and Hurtado Albir (2001: 281) stress the importance of classifying translation errors using a metalanguage that contributes to a better understanding of the nature of those errors and the cognitive processes involved in their generation. These authors point out that in translation error classification it is necessary to take into account the difference between: (i) errors relating to the source text (ST) and errors relating to the target text (TT), (ii) functional errors and absolute errors, (iii) systematic errors (recurrent) and random errors (isolated), and finally, (iv) errors in the product and errors in the

process.

Since the Declaration of Bologna in 1999, the process of convergence in European universities has highlighted a series of skills that are considered priorities in the academic, professional and personal development of students (Cózar Sievert 2003). Among these skills we find proactive learning, teamworking and quality assessment, all of which are developed to a higher degree through e-learning than in the traditional Spanish university classroom. In contrast with the traditional translation classroom, where knowledge would only pass from teachers to students, in this new teaching context, teachers act as facilitators in the learning process, and the classroom is not separated from the real world. The translation task is carried out under real-world conditions: there are clients and experts in the subject field, and students work in groups and rotate their roles. In the translation classroom, these principles have been applied for some time in approaches derived from social constructivism and collaborative learning (Király 2000, González Davies 2004, Olvera Lobo et al. 2005 and In press, Robinson 2003).

To achieve the highest possible degree of student involvement in their learning process, enhance their sense of responsibility for their own work and ensure the course provides students with adequate preparation for the professional world they are soon to enter, we have chosen to design this course around the concepts of self- and peer-assessment and tutor-moderation. To this end, we have structured the assessment procedures around a rating scale of criterion-referenced descriptors (See Table 1 and the Method section later in this paper).

The use of rating scales is frequent in the holistic approach to assessment, which aims to offer in a single score an overall impression of the quality of student work (Waddington 2001). They all consist of a numerical scale along which each number or range of numbers is labeled with the description that it represents. We consider that the most appropriate model for evaluating final year students is a holistic method. In this way, translation tasks that do not need a final revision will obtain a higher mark, whereas those that need thorough revision will fail. Rating scales are transparent, as they are normally available prior to the start of the learning process, meaning that all participants, students and tutor, know from the beginning how to assess quality. The initial learning objective of the course is to negotiate agreement on the interpretation of the criterion-referenced descriptors. All participants debate and discuss the criteria in order to reach mutual agreement on their meaning. Criterion-referencing is the means by which "performance is interpreted in relation to predetermined criteria. Emphasis is on attainment of objectives rather than on scores, and the descriptors are summaries of the degree of proficiency or type of performance expected in relation to a particular score" (ALTE 1998: 140, 142 and 146). Thus, the rating scale is the essential tool by which all performance is openly defined, debated, negotiated and measured.

This approach distinguishes our course from other University of Granada courses currently offered online, in which self-assessment takes the form of quizzes, often using specially developed software such as the HotPotatoes application (www.hotpotatoes.info). While this software is used in our course, it is limited to providing students with immediate feedback on their learning in the form of multiple choice, cloze or fill-the-gap activities. HotPotatoes is a valuable

Table 1. Criterion-referenced rating scale (adapted from Robinson 1998)

DECODING		ENCODING	
A. Content	B. Register, vocabulary, terminology	C. Translation brief and orientation to target text type	D. Written expression
0	The text fails to meet minimum requirements	The text fails to meet minimum requirements.	The text fails to meet minimum requirements.
1-2	Comprehension limited. Major content errors. Major omissions of ST content.	Choice of register inappropriate or inconsistent. Vocabulary limited, with some basic errors. Limited awareness of appropriate terminology.	Little or no evidence of orientation to TT type: formal or literal translation. Limited. Errors in basic structures.
3-4	Comprehension adequate. Minor content errors. Some omissions of ST content.	Choice of register occasionally inappropriate or inconsistent. Occasional mistakes of basic vocabulary. Clear awareness of appropriate terminology although some errors.	Ineffective. Errors in complex structures. Mistakes in basic structures.
5-6	Comprehension good. Minor omissions of less relevant ST content. Over- or under-translation distorts ST content or results in ambiguity	Choice of register mostly appropriate and consistent. Vocabulary effective despite mistakes. Terminology appropriate despite occasional errors.	Clear orientation towards TT type: appropriate use of TT type rhetorical devices
7-8	Comprehension very good. Over- or under-translation does not distort ST content or result in ambiguity.	Choice of register appropriate and consistent. Vocabulary effective despite occasional mistakes. Terminology appropriate despite mistakes.	Effective production of TT type: consistently appropriate use of many TT type rhetorical devices, with occasional errors
9-10	Comprehension excellent. ST content, including subtle detail, fully understood.	Choice of register consistently effective and appropriate. Sophisticated, highly effective choice of vocabulary. Terminology appropriate and wholly accurate.	Effective, sophisticated production of TT type, with few or no mistakes. Sophisticated. No errors. Almost mistake-free.

tool in that it allows the tutor to present a text with questions and answers online and to give students the opportunity to take the test a number of times until they are satisfied with their personal results. They can even request hints as to how they should improve their performance. Scores are recorded and available to students and tutor via the e-learning platform, but in our course they are used as another learning device rather than a part of the assessment.

The approach we describe in the present study implies a degree of risk for students: it is an innovation that may prove a bridge too far in that we are asking them to trust themselves in a role – that of translation reviser/tutor – which they are not normally expected to play. However, we believe it is essential they learn to accept the responsibility implied: once they have graduated, they will have no choice in the matter, so when and where better to learn?

Objectives

In this context, the current study evaluates an e-learning course module in which a rating scale of criterion-referenced descriptors is employed by students and tutor as the principle assessment instrument. The course establishes a translation quality assessment (TQA) routine in which students use self- and peer-assessment to score work produced individually or in teams. The scores they award to themselves and/or to their peers, and the underlying justifications for these scores are then moderated, i.e re-marked, by the course tutor.

The principal objectives of this study are

- To confirm, or not, that e-learning produces both quantitative and qualitative improvements in the learning of Scientific and technical translation from Spanish into English
- To test the current version of the rating scale of criterion-referenced descriptors (Tercedor Sánchez, López Rodríguez and Robinson 2005) adapted to Scientific and technical translation from Spanish into English
- To pilot the self- and peer-assessment procedures in an e-learning environment
- To assess the efficacy of the self- and peer-assessment by tutor moderation
- To evaluate student perception of criterion-referenced self- and peer-assessment.

E-learning at the University of Granada

The University of Granada Virtual Learning Centre, known by its Spanish acronym *Cevug*, has been working on the development of e-learning courses since 2001 (<http://cevug.ugr.es/web-cevug/index.php>). It provides the pedagogical and technical support for the conversion of current under- and post-graduate course modules into blended e-learning modules, that is courses that include both contact classroom teaching and online distance learning in subjects ranging from “Analysis, evaluation and training in sports technique” to “Binocular vision” and passing through “Audiovisual translation: localization, subtitling and dubbing” and “Scientific and technical translation”, taught by the authors of the present study.

The *Cevug* operates a strict selection process that involves teachers making

initial proposals for courses to be offered online (Cordón 2004). Courses that can be adapted to this mode of learning must be optional subjects, whether those pertaining to a specific degree program or free-choice options open to students of any discipline. Teachers make a written proposal documenting their motives for changing the course to an e-learning format as well as their experience and research achievements in e-learning. Proposals must receive the approval of the teachers' department and of the faculty in which the subject is to be taught prior to being accepted for processing by the *Cevug*. Course proposals are vetted internally, and teachers whose courses are accepted then receive training both in instructional material design, through an online course similar to that they will be preparing, and in the use of appropriate software (Macromedia Dreamweaver). Courses that have been prepared for e-learning are again assessed internally and then submitted to specialists at another university for external evaluation before receiving final approval. Approval is given for three years. Towards the end of the semester in which an online course is taught, the students participating are asked to complete an online questionnaire to assess their views of the materials and methodology of the particular subject. The results published by the *Cevug* offer a comparison of all courses taught in the same semester. We analyze the results for this study below, although we are aware that the value of the comparisons are limited: this module was taught at the same time as courses in the Science, Law and Education faculties, none of which used a model of assessment similar to ours.

The online learning platform used by the University is WebCT (www.WebCT.com). This platform provides all of the communication options needed for students to follow courses of this type, including access to course contents modules, e-mail, discussion and chat rooms, options to download and upload materials and to follow their personal progress through the course.

Benchmarks and instruments

To assess whether or not we have achieved our objectives, we have established the following benchmarks and instruments as parameters of success or failure.

In quantitative terms

- the final marks for the course module should represent an improvement on those of the previous academic year; at the very least, they should be the same or similar
- student use of the criterion-referenced rating scale should be in line with tutor moderation of scores and this should be reflected by a correlation between students' and tutor's scores within or greater than the 0.35-0.65 range

In qualitative terms

- the data gathered online from whole group discussions, team discussions, e-mail messages exchanged with the tutor, and "Time to reflect" activities should present positive opinions and attitudes on the part of the students towards the e-learning experience in general and to the process of self- and peer-evaluation and tutor moderation in particular.

Method

Study population

The population consisted of 78 students – 63 (80.8%) women and 15 (19.2%) men – registered for the course in Scientific and technical translation from Spanish into English. All of them had completed at least three years of the four-year first degree program in Translation and Interpreting. Some were in their fifth year, either because they had spent a year studying at an institution outside of Spain or because they had decided to take more than the minimum number of final year option credits, and had therefore extended their final year studies. Sample erosion was limited, and only five participants failed to complete the course: one was working full-time; another complained of an excessive workload as she was studying several options in the same semester; two were exchange students studying full-time outside of Spain; the fifth gave no explanation.

Teamwork organization

Much of the assessment involved students working in teams. These usually consisted of a maximum of five, and members were allocated to each team using a random numbers table to ensure the same people coincided as little as possible. The objectives of randomizing team selection in this way were to ensure that team scores represented individual capabilities as accurately as possible and to give students as wide an experience as possible of the positive and negative aspects of teamwork (Olvera Lobo et al., In press).

Course design

The language of instruction of the module was English, and all materials and instructions were published in English. Participants were deliberately given no specific guidelines as to the language they should use in their contributions in order to allow them to choose the language they were more comfortable with. It is not within the remit of this study to detail actual language use, but we can say that students overwhelmingly preferred Spanish for all types of interaction².

The basic course design involved three stages: a three-unit introductory stage during which students established the common ground of approaching Scientific and Technical translations; a five-unit-long intermediate stage which guided them through translation tasks focussing on the major components of the academic research article (RA); and a one-unit final stage in which they worked on a longer task completing the translation of a full text RA in the closest possible simulation of a professional environment.

The introductory stage of the course involved three so-called ‘learning units’: 1. “On common ground”, 2. “A level playing field” and 3. “Five’s company”. Unit 2 was dedicated entirely to presenting, explaining and instructing students in the use of the rating scale of criterion-referenced descriptors (Table 1). This included defining and distinguishing between the key concepts of error and mistake and discussing examples related to the application of the different descriptors and justifications for their use (Robinson 1998; Tercedor Sánchez, López Rodríguez and Robinson 2005). Students were then able to discuss the application of the descriptors via the open group discussion, team discussions and in individual written submissions when they were asked to reflect on prior learning.

The focus of the five-unit intermediate stage was the academic research article (RA) and the translation tasks centered on ± 350 -word fragments of RAs. The units began with the Abstract and then followed the traditional RA structure from Introduction through to Discussion and Conclusions. Thus, the texts presented a range of translation issues and were increasingly more demanding. In four units the task consisted of a team translation, and in the fifth it was an individual translation. Translation quality assessment (TQA) of the team translations was coordinated by one team member, but all participated; TQA of the individual translation was also an individual task, but students were encouraged to debate their decisions in the online team discussion forum. Teams of five carried out the final translation project, which consisted of a ± 3000 -word full text RA; marking was by peer-group.

Learning took place over the 15 weeks of the first semester of academic year 2005-6 (Table 2, below). The course was based on a blended e-learning model involving 40% contact classroom hours in a computer room with full internet access, and 60% e-learning. Students who did not have internet connections at home were able to use the Faculty self-access computer facility (open from 9.00 to 21.45, Monday to Friday) and had 4 hours/week reserved access during the timetabled class hours. Full group sessions were scheduled from the start and were timed to coincide with completion of the learning units, translation tasks and major translation project.

Table 2
Work schedule based on student learning hours per week

Week	Group sessions	E-learning time
1	2 x 2 hours	4 hours
2	1 x 2 hours	6 hours
3	2 x 2 hours	4 hours
4	1 x 2 hours	6 hours
5		8 hours
6	1 x 2 hours	6 hours
7		8 hours
8	1 x 2 hours	6 hours
9		8 hours
10		8 hours
11	2 x 2 hours	4 hours
12		8 hours
13	2 x 2 hours	4 hours
14		8 hours
15		8 hours

The course in its traditional format is worth 6 credits which, in Spanish universities, is equivalent to 60 contact classroom hours: 2 two-hour sessions per week over 15 weeks. Although the introduction of ECTS (European credit transfer system) credits has yet to take place, for some time now the Faculty of Translation

and Interpreting of the University of Granada has been involved in preparatory studies, and for the module described here, students were informed that they should consider the course required 8 hours per week *learning* time, whether or not contact classroom hours had been scheduled. During the presentation of the course and explicitly in the Study Guide, students were recommended to record and observe the amount of time they dedicated to the module.

Intermediate stage collaborative translation tasks were scheduled over 10-day periods, with 7 days for the documentation, terminology, translation and revision phases and then a further 3 days for the TQA phase. The final stage project was timed to last 3 weeks in line with normal, professional translation practice.

Assessment

The course included five assessment procedures (Table 3) in order to offer a range of opportunities to students in the belief that the more alternatives available, the better the chance that all students are able to play to their personal strengths.

Table 3
Assessment procedures

Participation (attendance at group sessions; active, constructive participation in forums and chats; some individual pre-translation tasks; "Time to reflect" tasks)	20%
Collaborative teamwork translation tasks	30%
Individual tasks	10%
Project work	30%
Final exam (1 text of \pm 350 words to translate in 2 hours)	10%
Total	100%

The inclusion of an assessment component for participation worth 20% of the final grade was motivated by a desire to focus student awareness on the value of active learning in a blended e-learning course and encourage them to be proactive. While our experience of teaching in this Faculty leads us to believe that students are highly motivated and proactive, we know that a minority tend to be more reactive in their approach to learning, which means that they respond to deadlines and tasks at the last minute and demonstrate a need for close tutor supervision. We predicted that this might lead to periods of inactivity interspersed with bursts of 'catching up' and that some students might find it more difficult to accept the increased responsibility that self- and peer-assessment entail.

The collaborative teamwork tasks were based on four of the five units centered on RA structure translation tasks. In each case, students worked in a team and all team members were awarded the same score. These scores represented 30% of the final grade. One of the collaborative teamwork tasks called for team preparation and individual completion of the translation task and TQA. This individual task represented 10% of the final grade.

A translation project at the third and final stage of the course called upon

students to complete a translation task centered on a full text RA of some 3000 words. Each team was required to complete a translation task and submit this to a different team for assessment. This peer-assessment represented 30% of the final score.

A final examination, representing 10% of the grade for the course, was included in the assessment model to fulfil three objectives: to provide those students who genuinely perform well under exam conditions an opportunity to do so; to provide one objective, individual assessment score; and to fulfil University requirements. The examination took place in the Faculty computer facility with full access to the Internet and other electronic resources such as dictionaries on CD-rom.

Assessment and moderation: application of the criterion-referenced descriptors

The most important innovation in this e-learning model is the use of student self- and peer-assessment and tutor moderation as pillars of the assessment procedure. The routine followed is an adaptation of the standard procedure for examination marking employed by exam boards such as the International Baccalaureate Organization (www.ibo.org) across the world (International Baccalaureate Organization 2004: 44)³.

The procedure involves five stages:

- Students complete a translation task and turn in their target text (TT)
- They are given access to a published translation of the text
- They apply the rating scale and score their TT
- They turn in the TT score with a written justification
- The tutor moderates the score.

Depending on the unit, students 1) made individual or team translations; 2) scored and justified their TTs individually or as a team; or 3) scored and justified the scores for their peers TTs. The staging of these variations was designed to facilitate student progress from naïve TQA to a more experienced, sophisticated application of the criteria presented in the rating scale.

Data collection

Data collection was conducted through the WebCT site and through the University of Granada webmail facility.

QUANTITATIVE DATA consisted of the scores awarded by individual students or teams of students to the completed translation tasks they turned in and the moderated scores for these pieces of work awarded by the course tutor. Pearson's *r* correlation coefficient was used to analyze the correspondence between these scores. The coefficient – also known as Pearson's product-moment correlation – calculates the relationship between two sets of data, in our case two scores awarded to each of the translations submitted by students: the students' own scores and those of the tutor. The coefficient indicates the strength of the relationship between the two sets of data (Henning 1987: 59; ALTE 1998: 156), and in the present study it can be interpreted in terms of the degree of agreement between students and tutor over the application of the rating scale.

In our courses we trained students in the use of the holistic grading system developed by Robinson (1998). The origins of the present rating scale and background to the design and development of this system have been described elsewhere (Robinson 1996 and 1998, Tercedor Sánchez, López Rodríguez and Robinson 2005). These criterion descriptors (cf. Table 1, above) allow the identification of the main areas where translation errors occur, relating them to the objectives of the translation process: decoding the source text and encoding the target text. In this way, there is a distinction between meaning errors and language errors (Gouadec 1989), covering the following areas from left to right: 1) Content, 2) Register, vocabulary, terminology, 3) Translation task and orientation to target text type, and 4) Written expression. Moreover, there is a vertical progression with regard to quality.

This assessment model is holistic, transparent, easy to understand and conducive to self-assessment. Students become aware of the positive and negative aspects of a translation and are able to classify errors and mistakes (Corder 1973, 1981) in terms of type (using the left-to-right criterion labels) and in terms of seriousness (using the vertical 0-10 scale). As a result, learners acquire 'editor-like' training.

In Unit 2, we present the criterion-referenced descriptors to students using the terminology defined by Corder in his ground-breaking work into error analysis (Corder 1973, 1981) rather than the more recent parallel terms defined by Martínez Melis and Hurtado Albir (2001: 281).

Following Corder, we consider *error* ("systematic error" in Martínez Melis and Hurtado Albir) to be a consequence of incomplete learning or a lack of knowledge: errors are usually systematic, and if you make an error, you make the same error more than once because the error represents something yet to be fully learned. We contrast this with the concept of *mistake* ("random error" in Martínez Melis and Hurtado Albir). A mistake is not systematic: it is usually the product of carelessness, tiredness or haste. You may make a mistake in one sentence and produce the correct version in the following sentence. Clearly, mistakes are bad and have to be avoided, but they are of less significance than errors, and this is reflected in the manner in which the two terms appear at different levels in the ordered descriptors (cf. Table 1).

Within the Department of Translation and Interpreting of the University of Granada, a number of teachers use criterion-referenced descriptors such as these in second, third and fourth year courses. However, as there will always be students who have not previously encountered this method of assessment, Unit2 of the Scientific and technical translation course is designed to introduce the concept and then teach something about how to apply it. To achieve this, we present fragments of source texts (ST) and target texts (TT) representing the different bands on the rating scale together with commentaries to explain why these examples have been scored in this way. All of the examples are authentic or semi-authentic, in that they have been drawn directly or adapted from student translations contributed in previous editions of this course (Tables 4-7).

Table 4
Extract from Unit 2: Teaching how to use the criterion-referenced rating scale with reference to Criterion A: Content

1-2	Major content errors. Major omissions of ST content.
3-4	Minor content errors. Some omissions of ST content.
5-6	Minor omissions of less relevant ST content. Over- or under-translation distorts ST content or results in ambiguity.
7-8	Over- or under-translation does not distort ST content or result in ambiguity.
9-10	ST content, including subtle detail, fully understood.

Each of the four criteria is treated separately, and to highlight the gradation from lower to higher scores we present the descriptors as above. On the website, the scores are color-coded with the failing bands 1-2 and 3-4, in yellow and the passing bands, 5-6, 7-8 and 9-10, in green.

Table 5
Extract from Unit 2. Criterion A: Content

ST	...34% no consumen fruta en todo el día. El 34% no consumen legumbres en todo el día.
TT1	... 34% did not eat either fruit or pulses in the whole day.
TT2	... 34% did not eat fruit and 34% did not eat pulses.

In Table 5 (TT1), the translator has probably committed this error because of the coincidence of the figures in the ST. The fact that both results refer to 34% of the sample does not mean that the children in each 34% are the same ones; the first translator's text states that they are. The translator probably believed that avoiding the repetition would improve the TT whereas in fact, the redundancy introduced by the second translator (TT2) was needed to state the information accurately. In assessing translation quality we consider this a major content error and would award a score in the 1-2 range.

Table 6
Extract from Unit 2. Criterion A: Content

ST	<i>Se realizó el estudio en un consultorio médico de familia perteneciente a un área de salud del municipio Centro Habana.</i>
TT1	The study was conducted in a family physician's office of Centro Habana municipality ...
TT2	The study was conducted in a family physician's office in Centro Habana, Cuba.

In Table 6 (TT1), the translator follows the ST and simply reproduces the town's name; in TT2, the translator identifies the country. The adaptation of geographical references to facilitate reading is as important in specialized texts as it is in general translation, but its absence neither communicates incorrect information nor – in this case – is it likely to confuse the reader. We consider this a minor omission of less relevant ST content and would score it in the 5-6 range.

Table 7
Extract from Unit 2: Criterion A: Content

ST	...(<i>grupo 8: bollería, y grupo 9: golosinas</i>).
TT1	... (group 8, pastries and group 9, sweets).
TT2	... (group 8, cakes and pastries and group 9, candies).

Here, the translators have been faced with an SL term that does not have a direct one-to-one TL equivalent and produced solutions that are partially correct. TT2 would earn a score in the 7-8 band, as it clearly shows that the translator has documented their work and taken a considered decision.

QUALITATIVE DATA was gathered from five sources: the whole group discussions, team discussions; e-mail messages exchanged with the tutor; "Time to reflect" activities; and the overall course evaluation questionnaire prepared and administered by the *Cevug*. The questionnaire cast light on wider issues and was more independent than our own instruments, as students were aware that the survey went beyond the confines of the specific subject.

Results

Quantitative data

The translation tasks completed individually or in teams and then moderated by the tutor constitute the quantitative data in this study. We compared scores by calculating Pearson's *r* correlation coefficient to identify the strength of agreement between students and tutor on the scores awarded.

Coefficients were calculated for each task and interpreted according to guidelines proposed by Cohen and Mannion (1989: 168-69), cf. Table 8. Our target for this initial study was set at 0.35-0.65 or better.

Table 8
Strength of correlation coefficients (Cohen and Mannion 1989: 168-9)

0.20-0.35	Limited use
0.35-0.65	Of little use unless supported by more than one study
0.65-0.85	"accurate enough for most purposes"
> 0.85	Strong correlation

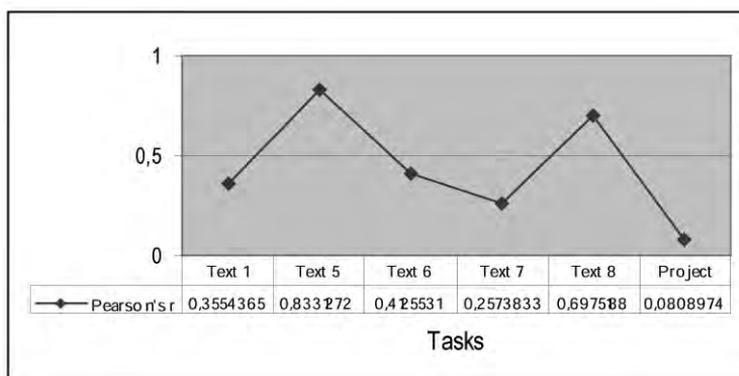
Students completed and self-assessed one individual translation, four team translations and one team translation project. The individual and team translations were of 350-word fragments taken from full text RAs; the project translation was of a ± 3000 -word full text RA. Table 9 summarizes the text types translated in each of the tasks, distinguishes between individual and team translations, and lists the corresponding coefficients calculated by comparing the scores awarded by the students with those of the tutor. Four of the six self- and peer-assessment tasks were successful: two of the coefficients fell within our target range and two were better than we had hoped for (the 'successful' figures are in bold in Table 9). None of the coefficients was negative.

Table 9
Student translation quality assessment and tutor moderation

Text type	Individual or team	Pearson's r coefficient	Target range 0.35-0.65
1 Abstract	Individual	0.355	On target
5 Introduction	Team	0.833	Above target range
6 Method	Team	0.412	On target
7 Results	Team	0.257	Below target
8 Discussion and conclusions	Team	0.698	Above target
Project	Team	0.081	Below target

In the present study, despite the overall success evident in the fact that four out of six correlations were within or above our target range, the fluctuations in the strength of the correlation (Figure 1) suggest that the principle we sought to establish, namely that self-assessment can be taught and used efficiently as a part of the overall assessment process, requires further consolidation.

Figure 1.
Task-on-task changes in Pearson's r



The correlation of $r=0.355$ for the first task (Text 1 Abstract) lies on the lower boundary of our target range. Given that it refers to the first task of its kind undertaken by students, this is encouraging. In contrast, the correlation recorded for the second task (Text 5 Introduction, $r=0.833$) is well above our target maximum. This result seems extremely high and could be explained by the fact that students were working in teams and that this first opportunity to analyze and apply the criterion descriptors together led to their fine-tuning of their perceptions to achieve such a successful result: they were on a learning curve and made every effort to 'get things right'. However, the erratic though still largely successful progress in the following three texts (Text 6 Method, $r=0.413$; Text 7 Results, $r=0.257$; Text 8 Discussion and conclusions, $r=0.698$) leaves the consistent consolidation of learning uncertain. In each case, the teams involved were different, which may mean that students varied their interpretations as a func-

tion of their interaction with different peer groups. Alternatively, the relative 'fall' on Text 7 – a coefficient below our target range – may be a function of the difficulty of the texts: each text was taken from a different part of the research articles: Text 6 belonged to the Method, Text 7 to the Results, which included tables. The least satisfactory of the correlations was that of the project ($r=0.08$), and two particularly important factors may have intervened here: this was the first time they had undertaken a project of these dimensions, and the project was scheduled for the end of the semester when many students were concerned to complete work for other courses and to prepare for the mid-year exam session. Perhaps the ± 3000 word project itself was, indeed, the bridge too far; in future editions of the module it may be better to limit translation tasks to the ± 350 word limit.

The qualitative data studied has been drawn from whole group discussions of contents and from submissions to the "Time to reflect" task that completes learning unit 2, "On common ground", in which the rating scale of criterion-referenced descriptors is presented.

Whole group discussion

Data relating to student contributions to the whole group discussion are drawn from WebCT messages recorded by students between Tuesday 18 and Monday 24 October 2005 (Table 10).

Table 10
Interventions in relevant online discussions

Nº interventions (Total=37)	Nº students (Total=17; 24%)
One	8
Two	5
Three	0
Four	1
Five	1
Six	2

All but one of the messages were in Spanish, and message length ranged from 19 lines down to a single word ("Right!"). Average message length was 6 lines (84 words).

The messages posted were threaded together in five themes, the titles of which were introduced by the first contributor in each case (Table 11).

The first thread discussed one of the key concepts introduced in the unit, namely that of the distinction between error, mistake (Corder 1973, 1981) and omission. The debate among students ranged over the reality of this distinction and over their personal interpretations. Much of the interaction centered on the question of what is deliberate and what is accidental and how omissions can be both deliberate and justifiable. The debate ended when one participant, somewhat cynically perhaps, described their experience of the professional translation marketplace and of the need to produce a financially acceptable product at the expense of a more professionally satisfying process.

Table 11
Contributions to online discussion threads

Thread topic	Nº of contributions
Opinion of Unit 2	15
Ambiguity	10
Ambiguities in English	7
Assessing translation quality	2
Criterion D: Written expression	3
Total	37

Ambiguity centered the debate in the second and third threads, and participants presented arguments ranging from the defence of an author-centered approach, in which ambiguity should be reproduced in the TT, to a reader-centered approach, in which the translator should strive to remove ambiguity. To this end, participants highlighted the importance of documentation, the distinction between scientific and technical translation and literary translation, the ambiguities inherent in languages, and the need to consult the author/client.

The thread headed “Assessing translation quality” contained only two messages from the same participant. The first message pointed to difficulties this participant had in following the contents of the unit and requested help. Curiously, one of the stated difficulties came from the use of metaphor. The content of the second messages clearly indicates that he received private responses – an option available in the discussions – to which we clearly cannot have access. This leaves us unable to assess the content of the reply or replies received.

The final thread was a straightforward request for help interpreting a sentence in the unit (reproduced in full in the participant’s contribution), a reply offering a contextualization of the sentence quoted, and acknowledgement of the helpfulness of the reply.

None of the contributions make really positive or negative comments about the rating scale itself, but the general tone of the debates is energetic and thoughtful and suggests that participants find the content on which the debate is centered and the debate itself to be productive and efficient. However, we cannot ignore the fact that only 17 out of 71 students participated in a discussion on a crucial unit (cf. Table 10). From the other activities undertaken, we can assume that the remaining 54 students had no difficulties arising from their initial learning about the rating scale. These contributions were considered when we awarded scores towards the Participation component of the final grade.

Submissions to “Time to reflect” activity

Throughout the course, units and teamwork tasks end with activities in which students are asked to reflect on their recent learning experience. They are encouraged to keep a diary and to make a note of the time they dedicate to different tasks and are sometimes given guidelines on topics to include in their reflections. Due to the configuration of WebCT, these activities are labeled compulsory and included among the assessment activities that students have to upload onto the website. This obliges students to respond and as a mechanism has a clear advantage over the open-ended invitation to (not) participate in the

discussion which we have mentioned above. It was made clear that submissions would not be 'marked' as such, but that they would be taken into account when scoring the Participation component.

The activity we have analyzed here took place after students had carried out two TQA tasks: the first was essentially naïve self-assessment without guidelines, designed to oblige them to focus on the issue without any input; the second, following the presentation in Unit 2, called upon them to apply the rating scale. It was hoped, therefore, that in their reflections they would compare and contrast the two tasks.

When asked to consider their first use of the rating scale, the prompts given were as follow:

- How are you going to assess your performance on this unit?
- Go back over the current unit and look carefully at all of the products of your work.
- Look back at the notes you made when you completed the unit and update your thoughts.
- Can you see differences in the process you have been through?
- Can you see differences in the products?

The references to the unit and the questions regarding differences were intended to elicit comparisons.

Although contributions to this activity were obligatory and students were given guideline prompts, this did not necessarily produce homogeneous responses from all participants. Table 12 shows the breakdown of responses in terms of those who did and those who did not make any reference to Unit 2 itself, the rating scale descriptors or the TQA task using these for the first time.

Table 12
Responses to "Time to reflect" activity

Reference to Unit 2, rating scales and TQA	
With	34 (46.6%)
Without	37 (50.7%)
No contribution	2 (2.7%)
Total	73 (100%)

Our analysis of these contributions involved identifying topics that centered on the self- and peer-assessment and the tutor moderation process: the unit, the rating scale, the TQA task, the whole group discussion and other; distinguishing between those that were positive, negative or neutral; and grouping contributions according to the keywords underlying their content: useful/future relevance, interesting, comprehensive/efficient, objective, learning/improvement, decision-making/professionalization and tolerance.

UNIT 2

- Only one participant specifically mentioned the manner in which Unit 2 was presented, identifying their own *learning* experience when reflecting

on the examples explained in the unit and candidly stating that they would have made the very same mistakes discussed.

THE RATING SCALE

- The bulk of the comments centered on the scale of criterion-referenced descriptors and made the hoped-for comparison between the naïve TQA task and the scale presented in the unit.
- Two students specifically referred back and outlined the criteria they had consciously or unconsciously adopted, finding similarities in the rating scale.
- Apart from expressing an interest in learning how to evaluate translations objectively (1 comment) in one way or another, 6 students pointed to the usefulness of the scale. Specifically, they described it as comprehensive in dealing with all aspects of a translation (6 comments); objective in providing criteria that were an improvement on those they had typically used previously (*That doesn't sound right. This sounds better) (5 comments); furthermore, they recognized the value of the scale as a tool they would be able to use in the future to evaluate and self-evaluate translations (3 comments).
- One student had found difficulty in understanding the descriptors and stated that they lacked precision. A further three alluded to this when reflecting on the whole group discussion, which they considered had been essential to clarify difficulties of interpretation of the scale itself and of aspects of the explanatory sections of the unit. All three were clearly appreciative of the help offered by their peers.
- The value of the descriptors as a professional instrument was clear to 3 students, as was its use to help with decision-making (3 comments). Seven students described the rating scale as an input to their learning. They stated that it would ensure they were more systematic and objective in their revision of translations and indicated that it had opened their eyes to the importance of revision, adequate documentation on content and the value of parallel texts to ensure an adequate approximation to the TT style.
- One student considered the criteria were "very strict".
- Three comments were written prior to completing the TQA task, and these students commented on the perceived benefits of the scale without having had the opportunity to use it. These were students who registered for the course after the start of the semester and who had some 'catching up' to do.

THE TRANSLATION QUALITY ASSESSMENT TASK

- The TQA task was new to students (1 comment), and this was the first time they had undertaken the activity. This in itself led to doubts about their preparedness for such an activity.
- Two students stated the time they had taken over the task: 20 minutes for one, 2 hours for the other.
- One affirmed that the descriptors had made their work more efficient.
- Twelve students wrote favorably about the task and indicated they had learned to be more tolerant of variety in the different translations of

the same text, although many of these (9 comments) indicated that the peer-assessment task was difficult because they were judging their companions.

- Four students identified the learning process that applying the rating scale involved as enriching and clearly felt the task was an appropriate challenge, however difficult.

OTHER

- Two comments reflected a reactive, teacher-centered orientation towards learning that is more typical of the traditional Spanish university classroom. These students called for individualized feedback and, in its absence, expressed a feeling of ignorance as to their learning progress. Their appreciation of self-assessment is that it is synonymous with no assessment.
- One student commented on the degree of similarity between peer-translated texts.

The majority of the comments included in the “Time to reflect” activity were positive and highlighted the benefits of the criterion-referenced rating scale as an objective instrument that would enable them to learn, to be more objective and more tolerant in their assessment of translation quality.

Cevug global questionnaire

The online questionnaire was written and processed by the *Cevug* and was aimed at participants in any of the e-learning modules. Statistics were calculated for the Scientific and technical translation module and comparisons made with 10 other modules taught online during the same semester of 2005-06, none of which was on translation, and none of which used a self- and peer-assessment and tutor moderation process similar to ours. Questions and answers were all in Spanish, and the items were either open-ended – with students having a space in which to write whatever they considered appropriate – or required responses on a Likert scale ranging from 1 to 6, where 1 was defined as “total disagreement” and 6 as “total agreement”. The latter item type gave no opportunity for adding comments.

The questionnaire was completed by 62 (85%) of the 73 students who finished this course, and 50 (81%) of the respondents were women, 12 (19%) men. Out of 31 questions, three focussed on topics related to the present study:

- Question 17. *The techniques and procedures used in assessment were consistent with the course objectives.
- Question 30. *Of all the resources available to you, which has most helped you to learn?
- Question 31. *What would you recommend to improve this course? (design, organization of content, methodology, etc.)

One item sought an overview of the course:

- Question 32. *Overall, what is your opinion of this course?

Details of responses are summarized in Table 13.

Table 13
Summary of student responses to the Cevug questionnaire

Students finishing the course	73 (100%)	
Students completing the questionnaire	62 (84.9%)	
Responses	Total	Relevant
Question 17	62 (84.9%)	
Question 30	49 (67.1%)	2 (2.7%)
Question 31	45 (61.6%)	10 (13.7%)
Question 32	62 (84.9%)	

QUESTION 17: ASSESSMENT TECHNIQUE AND PROCEDURES

This item elicited responses ranging from “total disagreement” (1) to “total agreement” (6), and no further comments were possible. The average score was 3.5, the same as the median for all 11 e-learning courses taught during the same semester, but below the 5.0, which was the mode for these modules. While a comparison of this nature is of limited value, it does suggest that some rethinking is required, although it gives no indication as to which aspects of the assessment technique and procedures failed to please students.

QUESTION 30: RESOURCES AVAILABLE

This item produced 49 open, written responses, of which 2 referred to the assessment of the course. One was essentially negative:

- *I suppose that the individual translations [helped me to learn most] but not a lot, given that we have not been corrected by the teacher but by our companions, and in my opinion that is not reliable (Response id 1045).

One was positive about ...

- *the way of working: the self-assessment and self-evaluation and the independence in our work (Response id 1197).

The data is clearly limited, Response id 1045 is the only one that is negative, so it is impossible to generalize, but clearly this aspect of the course may require rethinking if there are further indicators, via alternative modes of providing feedback, that it is unsatisfactory.

QUESTION 31: RECOMMENDATIONS

Question 31 elicited 45 open, written responses, of which 10 included references to aspects of the assessment procedure. These centered on two topics, neither of which was directly linked with the rating scale: 3 comments recommended scheduling the publication of scores awarded for translations, and 7 called for more teacher correction. Of these, 2 students specifically stated they needed more interaction with the teacher in correcting translations, whether in class or

online; another 2 pointed out their need to use corrections to judge their own improvement; 1 proposed corrected translations should be accompanied by a written commentary; and 1 complained that the teacher had placed all the responsibility for correction on the students.

QUESTION 32: OVERALL OPINION

This item elicited responses on a 1-to-6 scale, and the average score achieved by the module was 4.1.

Discussion

Year-on-year outcomes

To assess the efficacy of self- and peer-assessment, we compared the final grades awarded for the course taught online with those for the same module taught in a more traditional format during the previous academic year (Table 14). Our stated objective was to ensure that results were, at the very least, the same as previously, if not better.

Table 14
Comparison of year-on-year outcomes,
traditional vs. blended e-learning formats

Grade	2004-05 Traditional format	2005-06 E-learning format
Fail	0 (0%)	0 (0%)
Pass	11 (13.3%)	0 (0%)
Merit	64 (77.1%)	62 (84.9%)
Credit	7 (8.4%)	11 (15.1%)
Honors	1 (1.2%)	0 (0%)
Total	83 (100%)	73 (100%)

Clearly the most significant difference year on year is the 'bunching' of grades that has occurred in 2005-06, with no candidate attaining a Pass or an Honors grade. This is most likely to be due to two factors: one, an inherent part of the process of change, namely the conscious or unconscious wish on the part of the tutor not to prejudice students who are to some extent acting as 'guinea-pigs' in this first edition of the e-learning format of the course; the second is the inclusion of the Participation component, worth 20% of the overall grade. Inevitably, this must benefit students and is most likely to benefit weaker students more. The relationship between a change in the teaching mode and a desired improvement in the final grades is questionable. We would even suggest that it may be wrong to expect change to equate with improvement in results. After all, the students belong to the same population – i.e. the population of University of Granada undergraduate students of Translation and Interpreting with English as their first foreign language – and from our knowledge of the Faculty we are aware that this population has changed little over the last 12 years. Perhaps it would have been more appropriate to consider the consequences of change solely in qualitative terms.

Self- / peer-assessment vs. tutor moderation (using Pearson's r correlation coefficient)
 Correlation coefficients indicate the relative strength of the relationships between variables. They do not shed any light on the nature of the relationship, nor do they reveal its cause. In the present study we calculated Pearson's r to compare the variables 'Student score' and 'Tutor score' for each of six translation quality assessment (TQA) tasks. It was our belief that, if the criterion-referenced rating scale (Table 3) had been applied appropriately and consistently by students, there would be minimally useful correlations between their scores and those awarded by the tutor. In line with Cohen and Mannion (1989) we set our target on the 0.35-0.65 range.

The interpretation of Pearson's r is a topic of debate, and much depends on the relative homogeneity of the population sample. Cohen and Mannion consider 0.4 as the absolute minimum "crude" affirmation of the existence of a relationship, whereas Ehrman (1995) accepts 0.2 as the minimally significant correlation when dealing with a highly homogeneous sample. At the other end of the scale, the IBO uses $r=0.9$ as the minimally accepted correlation coefficient when moderating marking by examiners to ensure inter-examiner reliability. A coefficient lower than this means all of the examiner's work is remarked and would normally lead to termination of the examiner's contract (International Baccalaureate Organization 2004: 45).

While slightly erratic, student use of the rating scale has been more than satisfactory, and we can only hope that in future editions of the course we can make improvements in the areas indicated via student feedback that will enhance this.

Criterion-referenced rating scale

Our experience of the rating scale in previous years and that of other colleagues has demonstrated that the basic design of the instrument makes it an effective pedagogic tool. The quantitative and qualitative data on the current version is highly satisfactory, and we believe little need be modified prior to the next edition of the course. We believe that positive student perception of the scale is essential if it is to function adequately, and that we have achieved a high degree of acceptance in this course. We are aware that there are deficiencies in student learning about the rating scale, and we propose to remedy these by introducing activities and materials such as those described elsewhere (López Rodríguez, Robinson and Tercedor Sánchez, Forthcoming; Tercedor Sánchez, López Rodríguez and Robinson 2005) along with more typical e-learning quizzes using HotPotatoes software, in order to improve the way in which students use it.

Translation quality assessment (TQA)

The TQA task and the translation routine of which it forms a part is the systematization of our pedagogy and represents the fruits of the years of instruction and learning that students have experienced prior to this course, more than it does the course itself. Student perception of the task is overwhelmingly positive, and this encourages us to believe that our objectives of piloting and assessing the efficacy of the combination of self- and peer-assessment and tutor moderation have been achieved.

Student perceptions

The data on student perceptions has been gathered from the whole group discussions and “Time to reflect” activities and from the *Cevug* questionnaire. We believe that these sources of data have provided us with generally positive opinions and attitudes on the part of the students towards the e-learning experience in general, and to the process of self- and peer-evaluation and tutor moderation in particular. However, one of the issues that we need to address for future editions of the course is the question of participation in these modes of communication. Despite the fact that our assessment procedures included a participation component worth 20% of the final mark, contributions to the discussion were relatively few. This contrasted with submissions to the “Time to reflect” activities which, due to the configuration of the online platform WebCT, were obligatory. In both the discussions and the reflections, the quality of contributions was high, so we need to resolve the issue of whether we should make changes in the hope of bringing about an increased volume of participation in the discussions and, if we do decide to do so, how to achieve this. In the “Time to reflect” activities we have seen the value of using prompts to guide students’ submissions, and we could perhaps improve on these by including the concepts underlying the keywords we have identified in the contributions of our current batch of students, namely useful/future relevance, interesting, comprehensive/efficient, objective, learning/improvement, decision-making/professionalization, and tolerance.

The *Cevug* questionnaire was also a voluntary contribution, and the statistical data deriving from the comparisons with other courses taught in the same semester are of questionable value. However, specific comments and recommendations elicited by the open questions are of greater use: issues such as the publication of scores, teacher correction and corrections in general can be summarised as a demand for improved feedback. While the numbers of comments along these lines are limited, it is evident that there is a tension created by the self-assessment mode which imposes greater responsibility on students. This, in turn, leads some students to demand more attention from the tutor – attention that they do not feel they have received in the modes of interaction available via WebCT. This issue clearly takes us beyond the confines of the present study and deserves future empirical research.

Conclusions

Self-assessment in translator training is a logical component of any course designed to prepare translators for the professional marketplace. In the e-learning course we have described here, we have demonstrated that a self- and peer-assessment routine validated by tutor moderation can achieve satisfactory results both in quantitative and qualitative terms. Furthermore, we believe that the model we have developed is generally perceived by participants to be a pedagogically and professionally acceptable form of instruction. While there is clearly room for improvement in aspects of the course design and application – not the least of which are connected with the tension between student dependence and student autonomy, and the fostering of a sense of individual responsibility – we believe we have crossed a bridge of significant dimensions in our desire to advance in the training of translators.

Notes

1. This research is part of the project PUERTOTERM: *knowledge representation and the generation of terminological resources* within the domain of Coastal Engineering, BFF2003-04720, funded by the Spanish Ministry of Education.
2. All student texts originally in Spanish appearing in this article have been translated by the authors and are preceded by an *.
3. The author was the IBO Examiner responsible for Language B English from 1993 to 2001 and has been an assistant examiner since 2001.

Works cited

- Alcina Caudet, Amparo. 2002. Tutorizar trabajos terminológicos en un entorno virtual de trabajo colaborativo. *Séndebar*, no. 13: 169-81.
- ALTE 1998. *Multilingual glossary of language testing terms*. Studies in Language Testing, ed. Michael Milanovic, 6. Cambridge: Cambridge University Press.
- Castells, Manuel. 1996. *The Rise of the Network Society*. Vol. I. Malden, Maryland, USA: Blackwell.
- Cohen, L., and L. Mannion. 1989. *Research methods in education*. London: Routledge.
- Corder, S. Pit.. 1973. *Introducing Applied Linguistics*. Harmondsworth, England: Penguin Books Ltd.
- Corder, S. Pit. 1981. *Error analysis and interlanguage*. Oxford: Oxford University Press.
- Cordón, Oscar. 2004. Cevug 2004. Presentación del Plan de Virtualización de Asignaturas regladas de la Universidad de Granada para el curso 2004-05 (PVA04-05). *Jornadas de Presentación del Plan de Virtualización de Asignaturas 2004-05*. Available at www.ugr.es/cevug.
- Cózar Sievert, Ramón de. 2003. *Tuning educational structures in Europe*.
- Ehrman, Madeline E. 1995. *A Study of The Modern Language Aptitude Test for Predicting Learning Success and Advising Students*. Available from the author at ehrmann@aol.com.
- González Davies, María. 2004. *Multiple voices in the Translation Classroom*. Amsterdam: John Benjamins.
- Gouadec, Daniel. 1989. Comprendre, évaluer, prévenir. Pratique, enseignement et recherche face à l'erreur et à la faute en traduction. *TTR* 2, no. 2: 35-54.
- Henning, Grant. 1987. *A Guide to Language Testing*. Development Evaluation Research. Boston: Heinle & Heinle Publishers.
- International Baccalaureate Organization. 2004. *Diploma Programme assessment. Principles and practice*. Geneva: International Baccalaureate Organization. Available at www.ibo.org/diploma/assessment/documents/d_x_dpyyy_ass_0409_1_e.pdf.
- Kiraly, Donald C. 2000. *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*. Manchester: St. Jerome.
- López Rodríguez, Clara Inés; Bryan J Robinson, and María Isabel Tercedor Sánchez. Forthcoming. A learner-generated corpus to direct learner-centered courses. In: Thelen, Marcel and Barbara Lewandowska-Tomaszczyk (eds.) *Translation and Meaning, Part 7*.
- Martínez Melis, Nicole and Amparo Hurtado Albir. 2001. Assessment in translation studies: research needs. *Meta* XLVI, no. 2: 272-87.
- Olvera Lobo, María Dolores; María Rosa Castro Prieto, Ricardo Muñoz Martín, Eva Muñoz Raya, Miguel Murillo, Enrique Quero, Bryan Robinson, José A. Senso, Benjamín Vargas, and Cristina Domínguez-López. 2005. Translator training and modern market demands. *Perspectives: Studies in Translatology* 13, no. 2: 132-42.
- Olvera Lobo, María Dolores; Bryan Robinson, María Rosa Castro Prieto, Ricardo Muñoz Martín, Eva Muñoz Raya, Enrique Quero, Miguel Murillo, José A. Senso, Benjamín Vargas, and José Luis Díez Lerma. In press. A professional approach to translator training (PATT). *Meta*.
- Robinson, Bryan J. 1996. La evaluación de la traducción. In: Lozano, Wenceslao-Carlos and José Luis Vázquez Marruecos (eds.) *Actas de las I Jornadas Nacionales sobre Diseño Curricular del Traductor e Intérprete*. Granada, Spain: University of Granada: 199-208.
- Robinson, Bryan J. 1998. Traducción transparente: métodos cuantitativos y cualitativos en la evaluación de la traducción. *Revista De Enseñanza Universitaria*. Número extraor-

dinario: 577-89.

Robinson, Bryan J. 2003. *Proyecto docente*. Available from the author.

Tercedor Sánchez, María Isabel; Clara Inés López Rodríguez, and Bryan J Robinson. 2005.

Textual and visual aids for e-learning translation courses. *Meta L*, no. 4: 12.

Waddington, Christopher. 2001. Different Methods of Evaluating Student Translations: The Question of Validity. *Meta XLVI*, no. 2: 311-25.