Document generated on 10/12/2024 4:23 a.m.

Meta
Journal des traducteurs
Translators' Journal

META

Different Methods of Evaluating Student Translations: The Question of Validity

Christopher Waddington

Volume 46, Number 2, juin 2001

Évaluation : paramètres, méthodes, aspects pédagogiques / Evaluation: Parameters, Methods, Pedagogical Aspects

URI: https://id.erudit.org/iderudit/004583ar
DOI: https://doi.org/10.7202/004583ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)
1492-1421 (digital)

Explore this journal

Cite this article

Waddington, C. (2001). Different Methods of Evaluating Student Translations: The Question of Validity. *Meta*, *46*(2), 311–325. https://doi.org/10.7202/004583ar

Article abstract

This article examines the criterion-related validity of the results obtained by the application of four different methods of assessment to the correction of a second-year exam of translation into the foreign language (Spanish-English) done by 64 university students. These four methods are based on types currently used by university teachers, and the validation study is based on 17 external criteria taken from six different sources. In spite of this variety, a factor analysis reveals the presence of one main factor which is clearly identifiable as *Translation Competence* . The hypotheses regarding differences between the validity of the methods are verified as null, since all the systems, whether based on error analysis or a holistic approach, prove to correlate significantly with this main factor.

Tous droits réservés © Les Presses de l'Université de Montréal, 2001

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

Different Methods of Evaluating Student Translations: The Question of Validity

CHRISTOPHER WADDINGTON
Universidad Pontificia Comillas, Madrid, Spain

RÉSUMÉ

Cet article porte sur la validité des résultats obtenus à partir de l'application de quatre méthodes d'évaluation différentes à la correction d'un examen de thème (espagnol-anglais) fait par 64 étudiants inscrits en deuxième année de traduction. Les quatre méthodes se fondent sur les systèmes d'évaluation utilisés aujourd'hui par des professeurs d'université, et l'étude de la validité des méthodes repose sur 17 critères externes puisés dans six sources. Malgré cette variété, une analyse de facteurs révèle la présence d'un facteur principal qui peut s'identifier, de toute évidence, comme la *compétence de traduction*. Les hypothèses concernant les différences entre la validité des méthodes sont nulles, parce que les méthodes, aussi bien celles qui s'appuient sur l'analyse des erreurs que celles qui s'appuient sur une approche holistique, démontrent une corrélation significative avec ce facteur principal.

ABSTRACT

This article examines the criterion-related validity of the results obtained by the application of four different methods of assessment to the correction of a second-year exam of translation into the foreign language (Spanish-English) done by 64 university students. These four methods are based on types currently used by university teachers, and the validation study is based on 17 external criteria taken from six different sources. In spite of this variety, a factor analysis reveals the presence of one main factor which is clearly identifiable as *Translation Competence*. The hypotheses regarding differences between the validity of the methods are verified as null, since all the systems, whether based on error analysis or a holistic approach, prove to correlate significantly with this main factor.

MOTS-CLÉS/KEYWORDS

evaluating student translations, methods, Translation Competence, translation quality, translation errors

1. Introduction

The review of the literature on Translation Quality Assessment (Waddington 1999: chapters 1-5) shows that almost all the contributions have been descriptive or theoretical and have centred mainly on the following themes:

(i) Establishing the criteria for a "good translation" (Darbelnet 1977, Newmark 1991);
(ii) The nature of translation errors:
- Defining the nature of translation errors as opposed to language errors (House 1981, Nord 1993, Kussmaul 1995, Gouadec 1989);
- Drawing up a catalogue of possible translation errors (Gouadec 1981);
- Establishing the relative, as opposed to absolute, nature of translation errors (Williams 89, Gouadec 89, Pym 92, Kussmaul 95);

- The need to assess quality not only at the linguistic but also the pragmatic level (Sager 1989, Williams 1989, Hewson 1995, Kussmaul 1995, Nord 1996, Hatim & Mason 1997);

(iii) Basing quality assessment on text linguistic analysis (House 1981, Larose 1989);
(iv) Establishing various textual levels on a hierarchical basis and linking the importance of mistakes to these levels (Dancette 1989, Larose 1989);
(v) Assessment based on the psycholinguistic theory of "scenes and frames" (Dancette 1989 and 1992, Bensoussan & Rosenhouse 1994, Snell-Hornby 1995).

I came across very few contributions which submitted aspects of Translation Quality Assessment to some kind of empirical study. The following are exceptions to this rule:

(i) Campbell (1991) examines translation tests to see to what extent they examine translation competence and reveal translation processes rather than comparisons between source and target texts. The test papers of 38 candidates (belonging to four different ability groups) for a public examination in English-Arabic translation were analysed using ten criteria, such as lexical variety ratio, average word length, words omitted, etc. On the basis of the correlation matrix for the 38 subjects and ten analyses Campbell proposes the existence of three separate factors: *lexical coding of meaning*, *global target language competence* and *lexical transfer competence*.

This study follows in the steps of Séguinot (1989, 1990) in that it examines the processes of translation as reflected in the quality of the translator's work, although Séguinot concentrates on students' mistakes, whereas Campbell examines other aspects as well. However, these other aspects are mainly linguistic and Campbell ignores higher textual levels and the students' ability to manage the pragmatic elements in a translation. It is also surprising that Campbell does not use a factor analysis to determine the nature of the components of translation competence; perhaps this was because his study is based on too few subjects (n=38).

(ii) Stansfield et al. (1992) also aim to "identify the variables that constitute translation ability" (Stansfield et al. 1992: 455) and their study is based on work carried out for the U.S. Federal Bureau of Investigation (FBI) to develop and validate job-related tests of translation ability. The initial tests conducted with FBI employees indicated that translation competence should be divided into two different skills: (a) *Accuracy*, which is the degree of accuracy with which the translator transfers the content from the source to the target text; and (b) *Expression*, which refers to the quality of the translator's expression of this content in the target language. The team drew up translation skill-level descriptions for each of these aspects of translation competence and applied them to the correction of the subsequent tests held with the FBI employees. A study of the criterion-related validity of these results was then conducted by analysing how they correlated with the following external criteria: self-ratings provided by the examinees of their ability to translate different types of text, plus four other available measures of the candidates' performance in a Spanish oral interview, an English oral interview, a Spanish reading comprehension test and a Spanish listening comprehension test. Stansfield et al. claim that this validation study supports their division of translation competence into two different constructs, although it also indicates that *Accuracy* appears to be "the more valid measure of translation ability" (Stansfield et al. 1992: 461).

This study confirms what teachers of translation have long suspected, i.e., that translation competence consists of two different aspects, the ability to understand and transfer the content of the source text and the ability to express this content adequately in the target language. However, the study is based on tests of translation where only one third is devoted to translating three isolated paragraphs; the remaining two-thirds consists of multiple choice tests and the translation of isolated words, expressions or sentences.

The present paper differs from both Campbell (1991) and Stansfield et al (1992) in that it concentrates on summative testing in the university context. In order to find out the kind of translation exam and the kinds of methods of correction currently in use in Faculties of Translation, I sent out a questionnaire to 48 European and Canadian universities. A total of 52 teachers replied from 20 of these universities and their answers reflected the following situation:

(i) All the teachers said that they require the students to translate a text, although over half also include other complementary tests.

(ii) As far as methods of evaluating student translations were concerned, 36.5% of the teachers use a method based on error analysis, 38.5% use a holistic method, and 23% combine error analysis with a holistic appreciation.

In accordance with these findings, this paper considers the validity of the results obtained through applying these different types of method to the correction of translations of part of an authentic text done by students under exam conditions.

2. Description of experiment

2.1. *Description of the four methods of assessment*

(1) Method A

Method A is taken from Hurtado (1995); it is based on error analysis and possible mistakes are grouped under the following headings:

(i) Inappropriate renderings which affect the understanding of the source text; these are divided into eight categories: *contresens*, *faux sens*, *nonsens*, addition, omission, unresolved extralinguistic references, loss of meaning, and inappropriate linguistic variation (register, style, dialect, etc.).
(ii) Inappropriate renderings which affect expression in the target language; these are divided into five categories: spelling, grammar, lexical items, text and style.
(iii) Inadequate renderings which affect the transmission of either the main function or secondary functions of the source text.

In each of the categories a distinction is made between serious errors (–2 points) and minor errors (–1 point). There is a fourth category which describes the plus points to be awarded for good (+1 point) or exceptionally good solutions (+2 points) to translation problems. In the case of the translation exam where this method was used, the sum of the negative points was subtracted from a total of 110 and then divided by 11 to reach a mark from 0 to 10 (which is the normal Spanish system). For example, if a student gets a total of –66 points, his result would be calculated as follows: 110-66=44/11=4 (which fails to pass; the lowest pass mark is 5).

(2) Method B

Method B is also based on error analysis and was designed to take into account the negative effect of errors on the overall quality of the translations (Cf. Kussmaul 1995:129, and Waddington 1999: chapter 7). The corrector first has to determine whether each mistake is a translation mistake or just a language mistake; this is done by deciding whether or not the mistake affects the transfer of meaning from the source to the target text: if it does not, it is a language error (and is penalised with –1 point); if it does, it is a translation error (and is penalised with –2 points). However, in the case of translation errors, the corrector has to judge the importance of the negative effect that each one of these errors has on the translation, taking into consideration the objective and the target reader specified in the instructions to the candidates in the exam paper. In order to judge this importance, the corrector is given the following table:

TABLE 1

Typology of errors in Method B

Negative effect on words in ST:	*Penalty for negative effect*
On: 1-5 words	2
6-20 words	3
21-40 words	4
41-60 words	5
61-80 words	6
81-100 words	7
100+ words	8
The whole text	12

The final mark for each translation is calculated in the same way as for Method A: that is to say, the examiner fixes a total number of positive points (in the case of method B, this was 85), then subtracts the total number of negative points from this figure, and finally divides the result by 8.5. For example, if a student is given 30 minus points, his total mark would be 6.5 (pass): 85-30 = 55/8.5 = 6.5.

(3) Method C

Method C is a holistic method of assessment. Although, in the survey mentioned above, the teachers who answered were requested to send a brief description of the method of assessment they applied, I only received three descriptions of holistic methods. In addition to this, all three methods based their scales on the requirements of professional translation and were consequently of little use for judging the quality of translation into the foreign language. As a result, I had to design the following holistic method myself. The scale is unitary and treats the translation competence as a whole, but requires the corrector to consider three different aspects of the student's performance, as shown in the table below. For each of the five levels there are two possible marks, so as to comply with the Spanish marking system of 0 – 10; this allows the corrector freedom to award the higher mark to the candidate who

fully meets the requirements of a particular level and the lower mark to the candidate who falls between two levels but is closer to the upper one.

TABLE 2

Scale for holistic Method C

Level	Accuracy of transfer of ST content	Quality of expression in TL	Degree of task completion	Mark
Level 5	Complete transfer of ST information; only minor revision needed to reach professional standard.	Almost all the translation reads like a piece originally written in English. There may be minor lexical, grammatical or spelling errors.	Successful	9, 10
Level 4	Almost complete transfer; there may be one or two insignificant inaccuracies; requires certain amount of revision to reach professional standard.	Large sections read like a piece originally written in English. There are a number of lexical, grammatical or spelling errors.	Almost completely successful	7, 8
Level 3	Transfer of the general idea(s) but with a number of lapses in accuracy; needs considerable revision to reach professional standard.	Certain parts read like a piece originally written in English, but others read like a translation. There are a considerable number of lexical, grammatical or spelling errors.	Adequate	5, 6
Level 2	Transfer undermined by serious inaccuracies; thorough revision required to reach professional standard.	Almost the entire text reads like a translation; there are continual lexical, grammatical or spelling errors.	Inadequate	3, 4
Level 1	Totally inadequate transfer of ST content; the translation is not worth revising.	The candidate reveals a total lack of ability to express himself adequately in English.	Totally inadequate	1, 2

(4) Method D

Method D consists of combining error analysis Method B and holistic Method C in a proportion of 70/30; that is to say, Method B accounts for 70% of the total result and Method C for the remaining 30%.

2.2. *The three hypotheses*

(1) The first hypothesis was that "It is possible to improve the quality of methods of correction based on error analysis, if this analysis takes into account the possible effect of the errors on the rest of the translation."
To verify this hypothesis, the results obtained by Method A were compared to those obtained by Method B.

(2) The second hypothesis was that "Methods of assessment based on error analysis are more reliable and valid than holistic methods."
To verify this hypothesis, the results obtained by Methods A and B were compared to those obtained by Method C.

(3) The third hypothesis was that "The quality of a translation can be assessed more accurately if the method of assessment combines error analysis with a holistic appreciation."
To verify this hypothesis, the results obtained by Method A were compared to those obtained by Method D.

All three hypotheses present questions as to both the reliability and the validity of these four methods of assessment, but this paper only reports the study of validity. For the study of reliability, the reader is directed to Waddington (1999: chapter 8), which is in Spanish, or to the following two papers in English: Waddington (2000a) and Waddington (2000b).

2.3. *How the methods were applied*

To verify the three hypotheses, the four methods were applied to the correction of a second-year translation exam done by 64 students on the undergraduate degree course of Translation and Interpreting at the *Universidad Pontificia Comillas de Madrid*. The text of the exam paper which the students had to translate was an editorial from a Spanish newspaper (the *ABC*), entitled "Diálogo de la lengua," which discussed the present status of the Spanish language. The text was 330 words long and the students had 3 hours to translate it.

The three assessment methods were applied by five correctors; all were teachers at Comillas, but only two were teachers of translation; the remaining three were teachers of English as a foreign language and had virtually no experience of translation teaching. They applied each method to the 64 translations in a different order and with at least a month's interval between each method, and, before applying the method, they were trained in its use.

The results of the application of the methods by the five correctors can be found in Waddington (1999: 324).

3. The validity study

3.1. Description of the external criteria

To assess the criterion-related validity of the four methods, the results obtained by their application to the 64 translations were compared to the results obtained by the students in seventeen external criteria. These criteria can be grouped under six headings:

(i) knowledge of languages (criteria 1 – 3); (ii) results in intelligence tests (criteria 4 – 5); (iii) students' self-assessment (criteria 6 – 7); (iv) teachers' assessment of the students (criterion 8); (v) students' average mark in their translation course (Spanish-English) (criterion 9); (vi) marks in other translation exams (criteria 10 – 17).

(1) *Spanish Language*: this criterion is based on the marks obtained by the students in the final exam of the second-year subject in Spanish, which is part of the T&I syllabus.
(2) *English Language*: this criterion is based on the marks obtained in the final exam of the second-year subject in English, which is also part of the syllabus. The exam papers for both English and Spanish can be found in Waddington (1999: appendix 4).
(3) *English Composition*: this criterion is based on subjective marks awarded by the students' teacher of English in their second year and based on the assessment of composi-

tions handed in by the students in the course of the year. The teacher who awarded these marks is the same as the one who provided the marks for criterion 2.

(4) *Verbal Test*: verbal intelligence. Both this criterion and the next one are based on the results obtained by the students in the intelligence tests that form part of the entrance exam which the students had to take when they applied to enter the university two years before the translation exam which is the basis of the present study.
Example: …es a AGUA como COMER es a…
A. viajar – conducir
B. pie – enemigo
C. beber – pan
D. muchacha – industria
E. beber – enemigo
(The correct answer is C: "*Drinking* is to WATER as EATING is to *bread*.")

(5) *Series test*: mathematical series.
Example: Question: 3 5 7 9 11
Possible answers: A. 20 B. 18 C. 16 D. 13 E. 15
(The correct answer is D.)

(6) *Self-Assessment 2*: self-assessment made by the students of their ability to translate from Spanish into English the type of general texts that formed the basis of the second-year course General Translation (Spanish-English).

(7) *Self-Assessment 3*: self-assessment made by the students of their ability to translate from Spanish into English the type of specialised texts that formed the basis of the third-year course in Specialised Translation (Spanish-English).
Whereas the translation exam which forms the basis of this study was held in April 1997, these two self-assessments were completed by the students just before the first midterm exam in Specialised Translation (English-Spanish) in January 1998.

(8) *Teachers' Assessment*: Five teachers of the Faculty who knew the 64 students well (they have given them classes for at least one academic year) were asked to choose the ten best students and the ten worst and to order both lists from the best to the worst. The teachers were asked to consider overall excellence, which could include: ability in the teacher's own subject, ability to follow the degree, linguistic ability (in both Spanish and the foreign languages), attitude towards to the degree, willingness to participate, and any other aspects which distinguish the good students from those who are less good. This information was then converted in the following way: (i) each teacher's evaluation was recorded like this: the ten best students were numbered from 1 – 10 (from the best to the least good), and the ten worst from 21 – 30 (from the least bad to the worst); (ii) the following scale was applied to these numbers:

TABLE 3

Scale to convert teacher's assessment

Teacher's assessment	New number
1 - 5	1
6 - 10	2
no mark	3
21 - 25	4
26 - 30	5

In this way, each of the 64 students was given five marks (one by each teacher), which served in turn to provide the final single average mark which figures in the table in Waddington (1999:329).

(9) *Average translation mark*: this is the average mark which each student obtained in the translation course they did with me in second year. This mark is the average for all the translations which each student did throughout the course, whether for classwork or for exams; the only mark that was omitted was that for the second midterm exam, as this was the translation used in the experiment.

The remaining eight criteria are based on marks which the students obtained in different final exams of translation both from and into Spanish, and not only in the combination English-Spanish, but also in French-Spanish and German-Spanish:

(10) *General Translation B dir 1*: General Translation B in first year (English-Spanish). Because of the lack of a common term in English for *traducción directa* (translation into the native language) and *traducción inversa* (translation into the foreign language), I have retained the Spanish: *dir* for *directa*, *inv* for *inversa*.
(11) *General Translation B dir 2*: General Translation B in second year (English-Spanish).
(12) *General Translation C dir 2*: General Translation C in second year (French-Spanish or German-Spanish).
(13) *General Translation C inv 2*: General Translation C in second year (Spanish-French or Spanish-German).
(14) *Specialised Translation B dir 3*: Specialized Translation B in third year (English-Spanish).
(15) *Specialised Translation B inv 3*: Specialized Translation B in third year (Spanish-English).
(16) *General Translation C dir 3*: General Translation C in third year (French-Spanish or German-Spanish).
(17) *General Translation C inv 3*: General Translation C in third year (Spanish-French or Spanish-German).

3.2. *Factor analysis of the external criteria*

(1) Matrix of correlations between the four methods and the external criteria:

TABLE 4

Matrix of correlations between the four methods and the external criteria

	Method A	*Method B*	*Method C*	*Method D*
Spanish language	.244	.272*	.279*	.284*
English language	.503**	.542**	.471**	.542**
English composition	.495**	.545**	.488**	.550**
Verbal test	.301*	.261*	.269*	.274*
Series test	.261*	.240	.164	.226
Self-assessment 2	.498**	.423**	.475**	.457**
Self-assessment 3	.260*	.214	.317*	.254
Teachers' assessment	-.547**	-.526**	-.521**	-.545**
Average translation mark	.701**	.660**	.665**	.687**
Gen. Translation B dir 1	.576**	.508**	.540**	.538**
Gen. Translation B dir 2	.562**	.545**	.520**	.559**
Gen. Translation C dir 2	.492**	.461**	.439**	.472**
Gen. Translation C inv 2	.478**	.443**	.499**	.478**
Spec. Translation B dir 3	.499**	.453**	.434**	.467**
Spec. Translation B inv 3	.498**	.465**	.452**	.482**
Gen. Translation C dir 3	.439**	.406**	.351**	.407**
Gen. Translation C inv 3	.340**	.379**	.327*	.380**

**p>.001; *p>.01

(2) Matrix of Factor Analysis

As the 17 external criteria were so varied in nature, a factor analysis (Varimax rotational method) was made to determine their underlying structure, and the results are reported in the following table:

TABLE 5

Factor analysis matrix of the external criteria

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Specialized Translation B dir 3	**.771**	-.005	.244	-.009
Spanish Language	**.740**	.103	.342	-.008
General Translation C inv 2	**.725**	.167	.003	.388
Specialized Translation B inv 3	**.719**	.164	.206	.008
Teachers' Assessment	**-.708**	-.426	-.269	-.214
English Language	**.694**	.276	.115	-.264
English Composition	**.635**	.377	.008	.009
Average Translation Mark	**.612**	.482	.391	.158
General Translation C dir 3	**.569**	-.132	.532	.108
General Translation C dir 2	**.523**	.260	.493	.005
Students' Self-Assessment 2	.208	**.818**	.140	.112
Students' Self-Assessment 3	.004	**.780**	.008	-.112
Verbal Test	.119	.112	**.836**	.006
General Translation B dir 1	.267	.468	**.595**	.004
General Translation B dir 2	.490	.389	**.492**	.207
Mathematical Series Test	-.137	-.001	.141	**.837**
General Translation C inv 3	.541	.003	.008	**.628**

The following table shows these four factors account for 67.7% of the total variance:

TABLE 6

Total variance accounted for by the four factors

Factors	Total variance	% of the variance	% accumulated
Factor 1	7.639	44.934	44.934
Factor 2	1.501	8.830	53.764
Factor 3	1.335	7.855	61.619
Factor 4	1.037	6.101	67.720

The first factor is defined by ten variables: six translation variables, three language variables, and the one of the teachers' global assessment of the students' ability to do this degree. It would seem reasonable to identify this factor as *Translation Competence*, which is composed of linguistic ability (English and Spanish in the present case), of the ability to translate into and from these two languages, and of the ability to translate in other language combinations. It is also worth noting that two other translation variables also have a certain weight in this main factor, *General Translation C inv 3* (.541) and *General Translation B dir 2* (.490), which also supports this identification. In fact, the only translation variable which has a weight under .490 is that of *General Translation B dir 1* (.267).

The two variables which make up the second factor clearly identify it as *Students' Self-Assessment of their Translation Competence*.

The third factor is also clearly defined as *Native Language Competence* (Spanish); it is made up of the test of verbal intelligence and the subjects of general translation from English into Spanish from the first two years of the degree. The test of verbal intelligence, which forms part of the battery of entrance tests for the degree in Translation and Interpreting, uses very sophisticated vocabulary, and it is not surprising that the degree of competence in Spanish language which enables the students to do this test well should also enable them to translate well from English into Spanish in the first two years of their studies, as both these translation subjects naturally stress the fundamental importance of expression in the native language.

The fourth factor appears to relate to *Mathematical Intelligence* as it consists of only two variables, the main one being the intelligence test of mathematical series which also forms part of the battery of university entrance tests which the 64 students took to enter the Faculty two years before the date of the second-year translation. What is much more difficult to explain is the presence in this factor of the variable *General Translation C inv 3*, the third year subject of general translation from Spanish into the C language (either French or German).

Although it is difficult to understand this fourth factor, it is relatively easy to define the remaining three and I wish to underline the importance of the first one: *Translation Competence*. This factor is made up not only of language variables but also of different types of translation (general/specialized, from the foreign languages into the native language and vice-versa) with different language combinations (Spanish/English, Spanish/French and Spanish/German), and the composition of this factor would seem to indicate the presence of a global translation competence which affects students' performance in apparently different situations. What connection is there, for instance, between the third-year subject of Specialised Translation from the students' B language (English) into Spanish, on the one hand, and the second year subject of General Translation from Spanish into the students' C language (either German or French), on the other? This factor analysis replies that these two variables are connected by a common competence.

4. Validity Study

4.1. *Correlations between the three factors and the variable*

The following table shows the matrix of correlations between the three first factors, the variable *series,* and the four methods. In the case of the fourth factor, it was decided to use only the variable *series*, which has a clear identity as *Mathematical Intelligence*, and remove the variable *General Translation C inv 3* as it was difficult to explain its presence here.

TABLE 7

Correlation matrix of the four methods and the four factors

	Method A	Method B	Method C	Method D	transcom	nlancom	selfas	series
Method A	1.000							
Method B	.946**	1.000						
Method C	.854**	.822**	1.000					
Method D	.955**	.986**	.906**	1.000				
transcom	.630**	.609**	.578**	.623**	1.000			
nlancom	.672**	.568**	.555**	.587**	.672**	1.000		
selfas	.626**	.381**	.470**	.424**	.356**	.479**	1.000	
series	.261*	.240	.164	.226	.037	.106	.031	1.000

**p>.001

(transcom: Translation Competence; nlancom: Native Language Competence; selfass: Self-Assessment)

This matrix shows that there is a moderately high, statistically significant correlation (p>.001) between the first factor (*Translation Competence*) and the second and third factors (*Native Language Competence* and *Self-Assessment*). However, none of these three factors correlates with the variable *series*, which lends weight to their construct validity in the sense that, whereas the three factors are clearly related to translation ability, it is not surprising that they do not relate to the variable *series*, which is based on a university entrance test designed to test mathematical intelligence.

4.2. *Criterion-related validity of the four methods*

To determine the criterion-related validity of the four methods of assessment, a study was made of the correlation between the results of their application by the five correctors and the three main factors which underlie the 17 external criteria, according to the findings of our factor analysis. As above, the fourth factor was reduced to the variable *series*.

(1) Factor 1 (*Translation Competence*) and the four Methods

The following table shows the correlations between the mean scores reached by the five correctors applying the four methods and the factor of *Translation Competence*:

TABLE 8

Correlations between Methods and Translation Competence

	Translation Competence
Method A	.630**
Method B	.609**
Method C	.578**
Method D	.623**

**p<.001

These results show that there is a moderately high, statistically significant correlation between the correctors' application of the four methods and the factor of

Translation Competence. This means that, as far as the question of criterion-related validity is concerned, none of the three hypotheses has been verified, as all four methods have proved to be equally valid in spite of the considerable differences that exist between them.

However, maybe these results should not surprise us if we take into account the care with which the methods were prepared and applied. As far as the preparation is concerned, the error analysis Method A had been widely used in our Faculty since the degree in Translation and Interpreting was launched in 1993 and, by the time the experiment took place in 1997, it was well known and appreciated by members of staff and students alike. The case of Method B was very different, as I designed it myself in an attempt to take into account the variable effect of translation mistakes and I was the only teacher in the Faculty to have used it before the experiment took place. However, by this time I had been developing it for two years and the students were apparently satisfied with the final version that was applied in 1997. The holistic method C had not been used in the Faculty at all, as it was designed especially for the experiment; however, two members of staff (Alan Allott and Beverly Rising) helped me with the formulation of the descriptors, and Alan Allott also helped me to apply it to a number of student translations (of different texts to the one used in the experiment) before it was used by the team of correctors.

As far as the application of the three methods is concerned, all the correctors received training in how to use each one before applying it to the 64 translations of the experiment. In the case of the two error analysis methods A and B, this training entailed first understanding the instructions, and then learning to apply each method accurately. To do this, the correctors were required to apply the new method to a set of student translations of a text that was different from the one used in the experiment; then each corrector saw me individually to compare the errors detected and the penalties applied with my own corrections of the same translations. In this way, doubts were ironed out and I was able to make sure that the corrector had grasped two important principles: the first was to take points off for everything that was clearly unacceptable and not to make allowances for the fact that the students were not native speakers of English; the second was the opposite extreme: to give the students the benefit of the doubt whenever the corrector was not sure the English used was unacceptable. That is to say, the five correctors were asked to apply the rules of good English strictly, but not to be too severe when it came to what could loosely be called "errors of style." They were also asked to check carefully that the students had transferred the whole content of the original text but no more, although in the case of cultural problems they could be expected to make necessary additions for the English reader. In the exam, the students were given the following instructions: "Your translation will be published in the English newspaper *The Times* as part of an article on European languages. Your text will figure as a translation of an article from a Spanish newspaper. Overleaf you will find the complete article, in case you wish to consult it."

In the case of the holistic method C, the training was slightly different, in that the correctors were not only provided with instructions on how to apply the method and student translations to correct, but also with seven student translations (of this same practice text) which represented each one of the marks from 3 to 9 and served as illustrations of the descriptors. The correctors were advised to correct these first

(after covering the marks), before correcting the other ten student translations of the same text.

As I did not ask other correctors to apply these methods without undergoing any kind of training, I do not have any evidence of the effect this had on the reliability and validity of these results. However, in the two interviews that I held with each corrector before the application of each method to the 64 translations (the first to explain the method and the second to check the accuracy of its application), the following questions related to validity were inevitably discussed:

(i) The contextualization of the exam: that is, the fact that the students were writing for the readers of *The Times* and that their text would figure clearly as a translation of an article written on a stated date in the Spanish newspaper *ABC*. It was pointed out to the correctors that, during the course the students had received, all the translations had been contextualized in this way and in the exam they were expected to take these pragmatic factors into account.

(ii) The need for simplification and clarification: as the students were translating into their first foreign language, the translation course they had followed before taking this exam had advised them constantly to resist the temptation to offer literal renderings of passages in the source text that were especially difficult due to syntactic complexity or the obscurity of the underlying ideas or presuppositions on the part of the ST author, and to unravel the syntax or the ideas, to paraphrase, and to try to present the content in plain, straightforward English. The correctors had to be warned about this, as three of them were not teachers of translation and might penalise such attempts unfairly.

(2) The four methods and factors 2 and 3 (*Native Language Competence* and *Self-Assessment*)

The following table shows the correlations between the mean scores reached by the five correctors and these two factors:

TABLE 9

Correlations between Methods and Native Language Competence

	Native Language Competence	Students' Self-Assessment of Translation Competence
Method A	.626**	.454**
Method B	.568**	.381**
Method C	.555**	.470**
Method D	.587**	.424**

**p<.001

Once again, there are moderately high, statistically significant correlations between the four methods and the factors of *Native Language Competence* and *Self-Assessment*, which corroborates the criterion-related validity of all four ways of assessing the quality of student translations.

(3) The four methods and the variable *series*

TABLE 10

Correlations between the four methods and the variable series

	Series
Method A	.261*
Method B	.240
Method C	.164
Method D	.226

*p<.04

As was to be expected, the correlations between the four methods and the variable *series* are very low and only statistically significant (p>.04) in the case of Method A. As *series* measures a trait, *Mathematical Intelligence*, which would appear not to play an important role in the ability to translate, these correlations support the criterion-related validity of the four methods.

5. Conclusions

Although the 17 external criteria used in the experiment were varied in nature, the factor analysis revealed the presence of four underlying factors, and the main one was clearly identified as *Translation Competence*. The correlational study then showed us that this main factor related closely to the other two factors of *Native Language Competence* and the *Students' Self-Assessment of their Translation Competence*, although none of these factors related to the main variable in the fourth factor, *series*, which consisted of *Mathematical Intelligence*.

The fact that all four methods, in spite of the considerable differences between them, showed statistically significant correlations with the three main factors, and (except for Method A) not with the variable *series*, supports the criterion-related validity of all four systems of assessment. The fact that differences were not detected between the validity of the methods means that the three hypotheses were not verified, but this may be explained not only by the methods themselves but also by the fact that they were carefully designed, tested and applied. In any case, it is reassuring to know that, given the circumstances under which this experiment took place, the different types of methods of assessing student translations in university faculties in Europe and Canada can produce results which have criterion-related validity.

REFERENCES

BENSOUSSAN, M. and J. Rosenhouse (1994): "Evaluating student translations by discourse analysis," *Babel*, 36-2, pp. 65-84.

CAMPBELL, S. J. (1991): "Towards a Model of Translation Competence," *Meta* 36-2/3, pp. 329-243.

DANCETTE, J. (1989): « La faute de sens en traduction », *TTR*, 2-2, p. 83-102.

—— (1992): « Modèles sémantique et propositionnel de l'analyse de la fidélité en traduction », *Meta*, 37-3, p. 440-449.

DARBELNET, J. (1977): « Niveaux de la traduction », *Babel*, 23-1, p. 6-17.

GOUADEC, D. (1981): « Paramètres de l'évaluation des traductions », *Meta*, 26-2, p. 99-116.

—— (1989): « Comprendre, évaluer, prévenir. Pratique, enseignement et recherche face à l'erreur et à la faute en traduction », *TTR*, 2-2, p. 35-54.

HATIM, B. and I. MASON (1997): *The Translator as Communicator,* London, Routledge.

HEWSON, L. (1995): "Detecting Cultural Shifts: Some Notes on Translation Assessment," *Cross-Words. Issues and Debates in Literary and Non-literary Translating* (I. MASON and C. PAGNOULLE, eds.), Liège, L3, Liège Language and Literature, pp. 101-108.

HOUSE, J. (1981): *A Model for Translation Quality Assessment*, Tübingen, Gunter Narr.

HURTADO ALBIR, A. (1995): "La didáctica de la traducción. Evolución y estado actual," *X Perspectivas de la Traducción,* (P. Fernández, ed.), Valladolid, Universidad de Valladolid, pp. 49-74.

KUSSMAUL, P. (1995): *Training the Translator*, Amsterdam, John Benjamins.

LAROSE, R. (1989): *Théories contemporaines de la traduction*, Montréal, Presses de l'Université de Québec.

NEWMARK, P. (1991): *About Translation*, Clevedon, Multilingual Matters.

NORD, C. (1993): *La evaluación de errores en la enseñanza de traducción*, summary of postgraduate course given in the Universidad de Las Palmas (Canary Islands).

—— (1996): "El error en la traducción: categorías y evaluación," *La enseñanza de la traducción* (A. Hurtado Albir, ed.), Castelló de la Plana, Publicacions de la Universitat Jaume I, coll. «Estudis sobre la traducció», nº 3, pp. 91-103.

PYM, A. (1992): "Translation Error Analysis and the Interface with Language Teaching," *Teaching Translation and Interpreting. Training, Talent and Experience. Papers from the First Language International Conference, Elsinore, Denmark, 31 May–2 June, 1991* (C. DOLLERUP and A. LODDEGAARD, eds.), Amsterdam, John Benjamins, pp. 279-288.

SAGER, J. C. (1989): "Quality and Standards: The Evaluation of Translations," *The Translator's Handbook* (C. Picken, ed.), London, ASLIB, pp. 91-102 [This is the second edition of *The Translator's Handbook* (1983)].

SÉGUINOT, C. (1989): "Understanding Why Translators Make Mistakes," *TTR*, 2-2, pp. 73-102.

—— (1990): "Interpreting Errors in Translation," *Meta*, 25-1, pp. 68-73.

SNELL-HORNBY, M. (1995): "On Models and Structures and Target Text Cultures: Methods of Assessing Literary Translations," *La Traducció Literària* (Josep Marco BORILLO, ed.), Castelló de la Plana, Publicacions de la Universitat Jaume I, coll. "Estudis sobre la traducció," nº 2, pp. 43-58.

STANSFIELD, C. W., M. L. Scott and D. M. KENYON (1992): "The Measurement of Translation Ability." *The Modern Language Journal*, 76-iv, pp. 455-67.

WADDINGTON, C. (1999): *Estudio comparativo de diferentes métodos de evaluación de traducción general (Inglés-Español)*, Madrid, Publicaciones de la Universidad Pontificia Comillas.

—— (2000a): "Measuring the Effect of Errors on Translation Quality," paper presented at the Saarbrücker Symposium on *Translation and Interpretation: Models in Quality Assessment*, held at the Universität des Saarlandes 9th–11th March 2000.

—— (2000b): "Should student translations be assessed holistically or through error analysis?," paper presented at the CIUTI Research Seminar on *Text & Translation*, held at the Aarhus School of Business 30th May 2000 (to be published in the School's magazine *Hermes*).

WILLIAMS, M. (1989): "The Assessment of Professional Translation Quality: Creating Credibility out of Chaos," *TTR*, 2-2, pp. 13-33.