



**TRANSLATION AT INTER-GOVERNMENTAL ORGANIZATIONS THE SET OF  
SKILLS AND KNOWLEDGE REQUIRED AND THE IMPLICATIONS FOR  
RECRUITMENT TESTING**  
**Anne Lafeber**

**Dipòsit Legal: T. 879-2012**

**ADVERTIMENT.** L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

**ADVERTENCIA.** El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

**WARNING.** Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

Anne Lafeber

TRANSLATION AT INTER-GOVERNMENTAL ORGANIZATIONS: THE SET OF SKILLS  
AND KNOWLEDGE REQUIRED AND THE IMPLICATIONS FOR RECRUITMENT  
TESTING

DOCTORAL THESIS

Supervised by Dr. Anthony Pym and Dr. Christopher Scott-Tennent

Department of English and German Studies



UNIVERSITAT ROVIRA I VIRGILI

Tarragona

2012

UNIVERSITAT ROVIRA I VIRGILI

TRANSLATION AT INTER-GOVERNMENTAL ORGANIZATIONS THE SET OF SKILLS AND KNOWLEDGE REQUIRED AND THE IMPLICATIONS  
FOR RECRUITMENT TESTING

Anne Lafeber

Dipòsit Legal: T. 879-2012



Professor Anthony Pym

URV. Av. Catalunya 35

43002 Tarragona, Spain

anthony.pym@urv.cat

March 18, 2012

I hereby certify that the present study *Translation at inter-governmental organizations: The set of skills and knowledge required and the implications for recruitment testing*, presented by Anne Lafeber for the award of the degree of Doctor, has been carried out under the supervision of myself and the late Dr. Christopher Scott-Tennent at the Department of English and German Studies of the Rovira i Virgili University, and that it fulfills all the requirements for the mention "Doctor Europeus".

Professor Anthony Pym  
Intercultural Studies Group  
Universitat Rovira i Virgili  
Tarragona, Spain

President  
European Society for Translation  
Studies

UNIVERSITAT ROVIRA I VIRGILI

TRANSLATION AT INTER-GOVERNMENTAL ORGANIZATIONS THE SET OF SKILLS AND KNOWLEDGE REQUIRED AND THE IMPLICATIONS  
FOR RECRUITMENT TESTING

Anne Lafeber

Dipòsit Legal: T. 879-2012

## Acknowledgements

I owe my deepest gratitude to my supervisors at Universitat Rovira i Virgili, Dr. Anthony Pym and Dr. Christopher Scott-Tennent. Their guidance, their meticulous revision and their incisive comments have been invaluable. Through them I have discovered the fascinating world of academic research, and they have steered me deftly through the various stages, keeping me on course and patiently coaching me along the way. The momentum of the research was kept up due, in no small part, to the timely feedback received throughout from Dr. Pym and the enthusiastic encouragement provided by Dr. Scott-Tennent. The unexpected death of Dr. Scott-Tennent cast a dark shadow over the last stage of the preparation of this thesis, and it saddens me greatly that I cannot present the final product to him.

The research would have been impossible to carry out without the hundreds of professional translators who took the time to answer the surveys which I sent out to international organizations, and I am grateful to IAMLADP, the International Annual Meeting on Language Arrangements, Documentation and Publications, for circulating the different questionnaires among its members. I am also indebted to the students from the universities of Bath, Geneva, Leeds and Westminster and the Monterey Institute of International Studies who participated in the test trial and to their coordinators for administering the tests. Special thanks are due to the translators and revisers who willingly participated not only as test-takers, but also as graders in the trial, and particularly to my colleagues Nicole Maguire and Nicholas Whitehead for tackling a major portion of the grading.

I am also extremely grateful to my employer, the United Nations Secretariat, for funding some of the coursework and granting me study leave during the writing stage, and to my direct supervisors, Sally Shaw, at ECLAC, and Graham Hobbs, at UNOG, for their constant support. The expert help of the United Nations statisticians Dr. Humberto Soto and Dr. Muhammad Atiyat is also particularly appreciated. It should be noted that the views expressed in this thesis are my own and do not necessarily reflect the views of the United Nations.

I would like to make a special mention of my colleague and dear friend, Diane Frishman, whose advice and encouragement have been invaluable over the last four years. Finally, I would like to thank my husband, Roberto Correa, without whose unwavering support and enthusiasm this research project would never have been started and this thesis would never have been written. I owe him so much.



## Abstract

Translation at inter-governmental organizations (IGOs) is a branch of professional translation that calls for highly competent translators, and succession planning poses a major challenge. There is, therefore, both a practical and a moral imperative to make recruitment testing as fair and reliable as possible. The research focuses on the decisive element in the recruitment process: text-based translation tests. How effective are current tests as instruments of selection? Do they screen out the translators who are best suited to the translation work at the IGO in question? What combination of skills and knowledge should IGOs be looking for in candidates?

A survey was chosen as the tool for identifying and ranking the skills and knowledge that IGO translators need. Two questionnaires were distributed: one on the impact of the different components of the skills-knowledge set and one on the frequency with which new recruits lack the required skills and knowledge types. Over 320 translators and revisers working at 24 IGOs participated in the survey. The principles of qualitative risk analysis were applied to the results to draw up ideal candidate profiles for IGO translation work in general, as well as for individual translation services. The implications for recruitment testing were then explored by examining current testing practice in the light of the candidate profiles. A survey was made of testing at 18 IGO translation services, and the recruitment procedures of two individual organizations were analyzed in detail. A test trial, involving 40 M.A. students from five training institutions, as well as IGO translators and revisers, was then run to determine whether tests based on empirically-identified recruitment priorities might be better than traditional tests at identifying the most suitable candidates for a specific organization.

The findings suggest that IGO translators need far more than language skills. They also need research, computer, analytical and interpersonal skills, as well as extensive general knowledge and, possibly, specialized subject knowledge. Moreover, some skills and knowledge types are more important than others. Certain analytical skills are at least as important as, for example, source-language knowledge and target-language writing skills. The profile varies from organization to organization and by language service. Testing practices also vary considerably, and not all the important components of the required skills-knowledge set are covered in current recruitment examinations. The most significant omission is the non-measurement of analytical skills. The results of the test trial suggest that profile-adapted testing is possibly a more valid way to identify candidates with the skills-knowledge set sought by IGOs.





## Contents

List of acronyms used .....	1
<b>Chapter 1 Introduction .....</b>	<b>3</b>
<b>Chapter 2 Literature review .....</b>	<b>9</b>
2.1 Translation competence .....	9
2.1.1 The lack of empirical evidence .....	16
2.1.2 The translator-training focus of research .....	16
2.1.3 The absence of attempts to rank the components of translation competence ...	17
2.2 IGO translation.....	17
2.2.1 The institutional nature of IGO translation work.....	18
2.2.2 The emphasis on quality .....	19
2.2.3 The need for standardized language use .....	19
2.2.4 The influence of language policies .....	20
2.2.5 The range of subjects and document-types covered .....	20
2.2.6 The need for clarity .....	20
2.2.7 The poor quality of some source texts .....	21
2.2.8 The varied use of technology .....	21
2.2.9 Turnaround times .....	22
2.2.10 Summary of the review of the literature on IGO translation .....	22
2.3 Translation testing.....	24
2.3.1 The principles of test design .....	25
2.3.1.1 Authenticity and interactiveness .....	26
2.3.2 Task difficulty .....	29
2.3.3 Scoring .....	33
2.4 Summary of the literature review .....	39

**PART I.....41**

<b>Chapter 3 Methodology for the identification and ranking of the skills and knowledge IGOs should seek in their recruitment processes .....</b>	<b>43</b>
3.1. Choosing a methodology .....	43
3.1.1. How to approach importance? .....	43
3.2 The first pilot.....	45
3.2.1 The base list of skills and knowledge types presented for rating in both questionnaires .....	45
3.2.2 The phrasing of the questions in the first impact questionnaire .....	48
3.2.3 Other questions included in the impact questionnaire .....	49
3.2.4 The phrasing of the questions in the recruits questionnaire.....	49
3.2.5 Other questions included in the recruits questionnaire .....	49
3.2.6 The respondents .....	50
3.2.7 Other participants in the pilot .....	51
3.2.8 Assumptions.....	51
3.2.9 Procedure .....	52
3.2.10 Results of the first pilot of the impact questionnaire .....	52
3.2.10.1 The impact ratings.....	52
3.2.10.2 Skills and knowledge types reported as missing from the base list....	53
3.2.11 Analysis of the results of the first pilot of the impact questionnaire .....	54
3.2.12 Ways in which to improve the reliability of the impact questionnaire as a measuring instrument.....	55
3.2.13 Results of the recruits questionnaire .....	56
3.2.13.1 The frequency ratings .....	56
3.2.13.2 Skills and knowledge types reported as missing from the base list....	58
3.2.14 Analysis of the results of the recruits questionnaire .....	58
3.2.14.1 The understanding of “new” in “new recruit” .....	58
3.2.14.2 The in-house/external work factor .....	58
3.2.14.3 The revision style factor.....	59
3.2.14.4 The influence of anecdotal evidence.....	59
3.2.14.5 The answers to the context questions.....	60
3.2.15 Reactions to the questions on expectations and training .....	60

3.2.16	Ways in which to improve the reliability of the recruits questionnaire .....	61
3.2.17	Conclusions drawn from the first pilot .....	61
3.3	The second pilot.....	62
3.3.1	The revised questionnaires.....	62
3.3.2	The respondents .....	65
3.3.3	Assumptions.....	66
3.3.4	Procedure .....	66
3.3.5	Results of the second pilot of the impact questionnaire .....	66
3.3.5.1	Profile of the respondents .....	66
3.3.5.2	The impact ratings.....	67
3.3.5.3	Analysis of differences between the revisers' and the translators' answers.....	69
3.3.5.4	Mention of other skills and knowledge required .....	70
3.3.5.5	Expectations regarding when skills and knowledge should be acquired.....	71
3.3.5.6	General comments on the questionnaire .....	72
3.3.6	Results of the recruits questionnaire .....	72
3.3.6.1	Profile of the respondents .....	72
3.3.6.2	The frequency ratings .....	72
3.3.6.3	Other skills reported as lacking and needed .....	74
3.3.6.4	Document processing and translators' responsibilities .....	75
3.3.6.5	Comments made in the recruits questionnaire .....	75
3.3.6.6	Results of the semi-formal follow-up interviews with respondents .....	76
3.3.7	Analysis of the results of the second pilot .....	76
3.3.7.1	Comparison of the ratings awarded by the revisers who participated in both pilots .....	76
3.3.7.2	The influence of different variables on the ratings awarded in the two questionnaires .....	80
3.3.7.2.1	The influence of the in-house/external work variable. ....	81
3.3.7.2.2	The influence of the language service.. ....	83
3.3.8	Conclusions drawn from second pilot.....	86
3.3.8.1	The reliability and validity of the questionnaires as measuring instruments.....	86
3.3.8.2	The hierarchies produced by the questionnaires and their usefulness as input for identifying recruitment priorities.....	87

3.4	Methodology for identifying recruitment priorities .....	90
3.5	Conclusions on the methodology for identifying the skills and knowledge required in IGO translation.....	93
<b>Chapter 4</b>	<b>Results of the survey of the skills and knowledge required in IGO translation .....</b>	<b>95</b>
4.1	The revised questionnaires.....	95
4.2	The survey sample .....	95
4.3	Procedure .....	96
4.4	The results of the impact questionnaire .....	96
4.4.1	The respondents .....	96
4.4.2.	The impact ratings.....	98
4.4.3	The additional skills and knowledge reported in the impact questionnaire as being needed on the job .....	100
4.4.4	The skills and knowledge that can be acquired after recruitment.....	102
4.4.5	The profile(s) of the “ideal candidate” emerging from the findings of the impact questionnaire .....	103
4.5	The results of the recruits questionnaire .....	104
4.5.1.	The respondents.....	104
4.5.2	The frequency ratings .....	105
4.5.3	Comments on other skills and knowledge found lacking and general observations made about new recruits and training .....	108
4.5.4	The expectations of translators in the document production process .....	110
4.5.5.	The profile(s) of the “ideal candidate” emerging from the findings of the recruits questionnaire .....	112
<b>Chapter 5</b>	<b>Discussion of the results of the skills and knowledge survey.</b>	<b>113</b>
5.1	Analysis of the influence of different variables on the impact ratings... ..	117
5.1.1	The influence of the organization for which respondents answered .....	117
5.1.2	The influence of position: the perspective of the reviser versus the perspective of the translator .....	119
5.1.3	The influence of the in-house/external work factor .....	121
5.1.4	The influence of the target language involved.....	121
5.2	Analysis of the influence of different variables on the frequency ratings .....	122

5.2.1	The organization factor.....	122
5.2.2	The skills and knowledge most often lacking, by language service.....	125
5.2.3	The influence of the in-house/external work factor .....	126
5.3	The results as input for the identification of recruitment priorities at international organizations .....	126
5.4	Summary of the main findings regarding the skills-knowledge set required in IGO translation.....	134
<b>PART II .....</b>		<b>137</b>
<b>Chapter 6 Introduction.....</b>		<b>139</b>
<b>Chapter 7 The testing survey .....</b>		<b>141</b>
7.1	Method.....	141
7.2	Results of the testing survey .....	142
7.2.1	Procedure .....	142
7.2.2	The respondents .....	142
7.2.3	Findings relevant to identifying the coverage and weighting of the components of the skills-knowledge set .....	142
7.2.3.1.	Purpose of the examinations .....	142
7.2.3.2	Pre-test screening.....	143
7.2.3.3	Test content .....	143
7.2.3.4	Test length.....	144
7.2.3.5	Text types used in examinations .....	145
7.2.3.6	Test-text features .....	145
7.2.3.7	Test-text selection and preparation .....	150
7.2.3.8	Test conditions (use of computers, provision of reference material, access to the Internet and reference material).....	150
7.2.3.9	Test grading.....	151
7.2.3.10	General attitudes to errors in current testing practice .....	152
7.2.3.11	What organizations are looking for.....	155
7.2.3.12	Comments made.....	156
7.3	Discussion of the results of the testing survey and conclusions drawn about the coverage and weighting of skills and knowledge in current IGO testing practice ....	157

## **Chapter 8 Current testing practice: a closer look at two organizations.163**

8.1	Method for the analysis of testing practice at two organizations.....	163
8.2	The weighting of skills and knowledge types in testing at organization X .....	163
8.3	The weighting of skills and knowledge types in testing at organization Y .....	172
8.4	Conclusions of the close analysis of testing at two organizations .....	178

## **Chapter 9 The design of a test trial to compare performance on a typical recruitment test and a profile-adapted one.....181**

9.1	The design of the mock traditional test (Test One).....	182
9.1.1	The test task .....	182
9.1.2	Test conditions .....	187
9.1.3	Instructions for test-takers and mode of delivery .....	187
9.1.4	Grading scheme .....	188
9.1.5	Instructions for graders .....	188
9.2	The design of the profile-adapted test.....	190
9.2.1	Identification of the construct(s) to be measured.....	190
9.2.2	Preliminary considerations for the design of the test task .....	193
9.2.2.1	Authenticity .....	194
9.2.2.2	Activation power .....	195
9.2.2.3	The level of difficulty.....	195
9.2.3	Activating the A components .....	196
9.2.4	Activating all the components .....	197
9.2.5	Preparation of the ST .....	199
9.2.6	Test conditions .....	201
9.2.7	Instructions for test-takers and mode of delivery .....	201
9.2.8	Design of the grading scheme in light of the skills-knowledge hierarchy.....	204
9.2.9	Calculating the final scores.....	211
9.2.10	Fairness, validity and reliability considerations.....	211
9.2.10.1	Fairness .....	211
9.2.10.2	Validity considerations.....	212
9.2.10.3	Reliability considerations.....	213
9.3	The pilot.....	213

9.3.1	Participants.....	214
9.3.2	Test conditions and mode of delivery .....	214
9.3.3	Grading .....	215
9.3.4	Results of the pilot .....	215
9.3.5	Outcomes of the pilot.....	217
9.4	Organization of the test trial.....	217
9.4.1	The test-takers .....	217
9.4.2	The graders.....	218
9.4.3	The administration of the two tests .....	218
9.4.4	Feedback from participants.....	219
9.4.5	Preparation of the test TTs for grading .....	219
9.4.6	The organization of the grading .....	220
9.4.7	Preparation of the teams of graders .....	220
9.4.7.1	Those working with scheme A.....	220
9.4.7.2	Those working with scheme B.....	220
9.4.7.3	Feedback from graders.....	221
<b>Chapter 10</b>	<b>Results of the test trial.....</b>	<b>223</b>
10.1	The test-takers .....	223
10.2	The test scores .....	224
10.3	Comments made by test-takers.....	224
10.3.1	On the instructions.....	224
10.3.2	On the timing.....	225
10.3.3	On their experience as test-takers.....	226
10.4	Comments made by graders on the schemes used .....	227
<b>Chapter 11</b>	<b>Analysis of the results of the test trial.....</b>	<b>229</b>
11.1	The reliability of the grading schemes .....	229
11.1.1	The reliability of the scores obtained using scheme A (intuitive and points scoring) .....	229
11.1.2	The reliability of the scores obtained using scheme B (multi-trait grading) ..	232
11.1.3	Conclusions about the reliability of the scoring methods and treatment of the scores.....	234



11.2	Calculating composite scores using scheme B .....	235
11.2.1	The self-weighting of components in Test Two .....	236
11.2.2	Weighting options .....	237
11.3	Comparison of performance on the two tests .....	243
11.3.1	The relative influence of the test task and the grading scheme used .....	244
11.3.2	The implications for test-takers and test users .....	244
11.4	The influence of test-taker variables .....	249
11.4.1	The training institution .....	249
11.4.2	Age .....	250
11.4.3	Sex .....	250
11.4.4	Academic background .....	251
11.4.5	Professional experience .....	252
11.4.6	Self-assessment and ranking by teachers .....	255
11.5	Summary of the analysis of the test trial results .....	258
<b>Chapter 12</b>	<b>Main findings, limitations and suggestions for future research.....</b>	<b>259</b>
12.1	Main findings .....	259
12.2	Limitations of the research .....	261
12.3	Suggestions for future research .....	263
12.4	Final remarks .....	264
<b>References</b>	<b>.....</b>	<b>267</b>
<b>Appendixes</b>	<b>.....</b>	<b>275</b>
Appendix 1.	The impact questionnaire used in the first pilot .....	275
Appendix 2.	The recruits questionnaire used in the first pilot .....	278
Appendix 3.	How revisers become aware of translators' weaknesses. ....	281
Appendix 4.	The impact questionnaire used in the second pilot.....	282
Appendix 5.	The recruits questionnaire used in the second pilot .....	284
Appendix 6.	Final version of the impact questionnaire .....	286
Appendix 7.	Final version of the recruits questionnaire .....	288
Appendix 8.	Summary of the comments made in the impact and recruits questionnaires on IGO translation .....	290

Appendix 9.	Questions asked in the questionnaires on ad hoc tests and competitive examinations.....	292
Appendix 10.	Graders survey.....	295
Appendix 11.	The abbreviations used to refer to the skills and knowledge types rated in the second pilot and the changes made to the base list after the second pilot.....	296
Appendix 12.	Source text used in Test One. ....	298
Appendix 13.	Base texts used to develop the source text translated in the profile-adapted test.....	299
Appendix 14.	Source text used in Test Two (the profile-adapted test) .....	300
Appendix 15.	Questions included in the feedback questionnaires for test trial participants .....	302
Appendix 16.	Test One: scores and rankings obtained using grading scheme A (the intuitive and the points-deduction methods) .....	303
Appendix 17.	Test Two: scores and rankings, obtained using grading scheme B (multi-trait scoring) .....	304
Appendix 18.	Test One: scores obtained from the application of scheme B .....	305
Appendix 19.	Test Two: scores obtained from the application of scheme A .....	306
Appendix 20.	Test One and Test Two: mean intuitive scores .....	307
Appendix 21.	Test One and Test Two: mean component scores.....	308
Appendix 22.	Test Two: scores and rankings by weighting of A components and Non-A components .....	309
Appendix 23.	Test Two: composite scores obtained by weighting components according to impact and frequency ratings (profile weighting) and by weighting A components and non-A components 3:1 .....	310
Appendix 24.	Test One: composite scores obtained by weighting A and non-A components 3:1.....	311



## List of acronyms used

EBRD	European Bank for Reconstruction and Development
ECA	Economic Commission for Africa
ECLAC	Economic Commission for Latin America and the Caribbean
ESCWA	Economic and Social Commission for Western Asia
EESC	European Economic and Social Committee
EPO	European Patent Office
EU	European Union
FAO	Food and Agricultural Organization
IAMLADP	International Annual Meeting on Language Arrangements, Documentation and Publications
ICAO	International Civil Aviation Organization
ICRC	International Committee of the Red Cross
ICT	Information communication technologies
IDB	Inter-American Development Bank
IFAD	International Fund for Agricultural Development
IGO	Inter-governmental organization
IIC	Inter-American Investment Corporation
ILO	International Labour Organization
IMO	International Maritime Organization
IMF	International Monetary Fund
ITLOS	International Tribunal for the Law of the Sea
ITU	International Telecommunication Union
L1	Mother tongue/ native language/ language in which most comfortable working
L2	Second language/ source language from which most comfortable translating
L3	Third language/ second source language from which most comfortable translating
NATO	North Atlantic Treaty Organization
OAS	Organization of American States
SC	Source culture
SELA	Latin American Economic System
SL	Source language
ST	Source text
TC	Target culture
TL	Target language
TT	Target text
UN	United Nations
UNHQ	United Nations Headquarters
UNOG	United Nations Office at Geneva
UNON	United Nations Office at Nairobi
UNOV	United Nations Office at Vienna
WB	World Bank
WTO	World Trade Organization



## Chapter 1 Introduction

Translation is a complex activity in which many different abilities are called into play. In addition to essential comprehension skills in the source language and writing skills in the target language, translators will, to a greater or lesser extent according to the context of the task in question, need general knowledge, specialized knowledge, technical skills, research skills and a number of interpersonal skills as well. Translation at inter-governmental organizations (IGOs)<sup>1</sup> is a branch of professional translation that calls for highly competent translators. The stakes are high: IGOs rely on their translation services to facilitate communication among member States, and the effectiveness of those services in terms of the quality and speed of their output depends on the caliber of the translators they employ.

This has led IGOs to organize large-scale competitive examinations or to arrange ad hoc translation tests to recruit the best translators they can find. The reliability of those processes is becoming increasingly important in the face of the rising demand for competent translators. Succession planning for language staff is posing a major challenge to IGOs. At the United Nations Department for General Assembly and Conference Management, for example, an average of 382 staff will retire each year between 2011 and 2015 (DGACM 2011a). Another major translation service, the Directorate General for Translation of the European Commission, has 1,750 full-time translators on its pay-roll (DGT 2008: 4), yet according to Piet Verleysen, acting Director-General, with more pages to translate in a growing number of languages, certain language departments will find it increasingly difficult to recruit a sufficient number of highly qualified translators (Verleysen 2010). And these are just the core bodies of these two institutions. Both the United Nations and the European Union have dozens of specialized agencies and departments that also need the services of professional translators. Apart from the United Nations and the European Union, there are of course many other IGOs who employ language staff, even if not in such large numbers. Increasingly, IGOs are outsourcing their translation work to external contractors and hiring freelance translators to work temporarily in-house in order to meet demand. Some 28% of translation work is done externally at the European Commission (DGT 2011), for example, and over 20% at United Nations headquarters in New York (DGACM 2011b). The proportion

---

<sup>1</sup> Inter-governmental organizations (IGOs) are understood to be organizations comprised mainly of sovereign States that are usually referred to as “member States”.

ranges from 0% to 100% and averages slightly under 50% among the IGOs belonging to IAMLADP, the International Annual Meeting on Language Arrangements, Documentation and Publications.<sup>2</sup>

Outsourcing may broaden the pool of translators, but it does not completely alleviate the recruitment burden. Many organizations prefer to outsource to individual freelance translators than to agencies, and each one has to be tried and tested. Recruiting both internal and external translators consumes large amounts of time and financial resources. In the absence of any universally recognized accreditation, IGOs select translators almost exclusively through examinations that they set themselves. The formal competitive examinations held by the large IGOs have a reputation for being difficult to pass. For example, of the 38,231 persons who applied to sit the 55 competitive examinations for translators held by the United Nations between 2005 and 2009, 22,938 (60%) were called to the written tests, which are held at designated sites and times around the globe. Only 2,293 candidates (10%) passed the written tests and were called to the interview (the second part of the examination), and of those, only 583 (25%) were placed on the roster – a pass rate of 2.54% for the written and oral examinations combined (DGACM 2011a). Success rates at the European Commission are similarly low: they vary between 1% and 10%, with numbers of applicants for each session ranging from 250 to 1,500 depending on the target language (Wagner et al. 2002: 31-32).

Not all examinations have as many applicants or are so difficult to pass, of course. Some are much less formal procedures in which candidates take the examination online from the location of their choice, and sometimes at their own pace. In all recruitment testing the stakes are high, however, since mistakes in candidate selection are expensive and difficult to correct. Moreover, people's careers, livelihoods and futures are involved. There is therefore both a moral and a practical imperative to make recruitment testing as fair and reliable as possible. As the examination figures show, the problem is not so much finding translators as finding ones who make the grade. When supply is abundant, recruiters can afford to hire only candidates who perform brilliantly on difficult tests, secure in the knowledge they have identified translators of the highest caliber. When supply is running short, the bar may have to

---

<sup>2</sup> Report of the Drafting Group on Contracting for Language Services for IAMLADP 2010, Annex 1, restricted access. IAMLADP is a forum and network of managers of international organizations employing conference and language services providers (mainly translators and interpreters), whose membership includes the organizations in the United Nations system, European Union institutions, the Organization for Economic Cooperation and Development, the Organization for Security and Co-operation in Europe, the Economic Community of West African States and many other prominent international bodies.

be lowered and recruiters will possibly have to hire translators who perform less brilliantly on difficult tests. These candidates might not be so easy to identify with confidence, given that it is most often only the excellent or the hopeless who stand out. This will make it more important to be able to rank candidates in a valid and reliable manner. All recruitment tests for translators basically consist of having candidates perform one or more written translations. Some organizations also include written tests of general or specialized knowledge, verbal reasoning or other skills. A minimum level of education, mostly a Bachelor's degree in a relevant field, is required, and references and interviews are used to identify the soft skills, usually after successful completion of the written test(s). Decisive elements in all selection procedures, however, are the text-based translation tests, and these are the focus of the research presented here.

How effective are the text-based translation tests as instruments of selection? Do they screen out the translators who are best suited to the translation work at the IGO in question? What combination of skills and knowledge should IGOs be looking for in candidates? Those questions were the point of departure for the research. My experience as a translator who had taken several IGO recruitment examinations and later as an IGO in-house reviser with responsibility for setting and grading tests for translators suggested there was a lack of theoretical support for those involved specifically in the testing of translators for employment purposes. There were no recommendations or guidelines, let alone manuals or standards. My aim at the beginning of this research was to begin to fill that gap.

All test construction starts (or should start) with a needs analysis. Test developers require a clear idea of what the test is supposed to measure. My first area of research was therefore the combination of skills and knowledge that IGOs should be looking for in candidates. I conducted a survey of IGO translators and revisers and applied the principles of risk analysis to the results to classify skills and knowledge types according to their importance. The resulting skills-knowledge hierarchies would serve as a basis for identifying adjustments and changes that might need to be made to current testing practice in order to help IGOs find candidates with the right profile. I used the hierarchy obtained for a specific organization to devise a new type of text-based translation test, which I have called a "profile-adapted test". I also conducted a survey of translation testing at IGOs and analyzed testing practice more closely at two organizations to identify the main features of current IGO translation tests. That information was used to design a "mock traditional test". Finally, I compared performances on the "profile-adapted test" and the "mock traditional test" to



explore whether tests based on empirically-identified recruitment priorities might be better at identifying the most suitable candidates for a specific organization.

It should be noted that, as the focus was on text-based translation tests, social skills and psychological traits, such as the ability to work in a team or patience, were not included in the research. They may be as important as those associated with independent text production, but they possibly can only be properly assessed through on-the-job observation. The phrase “skills and knowledge types” in the thesis title thus refers to the skills and knowledge used by translators to produce written translations in isolation.

The specific questions the research set out to answer, the steps taken to try to answer them and the number of subjects involved are presented in Table 1.1, together with a timeline for the research.

Table 1.1. Outline of the research, research questions, steps and subjects

---

*Research question 1. Are certain skills and knowledge types more important in IGO translation than others? If so, which ones? and*

*Research question 2. Does the profile vary significantly according to the organization, language service or type of IGO translation work?*

August 2009      Step 1. First pilot of questionnaires on skills and knowledge types

Subjects involved: 2 revisers at 1 IGO

November 2009    Step 2. Second pilot of questionnaires on skills and knowledge types

Subjects involved: 4 revisers and 3 translators at 1 IGO

February 2010    Step 3. Skills and knowledge survey of international organizations

Subjects involved: 320 translators, revisers and heads of service working at 24 IGOs

*Research question 3. Do IGO text-based recruitment tests cover all the components of the skills-knowledge set that they should?*

March 2010      Step 4. Pilot of the IGO testing survey

Subjects involved: 2 revisers at 1 IGO

April 2010      Step 5. IGO testing survey

Subjects involved: 18 IGO language services





## Chapter 2 Literature review

The literature reviewed for this study covers three areas: translation competence, IGO translation, and translation testing.

### 2.1 Translation competence

IGOs are looking for people who can translate well to fill their translation posts. Scholars in the field of Translation Studies have yet to agree, however, on how to name, let alone define and conceptualize what it takes to translate well. Orozco and Hurtado Albir (2002: 375) identify six different names for the concept: *transfer competence*, *translational competence*, *translator competence*, *translation performance*, *translation ability* and *translation skill*. Gile (1995: 20) and PACTE (2011) meanwhile speak of *translation expertise*. Most, however, refer to *translation competence*.

As Orozco and Hurtado Albir (2002) and Pym (2003) note, translation competence has been used to mean many different things. At its simplest, it can be defined as “the knowledge and skills the translator must possess in order to carry out a translation” (Bell 1991: 43). Competence is associated with the ability to perform up to a certain standard, however, not just to execute an action, as pointed out by Englund Dimitrova (2005: 12). Englund Dimitrova distinguishes between translation competence and translation ability, which she sees as referring to the innate ability of any person with knowledge of more than one language. Competence is also relative: different translators may possess the necessary skills and knowledge to a greater or lesser extent: “[c]ompetence refers to qualities, skills and abilities, and it is not an absolute, but can be present in different degrees” (Englund Dimitrova 2005: 16). This approach is helpful in the context of testing, but what is the difference between skills and abilities?

A helpful definition of skills, which is used in this thesis, is provided in the European Qualifications Framework:

‘[S]kills’ means the ability to apply knowledge and use know-how to complete tasks and solve problems. [...] skills are described as cognitive (involving the use of logical,

intuitive and creative thinking) or practical (involving manual dexterity and the use of methods, materials, tools and instruments). (European Parliament Council 2008).

The Framework's definition of knowledge is equally useful: "knowledge' means the outcome of the assimilation of information through learning. Knowledge is the body of facts, principles, theories and practices that is related to a field of work or study" (op.cit.).

So what are the skills and knowledge that need to be measured in IGO translation tests? To answer that question and draw up an initial list of required skills and knowledge types, we have to return to the discussion of translation competence. Juliane House complained that "[a]lthough many authors talk of translation competence [...], few authors define the specific skills that constitute this competence" (1998: 199). Many authors have put forward models, however, which break translation competence down into a certain number of subcompetences: Campbell (1991), Cao (1996), Hatim and Mason (1997), PACTE (2003), Nord (2005), Kelly (2005), to name but a few. These have been discussed and their proliferation commented on by others in the field, such as Orozco and Hurtado Albir (2002), Pym (2003), Arango-Keeth and Koby (2003), Way (2008) and Angelelli (2009). Several, contrary to what House says, in fact go to great pains to specify all the skills involved in translating. The problem with trying to identify specific skills, as Pym (2003) points out, is that the list is potentially endless: it is in fact hard to think of a skill, knowledge-type, personal quality or experience that might not at some point be useful to a translator, especially with the increasingly wide range of tasks that translators are called upon to perform. Way agrees: "the first basic models and the minimalist approach have given way to increasingly complex models which have continued to be developed in recent years" (2008: 91). She goes on to say, as Pym also suggests, that this trend is both "in line with the demands of changing needs in society and the move towards a new model of higher education, particularly in the European Union" (2008: 91). For example, the Expert Group set up by the Directorate-General for Translation of the European Commission to make specific proposals for the European Master's in Translation presents a model of translation competence that consists of six "competences" and 48 different skills, knowledge types and abilities (EMT 2009). This model stands at the other end of the spectrum from Pym's back-to-basics approach which defines translation competence as just two abilities: "the ability to generate a series of more than one viable target text (TT<sub>1</sub>, TT<sub>2</sub> ...TT<sub>n</sub>) for a pertinent source text (ST);" and "the ability to select one viable TT from this series, quickly and with justified confidence" (2003). In between, there are numerous other schemes, which diverge to a

greater or lesser extent in their conceptualization, labeling and understanding of the interrelations of the components of translation competence. As Beeby points out, “[t]he choice of model will depend on the type of problem to be observed” (2001: 42). Likewise the extent to which the model is broken down into subcomponents and specified depends on its intended use.

A sample of the models found in the literature is presented in Table 2.1. The models are broken down into source, target and other components as far as this is possible. The greatest divergence is in the understanding of the role and composition of the non-linguistic “strategic” element. In some models this refers solely to the ability to transfer meaning; in others it encompasses all aspects of the translator’s psyche. Some authors see translation competence as a supercompetence that is the sum of all its parts (Wilss, for example), while others see it as a separate competence that complements or ties in all the other subcompetences that are called into play (PACTE).

Table 2.1. Sample of translation competence models

Author <sup>3</sup>	Source component	Target component	Other components	Observations
Koller (1979: 20)	SL language competence	TL language competence, text production skills	Creativity involved in finding and selecting equivalences	
Delisle (1980: 235)	Linguistic competence, comprehension competence	Linguistic competence, re-expression competence	Encyclopedic competence	
Wilss (1982: 58 and 118)	SL text-analytical competence based on comprehensive knowledge of the SL, including the text pragmatic dimension	TL text-reproductive competence based on comprehensive knowledge of the TL, including the text pragmatic dimension	Transfer competence is an “interlingual text synchronization capability”	The SL and TL competences are subordinate and integrated on a higher level (the text) by the transfer competence.
Roberts (1984: 172)	Linguistic competence (ability to understand the SL)	Linguistic competence (quality of expression in TL)	Disciplinary competence (subject knowledge), translational competence (accurate transfer of meaning), methodological competence (effective use of resources to find terminology), technical competence (use of word processors, dictaphones, etc.)	

<sup>3</sup> Page numbers are included as appropriate, when no page number is given, the information has been taken from several parts of the work.

<b>Author<sup>3</sup></b>	<b>Source component</b>	<b>Target component</b>	<b>Other components</b>	<b>Observations</b>
Bell (1991: 35-41)	SL (semantic, syntactic and pragmatic) knowledge and “decoding skills of reading”	TL (semantic, syntactic and pragmatic) knowledge and “encoding skills of writing”	Sociolinguistic competence, discourse competence, strategic competence	Translator competence is a multicomponent “communicative competence” consisting of grammatical competence and the “other” components. Of prime importance is “all-embracing linguistic knowledge”.
Campbell (1991: 339)	Lexical coding of meaning	Global TL competence	Lexical transfer competence and disposition (psychological qualities regarding risk-taking and persistence)	Lexical coding of meaning, global TL competence and lexical transfer together make up proficiency, which is independent of disposition. The three components of proficiency are closely related to each other, but lexical coding may be slightly independent.
Gile (1995: 20)	Passive command of working languages	Active command of active working languages	Knowledge of subject matter, world knowledge	These are components of “translation expertise”.
Kusssmaul (1995: 31-32)			Knowledge and awareness of the creative processes, self-awareness and self-confidence are desirable attributes	Translation competence is not described in terms of component skills and knowledge but as a process: comprehension of the ST, search for translations, deciding which translation to choose.
Cao (1996: 328-330)	SL organizational competence,(grammar and textual competence) and pragmatic competence (illocutionary and sociolinguistic competence)	TL organizational competence,(grammar and textual competence) and pragmatic competence (illocutionary and sociolinguistic competence)	Translational knowledge structures (general and specialized knowledge), translational strategic competence (includes cognitive skills and psychological mechanisms)	SL and TL organization competence together make translational language competence.
Hurtado Albir (1996: 34)	Communicative competence in SL. Text comprehension	Communicative competence in TL. Text production	Extra-linguistic competence (cultural and subject knowledge and research skills) and professional	Translator competence is beyond linguistic competence.

Author <sup>3</sup>	Source component	Target component	Other components	Observations
	competence (capacity to analyze, synthesize, deduce)	competence (clarity and wealth of expression, creativity)	competence (mastery of tools, translation market etc.)	“Translatorial” competence is the predisposition to switch from one language to another without interference.
Hatim and Mason (1997: 205)	ST processing competence	TT processing competence	Transfer competence	
Delisle, Lee-Jahnke and Cormier (1999: 152)	Comprehension of the “sense” and interpretation of the ST	Reformulation of the sense of the ST to produce the written TT in accordance with writing conventions	Ability to disassociate the languages, apply translation procedures and maintain textual coherence	“Translation competence” is not mentioned in this dictionary of translation terminology. The components here are listed under “linguistic competence”.
Adab (2000)	ST comprehension, comprehension of SL message	TL text production, revision of TT in relation to translation brief	Identification of translation problems, comparative and contrastive selection of solutions, comparative and contrastive language knowledge and cultural knowledge, research competence, subject knowledge, transfer strategies	The components are areas of sub-competence that contribute to the overall translation competence.
Beeby (2000)	Advanced reading skills in SL	Reformulation and composition skills in TL	De-verbalization skills, awareness of the multiple contexts involved in translation, awareness of the interdependence between micro and macro structures in the text and translation, contrastive linguistic competence, contrastive discourse competence; extra-linguistic competence	The focus is on the contrastive linguistic and discourse competences.
Kiraly (2000: 10-14)	Foreign language competence	Native tongue competence, ability to localize as well as translate	Ability to master new technologies, work collaboratively, research topics using electronic tools, acquire adequate knowledge in new areas, justify one’s work, negotiate and collaborate with other translators and subject matter experts	Also presented is an integrated model comprising: awareness of situational factors: translation-relevant knowledge (linguistic, cultural and specialized knowledge); translation-relevant skills (ability to produce adequate translations) (1995: 108).
Neubert (2000: 3-18)	SL language competence	TL language competence, textual	Subject competence, cultural competence	Translation competence is more than transfer



Author <sup>3</sup>	Source component	Target component	Other components	Observations
		competence (how to organize the content)		competence, but “the overarching transfer faculties and skills” dominate the rest.
Orozco (2000)	Comprehension competence, communicative competence	Re-expression competence, communicative competence	The ability to deverbilise and maintain the SL and TL in separate compartments, competence when carrying out the translation project. Extra-linguistic competence, professional-instrumental competence and psycho-physiological competence all feed into transfer competence. Strategic competence, which interacts with all the other competences.	“Transfer competence is the central competence of translation competence and integrates all the others” (p. 199).
PACTE (2000)	Communicative subcompetence		Extra-linguistic subcompetence, professional-instrumental subcompetence, transfer subcompetence, strategic subcompetence and psycho-physiological subcompetence	See also PACTE (2011).
Presas (2000:28)	Knowledge of SL	Knowledge of TL	Knowledge of the real world and the material, the ability to use tools such as dictionaries and other sources of documentation, cognitive qualities, such as creativity and attention, or the capacity to resolve specific problems	Competence refers to the “system of underlying kinds of knowledge, whether declarative or operative, which are needed for translation” (p. 28).
Robinson (2003)	Linguistic intelligence: ability to understand and manipulate languages		Interpersonal skills (“dealing with clients, agencies, employers; networking”) research, use of technology (p. 160), intercultural competence and awareness (p. 187), “the ability to analyze a source text linguistically, culturally, even philosophically or politically is of paramount importance to the translator” (p. 208).	The need for intelligence and analytical skills is stressed. The intelligence of translators and interpreters includes the “ability to learn foreign languages and to hear, sort out, produce, and manipulate the complexities of transfer among them” (p. 56)
Kelly (2005: 32-33)	Communicative and textual competence in SL	Communicative and textual competence in TL	Cultural and intercultural competence, subject area competence, professional and instrumental competence, attitudinal competence, interpersonal competence, strategic	

Author <sup>3</sup>	Source component	Target component	Other components	Observations
			competence	
Nord (2005)	Linguistic competence in L2, text reception and analysis competence	Linguistic competence in L1, text production competence	SC and TC knowledge, technical competence (for research) factual competence (specialized knowledge), competence of translation quality assessment; transfer competence, which comprises the skills of text reception, text production, and the use of translation tools, as well as the ability to 'synchronize' ST reception and TT production.	SL and TL competence are seen as part of being (ideally) bi-cultural, which means having a perfect command of both the source and the target culture (including language) (p. 12).
PACTE (2011)	Bilingual subcompetence (pragmatic, socio-linguistic, textual, grammatical and lexical knowledge in the SL and TL)		Extra-linguistic subcompetence, knowledge about translation subcompetence, instrumental subcompetence (use of document resources and ICT), psycho-physiological subcompetence, and a strategic subcompetence	Translation competence consists of expert, predominantly procedural (i.e. non-declarative) knowledge and inter-related subcompetences of which the strategic component is of particular importance.

All the models basically refer to a source-comprehension component, a target-production component, a thematic knowledge component, and a strategic component of some kind. The literature review thus provides a basic framework for starting to identify the skills and knowledge required in IGO translation. Moreover, regardless of the different underlying approaches, labels and classification systems employed, there seems to be agreement on the multi-componential nature of translation competence.

Three other points relevant to this research come to light in the review of the literature on translation competence: (1) the lack of empirical evidence to substantiate relations between the components of the competence models, as noted by Waddington (1999: 135); (2) the fact that most of the discussion of the subject takes place within the context of translator training at academic institutions; and (3) the absence of attempts to rank the components of translation competence.

### 2.1.1 *The lack of empirical evidence*

The lack of empirical data to explain relations between the various components of the translation competence models is bemoaned by the PACTE group, which is working on validating its own “holistic” model through empirical research (2008). Campbell has also conducted experiments to test the relations between the components of his own model (1991). Intriguing insights emerge from Waddington’s empirical study, which suggests that translation competence correlates closely with native language competence and self-appraisal, but not with mathematical intelligence (2001a). The other models seem to fall squarely into the speculative branch of Translation Studies for the moment. Criticism of the lack of empirical evidence underpinning them is not necessarily warranted, as Kelly points out:

The criticism is undoubtedly pertinent if the authors using or proposing the catalogue are claiming to describe the actual cognitive process of translating itself. If, however, they intend to provide a list of skills which a training course should provide, then the criticism would seem to be less well founded. (2005: 31-32)

Most of the speculative models, including Kelly’s own (pp. 32-40), derive from the observations of the theorists in the field, and thus necessarily have experiential input. As such, they also served as a point of departure for the research. The lack of empirical data did mean, however, that there would be little data to use as a point of comparison for any data generated by our research.

### 2.1.2 *The translator-training focus of research*

Both Way’s and Kelly’s comments above point to the underlying driver of the discussion of translation competence in Translation Studies: the need to design or justify the content of translator training programs. As Pym notes, “[i]n most cases, the complex models of competence coincide more or less with the things taught in the institutions where the theorists work” (2003: 487). This poses a possible problem in our case because the concerns and priorities of academic institutions are not necessarily the same as those of IGO recruitment officers. The literature on translation competence therefore needs to be complemented and contextualized with a review of the literature on IGO translation.

### 2.1.3 *The absence of attempts to rank the components of translation competence*

It is rather striking that none of the scholars mentioned seems to have tried to determine which components of translation competence might be more important than others. To build on Pym's metaphor (2003), they are willing to say what should go into the soup, but not in what measures. Of course nobody underestimates the complexity of translation work, which often involves the simultaneous activation of multiple interacting skills and knowledge, the combination of which can vary considerably from one task and even one moment to the next. Some skills and knowledge types must be called into play more often than others, however, and some must in turn be harder to find and some must have a greater impact on the acceptability of a translation produced in a given context.

The purpose for which scholars have drawn up their lists of skills probably explains why they have not attempted to weight them. Not only is it difficult, it might even also be inappropriate. Their aim has been to provide a theoretical framework, leaving curriculum developers free to determine the balance of contents among and within each subcomponent according to the identified needs of students. For syllabus planning or justification purposes, a catalogue of skills and knowledge types loosely grouped under various headings serves as a frame of reference. For IGO recruitment purposes, however, what is required is a hierarchy in which skills and knowledge types are at least broadly classified as essential, preferable or optional on the basis of a sound analysis of the organization's needs.

## 2.2 IGO translation

Considering the enormous volume of translation work that IGOs handle and the number of translators they employ, it is rather surprising that so few studies have been performed on IGO translation. This could be the result of lack of interest in the subject among translation scholars, whose reputation for ivory-tower attitudes is such that whole books have been published in attempts to address the problem.<sup>4</sup> It could also be the result of reluctance on the

---

<sup>4</sup> Consider, for example, Chesterman and Wagner's *Can Theory Help Translators? A dialogue between the ivory tower and the workplace* (2001) and Baer and Koby's *Beyond the Ivory Tower. Rethinking translation pedagogy* (2003).

part of IGO translation services to encourage enquiry into their workings. This might in turn stem from the confidential nature of some of the documents IGOs handle, time limitations and other practical constraints, the obligation of staff to protect the image of the IGO, a fear of tarnishing a reputation for high standards of excellence, or a combination of all these and other, unknown factors. In any case, the literature on IGO translation seems to have been produced mainly by insiders such as Wagner, Bech and Martinez (2002) from the European Union translation services, by people who managed to gain some access to IGOs, such as Koskinen (2000 and 2008), and by those working, if not with IGOs, then with large institutions: Mossop from the Canadian Government's Translation Bureau (1988) and Williams (1989), who has worked with the same Bureau as a consultant. To supplement the scholarly literature, I have also examined publications and statements issued by IGOs (EMT 2009, Hindle 1965/84, Sekel 2008 and Verleysen 2010).

The literature on IGO translation indicates that the most salient features of IGO translation work, as far as this study is concerned, are as follows:

### *2.2.1 The institutional nature of IGO translation work*

Translations are essential to IGOs, since it is through them that they communicate within themselves and with the world. IGOs are in fact both “translating institutions”, inasmuch as they produce textual translations, and “translated institutions” because their ideology is enshrined, captured and conveyed through their translations (Koskinen 2008). The goal of the institution determines its approach to translation (Mossop 1988: 65). The formality and import of diplomatic communication mean that IGO translation is a highly norm-governed activity, constrained by style manuals and strict usage rules. Every effort is made to preserve a climate of mutual respect, and careful attention is paid to tone and register and the use of politically neutral or sensitive terms in the target text. This includes opting for terms acceptable to most speakers of a language rather than regional preferences. As the former Chief of the United Nations English Translation Service points out, IGO translators need to be aware that:

Words have a significance beyond their obvious meanings [...] there is always a political dimension in the work of United Nations translators and interpreters. They must always be aware of, and sensitive to, the positions of the parties to a negotiation,

and must ensure that, to the extent possible, all translated versions of a text fully and accurately reflect the meaning and intent of the author or authors. (Sekel 2008: 1)

Ivanova similarly notes that “[k]nowledge of culture-specific and sensitive aspects or political concepts, of associated values and attitudes, as well as knowledge of the source and target culture are decisive elements of translation competence for political communication” (cited in Schäffner 2007: 142).

### 2.2.2 *The emphasis on quality*

Given the impact and permanence of official communications and the potential seriousness of errors, the highest standards of accuracy and style are pursued. Williams points out how errors can undermine national, organizational or personal safety and security (1989: 19-23). Misunderstandings can derail peace talks and delay projects, and poorly written documents can slow down negotiations and ultimately cost not just money but sometimes even lives. Hence the upstream and downstream control measures in place to optimize translation quality (editors, revisers, terminologists, proof readers, reference assistants, etc.) and the efforts to recruit highly competent translators (DGACM 2011a, DGT 2008).

### 2.2.3 *The need for standardized language use*

Since translations become official records of the organizations’ activities and communications, standardization and consistency are essential. “It is quite common in many institutional settings to have ‘document chains’ in the sense that each document is anticipated for (or regulated) in previous documents, and it in turn paves the way for new documents taking the issue further” (Koskinen 2008: 125). This means that each translation sets a precedent for future language use, and translators often have to glean terminology and phrasing from earlier documents on the particular topic on which they are working. Glossaries are created not to help new recruits as much as to ensure that all translators use the same terms. The requirement to adhere to in-house usage can interfere at times, however, with the readability of texts, especially for those outside the organization.

#### 2.2.4 *The influence of language policies*

Multilingual policies mean that translations are in some instances viewed as different language “versions” (Koskinen 2000: 54), upholding the principles of equal and multiple authenticity (Wagner et al. 2002: 7-8). This calls for a communicative, message-oriented approach to translation in which not only the accuracy of the translation but the idiomaticity of the target text are prime considerations to guard against accusations of linguistic imperialism (Schäffner 2007: 141). This idiomaticity is occasionally undermined, however, by the in-house usage rules referred to above, and is tempered by diplomatic considerations. Although the need for diplomacy encourages the domestication of translations since the use of foreign words in a translation could also smack of cultural imperialism, this does not extend to the use of foreign words that are commonly used in the target language. For example, Spanish-speaking economists may talk happily about “los hedge funds” but in IGO documents they will, ironically, probably be referred to as *fondos de cobertura* [funds of coverage], possibly with a footnote explaining that they are “lo que los angloparlantes llaman ‘hedge funds’” [what English-speakers call ‘hedge funds’].

#### 2.2.5 *The range of subjects and document-types covered*

Translators at IGOs can end up translating a wide range of text types on a wide range of subjects. At the European Union institutions, for example, apart from the obvious reports, translators work on policy documents, legal texts and resolutions, speeches, debates, minutes, press releases, petitions from the public, conference proceedings and material for websites (Wagner et al. 2002: 55). Although some translators specialize in certain topics, many have to handle documents on all kinds of subjects, making it essential to remain abreast of developments in many fields and to be able to master new subjects quickly (Sekel 2008).

#### 2.2.6 *The need for clarity*

As the product of a highly regulated and norm-governed activity with its style guidelines and in-house terminology databases, IGO translations (like IGO source texts) are at risk of becoming bogged down in in-house terminology, acronyms, abbreviations, excessive formality and political correctness to the point of being understandable only to those within the organization (Koskinen 2000: 53). The danger is being addressed with active campaigns

to promote clear and concise writing (Wagner et al. 2002: 73, Hindle 1965/84). The guidelines for the submission of documents issued by some organizations often include length limits and reminders of the importance of relevance.<sup>5</sup> Of course IGOs cannot tell member States or even their own authors not to obfuscate the message, when that may be the authors' very aim, and opacity may be a chosen feature. As Sekel notes, "Member States have ample scope to exploit the advantages of deliberate ambiguity or obscurity in the language they use. The often murky texts that emerge from the rough and tumble of multilateral negotiations pose difficult challenges" (2008: 1).

### 2.2.7 *The poor quality of some source texts*

Source texts at IGOs may also simply be poorly written, which poses additional challenges to the translator. This occurs for a number of reasons, not only haste, but the fact that texts are often drafted by non-native speakers, multiple authors and rewritten several times (Wagner et al. 2002: 70). Translators are therefore expected to detect inconsistencies and at times improve on the content as well as the style of the original. Mossop observes that "experienced translators tend to produce translations whose writing quality is much superior to that of the source text. To accomplish this they engage in what might be called mental stylistic editing and structure/ content editing *while they translate*" (2001: 33) (italics in the original). This means that, as Sekel notes (2008), translators at IGOs need to have the judgment to determine whether a lack of clarity or awkward syntax, for example, is deliberate and politically motivated or a stylistic slip.

### 2.2.8 *The varied use of technology*

In addition to editors, reference assistants, revisers and proofreaders, IGO translators usually have a range of technological support tools available to them (Verleysen 2010). At some organizations, you can still find translators who record their translations on dictaphones for later transcription by the text-processing unit, and revisers who mark changes with a red pencil on the hard copy. But at most IGOs, translations are produced on computers using word processing programs or speech-recognition software, and revisers work on-screen with

---

<sup>5</sup> See, for example, *Guidelines on the Form and Content of Initial Reports under Article 19 to Be Submitted by States Parties to the Convention against Torture* available at: <http://daccess-dds-ny.un.org/doc/UNDOC/GEN/G05/428/37/PDF/G0542837.pdf>.



custom-made tools. All organizations have converted their glossaries and style manuals to electronic format and most are either using or experimenting with translation memory software. Many have also developed their own online reference libraries or electronic referencing tools and are exploring the possible advantages of incorporating machine translation into the workflow. New recruits must be willing to master the technology used by the service in question (DGACM 2011a).

### 2.2.9 *Turnaround times*

Tight deadlines are common owing to the nature of some international communications and are often considered to be “unreasonable”, given the quality demands placed on translators (Wagner et al. 2002: 80; Cao and Zhao 2008: 41). It is therefore in the interests of IGOs to find translators who can work fast, not only to improve their productivity but to ensure that quality does not suffer too much on rush jobs.

### 2.2.10 *Summary of the review of the literature on IGO translation*

The literature on IGO translation provides a fairly clear picture of its main features, but is rather scarce and based on expert opinion (e.g. Mossop, Sekel and Wagner et. al) or controlled observations (Koskinen) rather than on empirical research. It does enable us, however, to appreciate IGO translation in the light of different theories of translation.

IGO translation, as a highly regulated activity with its reference assistants and revisers, its style guidelines and pursuit of political correctness, would possibly be a rich field for identifying norms, as suggested by Toury in his *Descriptive Translation Studies and Beyond* (1995: 55). His proposed law of standardization, according to which “textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of habitual options offered by a target repertoire” (p. 268), probably applies in most IGO translation. Chesterman’s expectancy norms (1997: 64) meanwhile could account for the obsession with detailed levels of accuracy and the careful checking of translations at IGOs. The reputation of an organization is so tied up in its communications with the outside world that high quality standards are maintained and consequently expected. His professional norms are also applicable, inasmuch as IGO translators work to ensure maximum communication (p. 69), and to some extent are accountable for their work. Usually a translation is seen as the product of the translation service or department, however: rarely

will individual translators be held responsible, and certainly not to the commissioner or reader. Translation decisions are also rarely taken, as suggested by Chesterman's "relation" norm, "according to text-type, the wishes of the commissioner, the intentions of the original author, and the assumed needs of the prospective reader" (1997: 69), since the translator may only actually be able to know the first, be unable to ascertain the third by any means other than an analysis of the source text itself and have no idea about the fourth. As far as text type is concerned, most documents are what Reiss refers to as "informative" texts that call for a content-focused approach, although translations of "operative" texts, such as speeches, also encourage other factors to be taken into consideration (1977/89: 113-114).

IGO translators take great care to reproduce the content of the source text as exactly as possible, and not just for political reasons. On the one hand texts often form part of document chains, as noted above, which makes consistency a guiding principle. On the other, organizations that work with several languages usually relay translations through just one or two of them, and care has to be taken not to shift the emphasis of the message or the results could be the same as in a game of Chinese whispers. IGO translation therefore largely subscribes to the theory of natural equivalence, as described by Pym (2010: 7-9), with its lack of directionality. The notion of "equal value" also sits well with the multilingualism policies of IGOs. The pursuit of the "closest natural equivalent to the source-language message", as proposed by Nida (1964: 166) is in fact carried out at the micro-linguistic level.

Although highly focused on reproducing the content of the source text, the purpose of the target text is an overriding concern, as advocated by the proponents of *Skopos* theory. In fact, Vermeer's idea that the nature of the target text is "primarily determined by its *Skopos* or commission" (Vermeer 1989/2000: 230) is applied in IGO translation. It also explains the only context in which IGO translators might diverge from the principle of rigorously reproducing the content and style of the source text: the need to protect the organization's image. In that regard, the coherence "rules" put forward by Reiss and Vermeer, as reported by Munday (2001: 79-80), are applicable and in some instances translation policy decisions (e.g. usage rules) determine linguistic choices, overriding the constraints of the source text. The plurality of purposes that *Skopos* theory presupposes rarely arises, however. IGO translations have two functions: (1) to perform the same function as the source text, and (2) to make the translation service (and hence the organization) look good. The former calls for the pursuit of the closest natural equivalent, the latter marks the boundaries for any divergence from that norm. In short, the norms and *Skopoi* of IGO translation coincide largely with traditional equivalence theory.

The review of the literature on IGO translation has served as a starting point for identifying the particular skills and knowledge required. It suggests that translators at IGOs need to be able to handle a wide range of document types, adhere to in-house style guides, work with reference documents, and handle electronic terminology tools and possibly translation memory software. They also need to be able to produce translations that fully and accurately reflect the meaning and intent of the author or authors, even if this means working from poorly written source texts and meeting a tight deadline. They must achieve a style that upholds the reputation of the organization and achieves its communicative aims, while writing clearly and idiomatically and in full awareness of any political undercurrents that may be at work. They also need general knowledge and, possibly, specialized technical knowledge as well, or at least the ability to master new subjects quickly.

### 2.3 Translation testing

If the literature on IGO translation is scarce, the literature on proficiency testing of translators is even more so. Scholars tend to focus on formative and summative assessment in the context of translator training at universities. Moreover, some, such as Nord (2005: 179), Kiraly (2000: 158-162) and Kelly (2005: 133-5), stress the importance of using alternative forms of assessment to testing. The pedagogical approach is not particularly helpful, however, from the recruitment perspective. Recruiters are not trying to *develop* candidates' translation competence or to see whether translators can apply what they have been taught in a particular program; they want a quick way to check whether a candidate has the skill-set required now; they want to *measure* translation competence. This is proficiency testing for selection purposes.

Language testing is the closest field to Translation Studies in which extensive research has been conducted into proficiency testing. Linguistic competence, as well as being a component of translation competence, poses the same main challenges as far as testing is concerned: it is multi-componential and impossible to measure directly, and tests may assess several distinct but interconnected abilities at once. The theory and practice of language testing therefore can be, and have been, applied to translation. Bachman's proposals for maximizing the validity and reliability of language testing (1990), with their thorough grounding in measurement theory, have been used for the design of a proposed rubric for the certification of translators by the ATA (American Translators Association) (Angelelli 2009),

and his model of communicative language ability underpins the translation competence model of Cao (1996) and Hatim and Mason (1997). In his *Fundamentals of Language Testing* (1990), Bachman highlights the limitations of language testing as: the problems of specification (owing to the interrelated abilities), indirectness (what is measured are manifestations of underlying skills), incompleteness (only a snapshot of performance is obtained), imprecision (of the scales of measurement and the quantification of assessment), subjectivity (of test design and scoring) and relativity (the problems of setting standards) (1990: 30-40). All of these are equally applicable to translation testing. He also explains four basic steps for addressing these limitations: (1) define the constructs theoretically,<sup>6</sup> (2) define them operationally, (3) define how to quantify the results, and (4) take measures to maximize the reliability and validity throughout (Bachman 1990; Bachman and Palmer 1996), which all constitute yardsticks for assessing testing practice.

The works of Bachman and other scholars from the field of language testing and of scholars from Translation Studies *per se* are examined in the review of literature on testing, which is divided, for convenience, into three subjects: the principles of test design, task difficulty and scoring.

### 2.3.1 *The principles of test design*

According to Alderson et al., the framework for test construction should be derived from a needs analysis based on the study of the domain and “the performance which the test is intended to predict” (1995: 22). This ties in with Bachman’s statement that the “value of tests lies in their capability for eliciting the specific kinds of behavior that the test user can interpret as evidence of the attributes or abilities of interest” (1990: 22). The distinguishing tasks in the target domain need to be identified with two objectives in mind: (1) to identify the skills and knowledge called into play, and (2) to identify the context in which they are used. In this, translation competence assessment may have a slight advantage over language testing, inasmuch as the type of performance we wish to generalize to is quite specific, and in IGO translation even more so. This makes identifying typical domain tasks relatively straightforward.

---

<sup>6</sup> Constructs are putative abilities, what Cushing Weigle refers to as “abstractions we define for assessment purposes” (2002: 41)

Once the construct to be measured has been defined, the next step, according to the literature, is to design the test task(s) that will produce a ratable sample of performance. Of course in text-based translation testing, what is rated is the evidence of the construct (translation competence) having been applied, not the construct itself. The question in the assessment of translation competence, as Campbell puts it in “Towards a Model of Translation Competence”, surrounds the “decision as to whether *processes* can be effectively assessed, or whether the assessor was happy to measure only the *product*. An assessment procedure to measure the former would have to be far more sophisticated than one that merely marked microlinguistic features” (1991: 339). The issue is not so much the validity of assessing translators on the basis of their product; after all, that is how people in skills-based professions, from bakers to surgeons, are evaluated: we are not necessarily interested, beyond basic ethical and safety concerns, in how they achieve results, as long as they can do so and show signs of being able to do so for as long as we need. The issue is *what product* we use as the basis for assessment. Campbell goes on to point out that “[a] test that aimed simply to assess translation competence once at a given moment could honestly operate by means of microlinguistic marking, as long as it were based on norms of item difficulty and did not hope to randomly assess competence with a text chosen with not much more than an arbitrary judgement as to its overall difficulty” (1991: 339). In recruitment testing, competence “at a given moment” (the time of the test) is the indicator of readiness, and it would seem that Campbell thus agrees with Hatim and Mason that “an unseen written translation test of a particular level of difficulty might serve as a (kind of) proficiency test” (Hatim and Mason 1997: 200). This suggests that text-based testing of translators is valid in recruitment, provided that the translation task is pitched at the right level of difficulty. Before worrying about the level of difficulty, what is needed, as Bachman points out, is to make sure the test will generate a ratable sample, in other words, sufficient evidence of competence, and according to him and also to his colleague Palmer, that means test tasks must be “authentic” and “interactive” (1996: 23-30).

### 2.3.1.1 Authenticity and interactiveness

Authenticity in testing, which is understood to refer to having test-takers perform domain tasks, might be thought to be particularly important in recruitment testing, which rests on the premise that performance on the test is indicative of future performance on the job. It would seem logical to select candidates on the basis of their performance of a task that forms part of the job description, to the extent that can be arranged, rather than on the basis of one that does

not. This logic appears to underpin Kiraly's rhetorical question: "Can the translation of a single text without advance preparation, without computer-based tools, without access to reference works or the chance to negotiate with a client really tell us anything credible at all about a translator's competence?" (2000: 156). He is commenting on translation testing in a different context, as the reference to "clients" shows, but his remark could equally be leveled at some IGO recruitment tests. Kelly, for example, criticizes traditional examinations, referring specifically to those used "for recruiting professionals at international organizations", from various perspectives, including because "[t]hey are unrealistic, that is they have little or nothing to do with real professional translation situations (prior knowledge of type of text, subject area, possibility for documentary research of all kinds, kind of time pressure, physical environment...)" (2005: 131-132).

Authenticity should not be an end in itself, however, but a means of producing a ratable performance. As Fulcher and Davidson, again from the world of language testing, point out:

Authenticity, even conceived of as matching test method facets to facets of similar tasks in the real world (Bachman and Palmer, 1996), does not make test tasks automatically valid through directness; it means only that we may be able to model test-taker behavior in ways that allow us to observe processes that would be used in real-world language use. (Fulcher and Davidson 2007: 63)

It is the ability to obtain manifestations of processes or skills in action that Bachman is after in his promotion of a trait-based approach and "interactiveness" (Bachman 1990: 329). "Interactiveness" in language testing, according to Bachman and Palmer, refers to "the extent and type of involvement of the test-taker's individual characteristics in accomplishing a test task" (1996: 25). The idea is to mirror reality based on the identification of the critical features of what we wish to test, in order to elicit the behavior sought. In other words, tests should be designed to measure specific abilities within a realistic context. This means making tasks "interactive" enough, i.e. ensuring they engage the skills and knowledge we wish to measure. A "real-life" task will not necessarily be interactive enough; it might not produce the range or sufficient examples of the behaviors sought for valid inferences to be made about abilities. Authenticity, for Bachman, thus refers to being authentic in type. An authentic test task, in this case, can be a modified real domain task or a wholly artificial imitation. The face validity of a genuine and a fake (authentic-like) task is the same, provided test-takers are not

aware of the difference. The important point is the affinity between the test task and the domain task on which it is based.

The discussion of authenticity is, of course, associated with validity claims. Chappelle's papers on validity theory (1998-1999, as reported in Fulcher and Davidson 2007: 16-17) point to the two extremes of the validity cline: trait theory and the new behaviorism, the former being reminiscent of Bachman's "interaction/ability" approach as it focuses on individual traits, and the latter to the real-life approach. The approach adopted affects the understanding of score meaning and generalizability. In the former, we generalize to the competence of the test-taker, in the latter to the context. Ideally, the two poles would be brought together in a trait-based authentic test with a pragmatic approach:

Chapelle (1998: 34, 44) describes an interactionist understanding of score meaning as 'the result of traits, contextual features and their interaction' and says that 'performance is viewed as a sign of underlying traits, and is influenced by the context in which it occurs, and is therefore a sample of performance in similar contexts'. In this approach we acknowledge that the test contains only a sample of the situation or situations to which we wish to generalize. Part of investigating the validity of score meaning is therefore collecting evidence to show that the sample is domain-relevant and predictive of the wider range of abilities or performances that we wish to say something about. (Fulcher and Davidson 2007: 17)

This brings us back to domain-relevant (authentic-like) evidence-based design, but also highlights the importance of trying to validate the inferences drawn from test scores by some other means, i.e. construct validation. Cushing Weigle applies Bachman's principles in her work on the assessment of writing:

In testing writing, construct validity must be demonstrated in at least three ways: (1) the task must elicit the type of writing that we want to test [...]; (2) the scoring criteria must take into account those components of writing that are included in the definition of the construct; and (3) the readers must actually adhere to those criteria when scoring writing samples. (Cushing Weigle 2002: 51)

The same principles can be applied to assessing translation. Setting aside scoring considerations for the moment, as far as test tasks are concerned, authenticity (affinity with

domain tasks) is required to elicit the right type of sample, and interactiveness (or its equivalent for engaging the components of translation competence) is required to elicit a sample that is sufficient. Authenticity and interactiveness alone are not enough, however: they might ensure the test task produces domain-relevant and construct-relevant evidence, but that evidence must also make it possible to discriminate between test-takers and identify the most proficient among them. That means, as mentioned earlier, that the test task must be pitched at the right level of difficulty.

### 2.3.2 *Task difficulty*

Several aspects of a translation task may determine the difficulty it poses: e.g. time allowed for completion, access to resources, the equipment made available for producing the translation, and the physical environment. All these elements are relatively easy to control and hence to standardize, and Howard points out how testing systems can be conducive to failure if they do not take these factors into account (2011: 21-22). The key input of any translation test, however, is the source text, and there is surprisingly little literature on the texts that should be used in translation tests and how they should be obtained. Recommendations on selecting texts that pose the right type of linguistic and transferability challenges and are suitably inoffensive can be found in the works of scholars who focus on translator training (e.g. Hurtado 1995: 60; Kussmaul 1995: 51 and Nord 2005: 147), and there are tips on the wealth of suitable challenges to be found in certain text types (e.g. Kelly 2000: 158-166), but they all refer to formative or summative assessment, not recruitment testing, where fairness has a very different meaning. Fairness in recruitment refers only to test conditions, not to the content of a particular training program. It may be unfair to include something in a university exam that has not been covered during the course or is beyond the capabilities of most test-takers, but that is not a concern in recruitment testing.

Also the focus of the literature is invariably on text-selection, as if texts have to be found and cannot be produced specifically for testing purposes. Kelly reports that “[c]onsensus in the field is that texts should, as far as possible, be authentic, unmanipulated and presented in their original form” (2005: 119). Does this apply only to classroom activities or to testing as well? Test tasks need to be domain-relevant, not necessarily taken from the domain. The exclusive use of genuinely “authentic” texts (ones not created for translation testing purposes) would seem to undermine the validity of testing; it suggests the “arbitrary” selection referred to by Kiraly and Campbell. Unless only one highly particular type of text is



translated, of course, then any example of that text type will be typical. In most domains, however, how likely is it that a source text has already been written that contains all the features required to produce the evidence needed to make the most valid inferences possible about abilities? Nord, who advocates the use of what she refers to as “real texts-in-situation”, comments that it is “often difficult to find authentic texts which are not too long and have exactly the degree of difficulty appropriate for the students” (2005: 163). She does not offer a solution, however. Maybe it does not matter so much in the context of translator training. Teachers can possibly get around the problem by grading on a curve, i.e. interpreting scores solely with reference to the performance of the group, as in norm-referenced testing (Bachman 1990: 72-73). This cannot always be done in IGO recruitment testing, however. Recruitment testing is (or probably should be) criterion-referenced, in other words, candidates must meet a certain minimum standard. We would not hire surgeons to operate on our children only because they were the best in the bunch unless they also met particular criteria; the same should apply to any profession in which poor-quality work can have serious repercussions. In other words, test tasks must be pitched correctly not only to ensure general authenticity and discriminate between candidates, but to ensure that the selected candidates are good enough on a series of key points.

In criterion-referenced tests, the level of difficulty is vitally important. As Bachman notes, “[t]he primary concerns in developing a CR [criterion-referenced] test are that it adequately represent the criterion ability level or sample the context domain, and that it be sensitive to levels of ability or degrees of mastery of the different components of that domain” (1990: 74). Criterion-referenced tests must also be adequate in scope: tests must elicit ratable samples that provide an opportunity for the full range of components to be rated to occur, and must enable test-takers to perform at their highest level of ability (Bachman and Palmer 1996: 218-9). In recruitment testing, this means that tests must cover all the components of the skills-knowledge set required, and the level of difficulty, as determined by the needs of the job, must be such as to allow not only that the adequate candidates may stand out, but to ensure that the best candidates do stand out. In this sense, only texts based on some notion of item difficulty, as Campbell said, and that produce sufficient evidence of the candidates’ handling of those items, can be valid. And then, only if those items are standardized, can the test become a reliable measure across time.

Nord sees difficulty being determined not only by the source text but by the translator’s competence. She therefore refers to translation *problems*, which she defines as a “transfer task which every translator (irrespective of their level of competence and of the technical

working conditions) has to solve during a particular translation process” (2005: 166-167). These can be pragmatic, convention-related, linguistic or text-specific. Nord’s focus is mainly on translator training, but she comments specifically on the selection of texts for assessing professionals, saying that that beyond stipulating whether a source text should be “general” or “specialized”, it makes no sense to establish other criteria for examinations when assessing professionals because “a professional translator should be in a position to cope with any conceivable translation problem” (2005: 178). This may apply to proficiency testing for accreditation purposes, but not to IGO recruitment, and it fails to appreciate that within the world of professional translation there are those who cope far better than others.

Awareness of what makes a source text difficult is important for optimizing the validity of a test. As Cushing Weigle notes, “construct validity is enhanced when the factors that contribute to difficulty are those that are included in the definition of the construct” (2002: 50). The difficulty drivers should be items that are expected to activate the components of translation competence that the test is intended to measure. And sufficient difficult items must be included to obtain a ratable sample. This brings us back to the “interactiveness” of the test task: the higher the density of difficult items, the greater the “interactiveness” of the source text.

Applying what all the aforementioned scholars from the field of language testing say about comprehension tasks, it is clear that some aspects of translation-task difficulty can be controlled relatively easily: the amount of technical knowledge required to understand the source text, for example, and the time allowed for completion. Just like reading-comprehension texts in language testing, source texts can be on topics expected to be wholly familiar or wholly unfamiliar to all candidates, if the goal is to level the playing field. Alternatively, they can be on a particular subject when knowledge of that subject is an essential component of the skills-knowledge set. How much that knowledge or lack of it might interfere with the other components is difficult to determine, however, and it may in fact be preferable to test subject knowledge separately.

Clear instructions can also be given to optimize the chances of obtaining samples of top performance:

It is clearly important that candidates are given adequate information to enable them to know exactly what the test will look like [...]. The intention of such specifications for candidates should be to ensure that as far as possible, and as far as is consistent with

test security, candidates are given enough information to enable them to perform to the best of their ability. (Alderson et al. 1995: 21)

Nord, as an advocate of a *Skopos*-based approach, notes that instructions can be used to adjust the level of difficulty, especially as they have a direct bearing on awareness of the purpose of the translation. Cushing Weigle meanwhile points out that “personal characteristics and affect are usually explicitly excluded from the construct, and we try to avoid inadvertently measuring those factors in a language test” (2002: 46). This means not only avoiding topics that give an unfair advantage to a certain group of test-takers, but avoiding ones that are potentially upsetting or controversial because “affect can have a facilitating or a debilitating effect” (2002: 46).

Determining the difficulty of a source text in terms of its linguistic and rhetorical features is far harder. Stuart Campbell and Sandra Hale have written several articles on their efforts to determine what makes a source text difficult. To them, “[t]ext difficulty for the purpose of translation, then, is seen as a function of the cognitive efforts required to process the item in question and convert it into the target language” (Hale and Campbell 2002: 15). They report that Reiss (1982: 12) says that difficulty depends on the subject (the semantic aspect), register (the material aspect), the type of language (functional aspect), pragmatics, and the historical-cultural context, and identifies four levels of difficulty within each aspect, but still does not specify the text features or characteristics of the corresponding source texts, and the determination of levels is not empirically based.

Campbell and Hale’s research has shown that certain item types in English pose more difficulties than others (abstractions, metaphors, official names, complex noun phrases, for example) but for different reasons (Campbell and Hale 1999), and that a text is difficult to translate when it includes a high number of incidents in which the translator faces “choosing a semantic interpretation in the face of ambiguous syntax” (Hale and Campbell 2002: 29). They admit, however, that much research still needs to be done into the notion of accuracy and the identification of the features that affect the difficulty of source texts: “To this day there is not a single acknowledged procedure for grading texts used for translation teaching, or for measuring the reliability of translation examinations” (2002: 30).

Managing task difficulty is therefore not just a validity concern; reliability cannot be measured unless task (and hence source-text) difficulty can be determined and controlled. In translation testing, measuring reliability using traditional parallel test methods would require having equally difficult and independent source texts: “this would involve finding or

comparing two examination texts of exactly the same degree of complexity in lexis, grammar, content, style and rhetorical structure” (Campbell and Hale 2003: 219). Note that Campbell and Hale here refer to finding source texts, although in their paper “What makes a text difficult to translate?” (1999) they speak of creating materials as well. If difficulty drivers can be identified, then it should be possible to construct source texts around them. This would mean not only identifying the items that cause the most difficulty, but measuring their relative discriminatory power as well. The problem, of course, is that it is impossible to pinpoint discrete, equal items for applying split-half methods of statistical measures of reliability. The complex interrelation of skills and knowledge in translation are not to be underestimated. But it should be possible to come up with task specifications based on the number and type of difficulty drivers that at least maximize the comparability of examinations and hence make it possible to gauge, albeit approximately, their reliability. Nord seems to be the only one attempting to do this. She proposes identifying “translation problems” (defined above) in a text and rating translators on the percentage they solve adequately (Nord 2005: 188-189). She does not explain how to standardize the type of problem, however. Nevertheless, it is a start, and her approach could possibly be developed, provided that the problems are identified in terms of the construct the test is intended to measure, i.e. which component(s) of the skills-knowledge set they engage.

Given the problems of standardizing source-text difficulty, Campbell and Hale report that current practice seems to be to focus efforts to enhance the reliability of translation tests on the scoring (inter-rater and intra-rater reliability) rather than on the test tasks. In response to the question they set out to answer, “Is it possible to create or select texts of a given level of difficulty?”, they find that “the most profitable avenue is to weight items in a translation text [...]. The point of weighting texts would be to create materials to meet the needs of students and examinees at different levels of development or proficiency” (1999: 2). This is one avenue that could possibly be explored in IGO recruitment testing.

### 2.3.3 *Scoring*

The most objective form of scoring in language testing, according to Bachman and Palmer (1996: 194-229), is to calculate scores on the basis of the number of tasks successfully completed, which is the principle underlying Nord’s grading scheme referred to above. Multiple-choice and cloze exercises might work for discrete testing of some aspects of translation competence, and Nord’s scheme might be appropriate in the context of translator

training, but in performance assessment, and especially in recruitment testing, the goal is to gain an accurate impression of a candidate's overall competence, which means assessing as many aspects as possible, or at least the most important aspects. This is done either by rating performance as a whole, i.e. holistically, or by rating different aspects of the performance separately, i.e. grading analytically (Alderson, et. al, 1995: 107-8). This type of grading is also referred to as multi-trait scoring (Fulcher and Davidson 2007: 254-257).

One holistic approach is to grade intuitively, in other words, read the target text and pass or fail the candidate or award the performance a score on a given scale. This method is sometimes referred to as impression grading (Alderson et al. 1995: 108). The problem of course is that graders may have different criteria for assigning a score (Bachman and Palmer 1996: 117). Providing detailed descriptors for each level on the holistic scale may not help because that would presume all test-takers reach the same level in each aspect of performance. Where to place a candidate who scores high on comprehension but low on writing, for example? Bands can be appropriate when minimum standards are required, such as in certification proceedings, but not in recruitment testing at IGOs, however, unless organizations are looking for solid all-rounders. Recruiters may be assessing proficiency but will settle for excellent potential as well, particularly when demand outstrips supply and, in IGOs, with their well-established revision and mentoring procedures, recruiters are willing to overlook a major error if the translator shows brilliance in other areas.

In nearly all cases, the rating of the components of translation competence or of translation competence as a whole is based on the evidence found in the target text. One common scheme consists of deducting points for errors, on the premise that an error-free translation deserves a perfect score. Waddington found in his survey of 20 European and Canadian universities that 36.5% of teachers evaluated student translations on the basis of error analysis alone (1991a: 16). This might actually seem a valid approach as far as authenticity is concerned, since the quantity of errors a new recruit produces directly affects the amount of revision work they generate. Of course, some errors are more time-consuming to correct than others, which would mean having to adjust points according to the error type. Kiraly, a strong advocate of the incorporation of professional considerations in translator training, proposes grading precisely on the basis of the time required for external revision. He claims that “[a]nchoring texts in realistic translation situations and using credible professional norms of editing time for their evaluation can help provide the necessary authenticity to result in trustworthy assessment procedures” (2000: 162-163). With regard to using revision time as

a measuring stick, he admits “[t]here is no suggestion of objectivity in the evaluation procedure, but there can also be no claim against this approach for arbitrariness” (2000: 159).

Arbitrariness is precisely the criticism leveled at the more common points-deduction schemes, such as that put forward by Adab (2000: 218), which is based on grading on a scale of A-F for coherence and readability and then deducting 1-3 points for errors regarding language accuracy, accuracy of messages, target-language reader’s needs, intertextual elements and acceptability-readability, and then providing bonus points for clever solutions. Such schemes have drawn scathing comments. Campbell and Hale write: “The upside-down marking scheme that seems to be commonly used (error marks being deducted from a perfect score) is so odd as to defy categorization” (2003: 210). One major objection is that, if a norm-referenced approach is used, since there is no bottom to the scale, “there is no way to assess the relative achievement of the top scoring candidates; depending where the bottom of the scale finds itself, one may be very good and one exceptional or perhaps they are separated by a whisker” (2003: 210-211). If a criterion-referenced approach is used with a points-deduction scheme, the problem lies with the arbitrariness of the number chosen to represent the perfect score and the number of errors chosen as the cut-off point. Colina also remarks on the subjectivity of these grading schemes, which she says “fall under the category of experience-based and anecdotal approaches” (2008: 99). Williams comments that points-deduction schemes are only valid in examination contexts when everyone is translating the same text (1989: 18). In other words, comparisons of performances across tests are impossible. He also highlights the problem of bonus points: “Three paragraphs of brilliant translation containing masterstrokes of formulation and interpretation cannot compensate for one critical error” (1989: 19). Kelly states her objection as follows:

[T]he concept of the perfect translation, however implicit, sits uncomfortably with any modern view of translation theory [...] calculating the exact value of a particular error is an immensely complicated task, and can easily end up being arbitrary, hard to justify and far from transparent [...] seemingly objective numerical values are in essence extremely subjective as the consideration of each individual error involves a large degree of subjective evaluation and decision-making. (2005: 140)

In short, the fundamental flaws of such schemes are the arbitrariness of the points “in the bank” that test-takers start with and the difficulty of dealing sensibly with translators who do so well that they score more points than they started out with. Negative points schemes

might be valid if the idea is just to rank test-takers on the same task i.e. in norm-referenced testing, or on a simple pass/fail basis, but they can (should) never be used to compare performance over time because they do not take task difficulty into account.

In terms of inter-rater reliability, schemes based on error-analysis may be more reliable than holistic grading schemes. Trials conducted by Waddington (2001a and b) compared the scores obtained using four methods of scoring: method A was a points-deduction scheme devised by Hurtado Albir which rated errors (-1 or -2) associated with misunderstandings of the source text, inappropriate target-language renderings, and their impact on the transmission of the source-text functions; method B tried to factor in the seriousness of errors by deducting 2 points for a translation error or 1 point for a language error, and adjusting the points deducted according to the number of words in the text affected by the error; method C was a holistic 1-5 scale which took into consideration accuracy of source-text content transfer; quality of expression in the target language, and degree of completion, measured by the amount of revision required; and method D combined methods B and C in a proportion of 70/30. The study showed that, among the factors taken into consideration, translation competence correlated closely with native-language competence and self-assessment of translation competence, but not with mathematical intelligence, and this result was observed regardless of the scoring method used, suggesting the criterion-related validity of all four systems of assessment (2001a: 324). The analysis of inter-rater reliability found that method A, which is based on error analysis, was more reliable than the holistic methods, and interestingly, that the combined method (D), was more reliable than A. Waddington concludes, however that “[a]lthough the statistical analysis indicates the superiority of methods based on error analysis over those based on a purely holistic appreciation, it also shows the limitations of error analysis by itself, and the benefits of combining both approaches” (2001b: 35). Also interesting for our purposes is the fact that methods A and B ranked candidates in very different ways. By my calculations, rankings differed by over 5 places in over a third of the cases (23/64) and by 8-22 places in 11 cases. The choice of grading scheme is therefore critical in recruitment testing.

Scholars have also been quick to point out that the problem of assigning value to different errors is not unique to points-deduction schemes. The problem of “how to distinguish whether an error should be worth 1, 2 or 3 negative points” (Hatim and Mason 1997: 199) is not just a problem of quantification but of judgment and hence of subjectivity. The problem when assessing translators on the basis of their translations is that “all the skills involved in translating are tested at once and errors do not necessarily show which skill is

deficient” (1997: 198). By extrapolation, successful solutions do not necessarily show which skill or knowledge type has been effectively applied. Hatim and Mason suggest that, when assessing, errors be handled according to whether they constitute significant mismatches of denotational meaning between the source text and the target text or breaches of the target-language system (e.g. spelling and grammar). For the rest,

[I]t is a matter of making judgements about the relative acceptability of the range of options from what the translator chooses. Such judgements can, of course, never be completely objectivized. But those who are professionally involved in translating might expect to achieve a considerable level of consensus in assessing the relative adequacy of variant translations – especially if, as suggested earlier, a well-defined focus is provided for each translation task set as a test. (1997: 203)

This means that the more qualified and more informed the graders are of the supposed focus (*Skopos*) of the test task, the better their judgments will be. In other words, it takes a professional to judge a fellow professional, or to judge one in the making. Hatim and Mason suggest cross-referring findings based on evidence produced in the translation to get a feel for the translator’s strengths and weaknesses, and complementing the assessment of the ability to translate whole texts with the testing of discrete skills (1997: 201). This would pave the way for construct validation and contribute to the exploration of item difficulty.

In the meantime, Bachman and Palmer (1996), Alderson et al. (1995) and Hamp-Lyons (1991) from the field of language testing, and Colina (2008), Hatim and Mason (2007), Kelly (2005) and Angelelli (2009) from Translation Studies, among others, all propose rating performance using analytic scales. For the sake of construct validity, the scales must refer to the components of the construct the test is intended to measure: “a prerequisite of a multiple trait instrument is that there is a close match between the writing to be done and the skills and text facets to be evaluated” (Fulcher and Davidson 2007: 255). In translation testing this means creating separate scales to rate the components of translation competence we wish to assess. If a single score is required, the scores for each component can be combined according to the profile sought. This ability to weight items is useful for shifting the emphasis of the test and adjusting task difficulty.

Hamp-Lyons (1991) highlights three advantages of analytic grading or what she calls multi-trait scoring: “salience” (the method enables test-writers to determine what aspects of performance constitute salient traits worthy of scale development); “reality and community”



(“judgements are linked to something that is at once external and internal” through the training and fine-tuning of the scale descriptors the method calls for) and “increased information” (a profile of test-takers is obtained). Another point in the scheme’s favor is that assessing individual components is what raters do anyway (Bachman and Palmer 1996: 211-212).

Alderson et al. state that “[s]cales can help increase the reliability of subjectively judged skills, especially of the productive language skills, and provide a standard and meaning for such judgements” (1995: 84). This holds true in translation as well: it is often easier to judge the target-language writing skills that went into producing an effective solution than the source-language knowledge, for example. Was the correct interpretation of the source-language term the result of applied knowledge or a lucky guess? This suggests that source-language comprehension skills should perhaps be measured separately in translation tests.

Cushing Weigle also notes that the breakdown into components and the ability to train raters permitted by an analytic scoring scheme can help reduce the subjectivity of the assessment process. The main disadvantage is that it takes longer to apply than other schemes (2002: 120). She reports on research into writing assessment that found analytic schemes distinguished professional writers from college writers, but holistic schemes did not. Moreover, professional writers consistently scored higher on analytic scales whereas college writers received approximately the same scores regardless of the scale used. Other research into rating scales has found that analytic scores are more reliable than holistic ones (op. cit. 72-73). The procedure is outlined in Alderson et al.:

In order to produce reliable ratings, it is often necessary to provide series of rating scales, each focusing on a different aspect or combination of aspects of the performance. When raters score performances according to these aspects or criteria, they refer to descriptions of levels and assign scores accordingly. However, the scores that result from the application of such criteria then have to be aggregated and are often weighted according to perceived importance or relevance. The resulting “final” score bears no direct relation to the original scores or to the individual score descriptions, but still has validity as a summary of that performance. (1995: 77)

The use of analytic scales can thus provide test users with both a profile of the test-takers and a single composite score that can be weighted according to needs to identify the

most suitable candidates. This makes them particularly useful in recruitment testing. The only remaining challenge then is the drafting of the descriptors for each component level, the design of the corresponding scale and the training of the raters. As for cut-off scores, Alderson et al. suggest the use of a panel of expert judges to determine the appropriate weighting of component scores and to determine the pass/fail mark (1995: 158-159). This could be easily arranged in the recruitment testing of translators at IGOs, where “experts” abound.

## **2.4 Summary of the literature review**

Many ideas have been produced by Translation Studies scholars on the skills and knowledge that translators, including IGO translators, need to have, but hardly any empirical data have been analyzed, and there appears to be no data on the relative importance of items. The above literature review has made it possible to define broad categories of skills and knowledge and identify several skills and knowledge types specific to IGO translation. These provided the starting point for our research and the design of the method devised for ranking the components of the skills-knowledge set described in Part I. Meanwhile, the literature on language testing and the insights of translator-training scholars on assessment highlight the importance of various aspects of translation test design: (1) test tasks should be authentic (domain-relevant), interactive (able to activate the components of the construct to be measured) and pitched at the right level of difficulty to be able to discriminate among candidates; (2) the difficulty drivers of the task should be source-text features or test method facets that are expected to activate the components of the construct; (3) separate scales should be designed for rating the components of the skills-knowledge set on the basis of the evidence produced in the test; and (4) measures to maximize the validity and reliability of the test should be taken at all stages of the process. These key considerations are taken into account in the examination of testing practice at IGOs, presented in Part II below.



**PART I**  
**IDENTIFICATION OF THE SKILLS- KNOWLEDGE SET REQUIRED BY**  
**TRANSLATORS AT INTER-GOVERNMENTAL ORGANIZATIONS**

*“Let us dispel the notion that the translator or interpreter must simply be a good linguist, that is to say must simply have a good command of another language.”*

Stephen Sekel<sup>7</sup>

---

<sup>7</sup> Former Chief of the English Translation Service, United Nations Headquarters. Words taken from the keynote address delivered at the 18th FIT World Congress, 4-7 August 2008, Shanghai, China.



## **Chapter 3 Methodology for the identification and ranking of the skills and knowledge IGOs should seek in their recruitment processes**

### **3.1. Choosing a methodology**

Competence cannot be measured directly but must be detected through product and process analysis, i.e. through the manifestations and displays of that competence. Product (text) analysis can only suggest the skills and knowledge that translators apply when they translate; it cannot rank them in order of importance. Process analysis, meanwhile, is not practical on a large scale as a research method. An alternative way to find out about the skills and knowledge used in IGO translation is to consult the people who have the greatest first-hand experience in using and observing their application and repercussions and hence have the greatest awareness of their relative importance to an organization. Those people are the professional translators and revisers who work exclusively for IGOs. A survey was therefore chosen as the tool for collecting data that could be used to determine and weight the skills-knowledge set(s) of ideal candidates for IGO translation work.

#### *3.1.1. How to approach importance?*

Having decided to use a survey to obtain data on the relative importance of the components of the skills-knowledge set, the next step was to determine how to approach the matter of importance. The first consideration was: important to whom? Answer: the IGO, and more specifically the translation service. The second consideration was: important for what? Answer: for producing translations of the type and in the manner required. From the recruitment angle, the relative importance of skills and knowledge could be understood to depend (1) on their impact on the translations produced by the organization, or (2) on their relative scarcity among new recruits. If new recruits lack the skills and knowledge needed on the job, they cannot produce translations of the type and in the manner required. In IGOs this affects output, either by lowering the quality (by their standards) of the translations published or by increasing the revision load. The logic of what we might call approach 1 is that the skills and knowledge types with the highest impact are the ones the service needs to award priority to in its recruitment procedures, in order to maximize their positive influence. The logic of approach 2, on the other hand, is that the skills and knowledge most often found

lacking among new recruits are the ones the service needs to find in greater abundance, in order to minimize their negative influence.

A questionnaire based on the first approach would collect largely self-report data. Respondents would probably answer on the basis of a positive assessment of the skills and knowledge they think they apply in their work. This would be an evaluation of the situation from the demand side – what is currently needed by the organization. A questionnaire based on the second approach would, if answered by revisers, collect data based on the performance of others, generated by the negative assessment of the components of translation competence that recruits do not have. This would be an evaluation of the situation from the supply side – what is currently available (or not available) among recent recruits.

It was decided to pilot both types of questionnaire to determine which approach was more appropriate for determining recruitment priorities. The first would be called the “impact questionnaire”, and the second the “recruits questionnaire”.

Data on what IGOs expect from new recruits was also sought to contextualize the findings, since a high frequency rating in the recruits questionnaire (i.e. the suggestion that a skill or knowledge type is often lacking) might not be relevant if the IGO does not expect new recruits to start out with the skill or knowledge in question because that component can be (or has to be) acquired after recruitment through on-the-job training or experience.

The main hypotheses being tested at the outset in Part I were thus as follows:

*H<sub>1</sub>: Certain skills and knowledge types are more important than others in the context of IGO translation.*

and

*H<sub>2</sub>: Certain desired skills and knowledge types are more often lacking than others among new recruits at IGOs.*

Where:

1. “Important” refers to two aspects of the work of IGO translation services: their productivity and their reputation.
2. “New recruits” are translators who have been working for the organization for less than one year.

These hypotheses obviously rest on the assumption, suggested in the literature on translation competence, that IGO translators use several types of skills and knowledge in their work.

## 3.2 The first pilot

The aim of the first pilot was to answer the following questions about the two questionnaires:

- Will the answers to the questionnaires produce hierarchies that are useful input for recruitment test design?
- Can the questionnaires be developed into reliable measuring instruments?

### 3.2.1 *The base list of skills and knowledge types presented for rating in both questionnaires*

The list of skills and types of knowledge presented for rating in both the impact and the recruits questionnaires was drawn up on the basis of the broad categories identified in the literature on translation competence and the specific skills and knowledge suggested by the review of the literature on IGO translation. Since the purpose of the study was to obtain a hierarchy of skills and knowledge types as input for the design of text-based translation tests, the level of specificity of the text-production skills was greater than other types, and interpersonal and other skills and knowledge types that do not lend themselves readily to text-based testing were excluded from the base list.

The skills and knowledge types as grouped in the first pilot are presented below, together with the reasons for their inclusion. For convenience, the following acronyms have been used here (and in other parts of the thesis): SL for source language, TL for target language, SC for source culture, TC for target culture, ST for source text and TT for target text. The numbers in parentheses alongside each skill or knowledge type are those used to identify them in the questionnaires and during the first pilot.

1. Knowledge types: *knowledge of (1) the SL (vocabulary, expressions, rhetorical devices), (2) the SC (including sensitive issues, social, economic and political situations), (3) the subject, and (4) the organization and how it works.*

SL knowledge was obviously included as an essential comprehension tool. Ivanova has pointed out the importance of SC knowledge when translating political texts (cited in Schäffner 2007: 142). Knowledge of the organization and how it works could be seen as a type of SC knowledge and also as important for understanding the IGO's approach to translation, referred to by Mossop (1988: 65). The strong subject focus at some IGOs (e.g. the European Court of Human Rights) or the highly technical nature of the texts translated (e.g. at the International Telecommunications Union) made subject knowledge a possible priority.



2. Comprehension and analytical skills: (5) *understand complex subjects*, (6) *master new subjects quickly*, (7) *work out the meaning of obscure passages in the ST*, (8) *detect inconsistencies contradictions, nonsense, unintended ambiguities, etc. in the ST*, and (9) *detect logical, mathematical or factual errors (“eye for detail”)*.

One of the challenges in some IGOs, as pointed out by Sekel (2008), is that translators need to remain abreast of developments in a broad range of fields, some of which may be extremely complex. Hence the inclusion of skills (5) and (6). Translating obscure passages is one of the challenges of translating texts written by individual or multiple authors, who may have an imperfect command of the language or insufficient knowledge of its norms of discourse (Cushing Weigle 2002: 20-22) or have been working in a multilingual environment for so long that they have lost their linguistic bearings (Mossop 2001: 42). Being able to detect inconsistencies of various types (skills 8 and 9) is important when working from unedited STs since the reputation of an IGO is transmitted and preserved in its translations, and IGO translators are contractually, if not morally, bound to protect that reputation (Koskinen 2000: 68).

3. Target-text production skills: (21) *write and convey the ST message clearly*, (22) *tailor language to the targeted reader*, (23) *capture every nuance of the ST in the translation*, (24) *bring translations into line with new versions, i.e. input changes smoothly making any modifications necessary to surrounding text*, (25) *achieve the right tone and register*, (26) *improve on the style of the ST (e.g. remove tedious repetitions, pare down convoluted sentences, eliminate verbosity)*, (27) *produce translations that flow smoothly even when the ST does not*, and (28) *improve on the content of the source text (remove, correct or at least point out redundancies, unintended ambiguities, contradictions, unexplained acronyms, misleading headings, etc.)*.

The need to produce well-written TTs even when working from poorly written STs was behind the inclusion of skills (26) and (27), which refer to improving on the style and even the content of the original. The high level of accuracy sought and the need for political correctness explain the presence of the abilities to capture nuances (22) and achieve the right tone and register (25). The need for clarity (21) was highlighted by Wagner et al. (2002: 73) and Hindle (1965/84), and inserting changes to TTs (24) responded to an aspect of IGO translation work pointed out by Wagner et al. (2002: 74).

4. Research skills: (29) *find the correct terminology*, (30) *track down sources of relevant background information (mining for information)*, (31) *work with reference documents (to find accepted phrasing and terminology)*.

These skills were held to possibly be important given the technical and complex nature of the subjects covered by some IGO documents, as well as the archival purpose of conference documents, for example, and the fact, pointed out by Koskinen (2008: 25), that IGO documents often form part of a series. This makes consistent terminology essential.

5. Computer skills: (32) *type accurately and fairly fast*, (33) *reproduce the formatting of the source text*, (34) *work with more than basic Word functions (track changes, tables, autocorrect, etc.)*, (35) *work with translation memory software*, (36) *work with Excel documents*, (37) *work with PowerPoint presentations*.

All these skills could be important for translators in any field today. The idea was to see where they would fit in in relation to the others.

6. Other skills related to translation work: (38) *meet tight deadlines (translate fast and under pressure)*, (39) *explain translation decisions or problems with the source text (e.g. to authors, users or revisers)*, (40) *follow complicated instructions about what needs to be done with a text (additions that need translating, parts that need relocating, patching together, revising against new versions, etc.)*, (41) *adhere to in-house style guidelines*.

Like all employers of translators, IGOs could be expected to be interested in the productivity of potential new recruits, especially if they are to work in-house and be paid by the day. Hence the inclusion of skill 38. The inclusion of the other three skills in this category reflected particular aspects of IGO translation work mentioned by Koskinen (2008) and Wagner et al. (2002). The ability to explain decisions or problems is strictly speaking not a text-production skill and should probably not have been included. It clearly falls under interpersonal skills, not usually assessed in text-based examination processes. It was originally included because writing e-mails to authors, editors, users and revisers explaining why a sentence could not be translated without more information or changes being made to the wording in the ST, for example, was a frequent task at the pilot IGO, and there had been vague plans to devise a text-based test in which candidates would be asked to prepare e-mails to explain the problems posed by the ST or to include such explanations as annotations.

In addition to asking respondents to consider these skills and knowledge types, questions were included about different types of errors. In the impact questionnaire, the idea was to see how serious the errors were considered to be (i.e. how important they were), and in the recruits questionnaire the idea was to see how much revision work they generated in relation to one another (i.e. how frequently they were made). The errors included were as follows: (10) *slight meaning errors (inaccuracies)*, (11) *spelling mistakes*, (12) *incorrect grammar*, (13) *punctuation errors*, (14) *formatting errors*, (15) *unwarranted omissions*, (16) *unwarranted additions*, (17) *errors of fact (e.g. names that have not been checked)*, (18) *unidiomatic (unnatural-sounding) language*, (19) *poor word choices (inexact terms, inappropriate language)*, and (20) *stylistic errors (e.g. dangling modifiers, faulty parallelism, poorly constructed sentences, excessive use of the same high-impact words, mixed metaphors)*.

The errors reflect those identified in the scheme provided by one IGO (under strict terms of confidentiality) and follow those commonly used when accounting for revision work (Mossop 2001). Meaning errors, poor word choices, and unwarranted omissions and additions, together with the mechanical errors (grammar punctuation, spelling and formatting) and stylistic errors undoubtedly make up the bulk of revision work. The need for idiomaticity was pointed out by Schäffner (2007:141). Errors of fact refer to a more IGO-specific requirement: the need to preserve the image of the institution by ensuring not only good style, but accurate content as well. The concept of error was included because it was thought to be readily understandable to revisers and because it would be useful when cross-referring the findings with data on translation test grading.

### 3.2.2 *The phrasing of the questions in the first impact questionnaire*

The questions on skills were phrased “To what extent are the following skills important?” The questions on error types were phrased “How important is it that translations at your organization do not contain...?” In both cases, respondents were asked to reply using the following scale: irrelevant, not very important, desirable, very important, essential. The questions on knowledge were phrased “How much knowledge is required?”, and respondents were asked to reply using the scale: none, minimal, some, sound, extensive. The option of answering “irrelevant” or “none” was included to ensure the content validity of the questionnaire, and for the same reason, respondents were invited to list and rate any skills or

types of knowledge or error they thought should have been included. A nominal scale was used on the assumption that it might be easier to use.<sup>8</sup>

### *3.2.3 Other questions included in the impact questionnaire*

Context questions were included on the type of translation work revised (i.e. in-house or external), and respondents were asked which type of translators they would base their answers on. The idea was to get them to think in terms of “translators” rather than “I”. Questions on experience in IGO translation were also included. The version of the impact questionnaire used in the first pilot is presented in Appendix 1.

### *3.2.4 The phrasing of the questions in the recruits questionnaire*

The questions on skills and knowledge were phrased “How often do new recruits lack...?”. The questions on error types were phrased “How often do the translations by new recruits contain...?”. Respondents were asked to reply in both cases using the following scale: never, rarely, sometimes, often, and always. In each case they had the option of answering “not required” or “not applicable”. As in the impact questionnaire, respondents had the opportunity to list and rate any skills or types of knowledge or error they thought should have been included.

### *3.2.5 Other questions included in the recruits questionnaire*

Respondents were also asked to indicate whether each skill or knowledge type was expected to be acquired before or after recruitment, and what kind of in-house training was provided for new recruits. Skills not expected to be acquired until after hiring, regardless of how often they may be lacking among new recruits, do not need to be tested in recruitment. The in-house training questions were included to see how soon after hiring a new recruit might be

---

<sup>8</sup> “Although agreement with attitude items offering five or seven response categories is technically a discrete ordered variable, an analyst can make use of very powerful statistical techniques by assuming that these categories have continuous properties. The rationale behind this assumption is that “strength of attitude” is fundamentally a continuous property that our measuring devices only crudely tap. Our major concern should be whether treating ordered categories as continuous results in distorted conclusions from statistical analyses. So far, most evidence suggests that statistical techniques are fairly “robust” under such circumstances.” (Knoke and Bohrnstedt 1994: 21)

expected to acquire some of the skills. If an intense induction course on in-house terminology tools and style rules is given upon arrival, for example, new recruits can be expected to know how to apply them from early on. The version of the recruits questionnaire used in the first pilot is presented in Appendix 2.

### 3.2.6 *The respondents*

For the first pilot, two revisers working for the same IGO were asked to participate. Revisers were chosen rather than translators as they are in the best position to rank the items in both questionnaires. By producing translations themselves and revising those done by others, they gain awareness of the skills involved and of their direct positive impact on the effectiveness of the translations produced (our definition of importance). Revisers are probably the most aware of the nature of the work and the standards required as they are almost always the most experienced translators in IGOs.

In addition to making translations ready for publication, revisers are responsible for mentoring new recruits. Through their interaction with junior translators in feedback sessions and the direct manifestations of a lack of skill or knowledge encountered in the translations produced by new recruits, revisers gain a picture of the shortcomings of new recruits and how often certain gaps in skills or knowledge account for the revisions they have to make (transcripts of such interactions are provided in Appendix 3). Moreover the revisers' opinions were likely to be more valid than self-report data (the opinions of new recruits themselves) or self-report data with hindsight (the opinions of translators trying to remember what they were like when they first joined the organization). As mentors and supervisors, revisers are used to trying to determine the strengths and weaknesses of new recruits as objectively as possible in relative and absolute terms, with a view both to helping their junior colleagues and to assessing their value to the organization as a human resource. Self-report data, meanwhile, may be marred by ignorance and/or subjectivity. Novice translators could either underrate or overrate the importance of a skill they know they lack, for example, depending on their self-awareness, their understanding of the demands of the job and how they respond to negative feedback.

Revisers from the same organization were chosen in order to minimize the influence of one possible context variable: the organization for which respondents worked. The IGO in question was chosen out of convenience: the researcher was familiar with the work of the

translation department and had constant access to its members. Respondents requested, and were assured, anonymity.

The profile of the two respondents was as follows:

Reviser A: 16 years' experience in translation since obtaining a degree in Translation and Interpreting. Translates from French, Portuguese and Spanish into English. Has been working for IGOs for fourteen years and as a reviser for the last nine.

Reviser B: 33 years' experience in translation since obtaining a degree in Linguistics. Translates from French, Russian and Spanish into English. Has been working for IGOs for over sixteen years and as a reviser for the last six.

These profiles are fairly typical of IGO revisers. Few are hired from outside the organization unless they have experience at other IGOs. Most are translators who have worked their way up the rather military-style organizational structure of IGO translation services.<sup>9</sup>

### *3.2.7 Other participants in the pilot*

To contextualize the results of the two questionnaires and thus interpret the findings better, consultations were held informally at various stages during the pilot with six translators (with 3 months, 7 months, 2 years, 4 years, 5 years and 14 years experience at the IGO) and the head of the department responsible for all translation and editorial services. The purpose of these consultations was to obtain background information on the work of the unit, testing practice, the revisers who completed the questionnaires and in-house training arrangements.

### *3.2.8 Assumptions*

The following assumptions were made in the first pilot:

1. The revisers would understand the instructions and the questions and be able to use the rating scales.
2. The revisers would answer the recruits questionnaire honestly because they had nothing to gain or lose: they were not directly responsible for the shortcomings of new recruits. Any misgivings they might have about their answers reflecting badly on their employer should be overcome by the promise of anonymity.

---

<sup>9</sup> Personal observations confirmed through informal consultations with revisers at several IGOs.

3. The revisers would answer the impact questionnaire honestly because they had nothing to gain or lose: the impact of the skills and knowledge would not reflect on them in any way.
4. The revisers would answer honestly about the possibilities of post-recruitment acquisition because this information could be easily verified by consulting other sources.

### *3.2.9 Procedure*

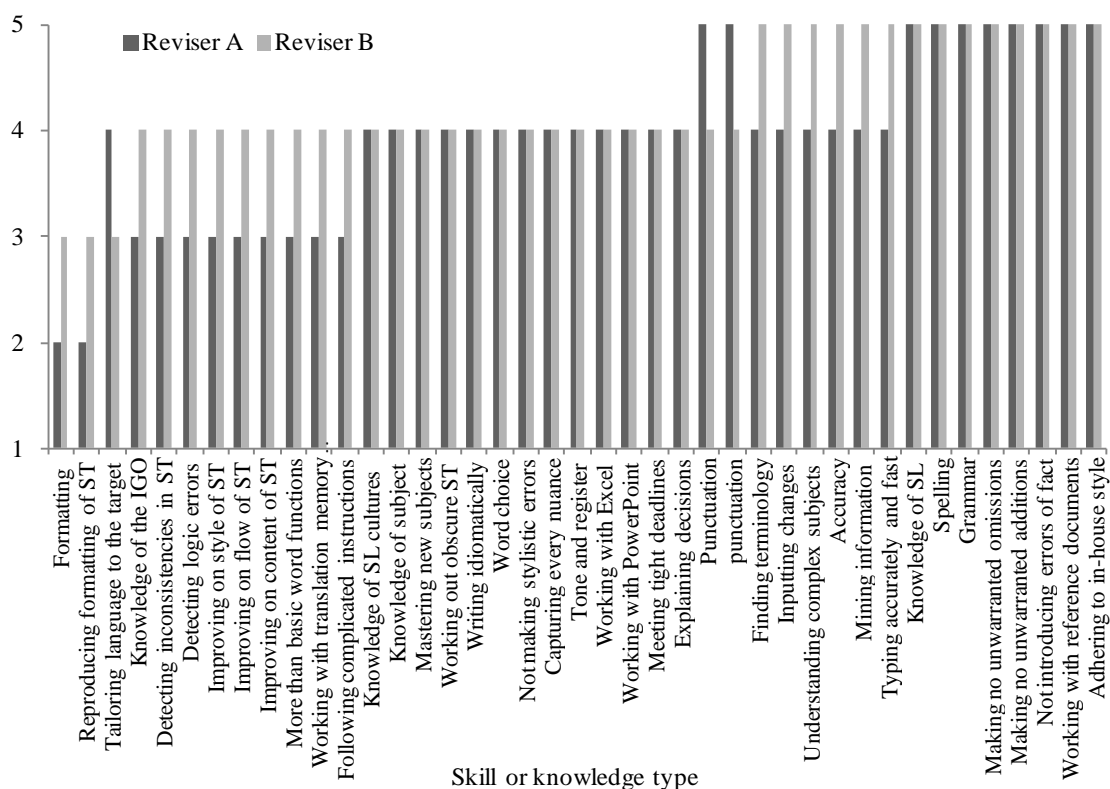
Both questionnaires were completed by both respondents in late August 2009, with a five-day gap between one set of questionnaires and the other in an attempt to ensure that responses in the second were not strongly influenced by those given in the first. Informal follow-up interviews lasting about 20 minutes were conducted with each reviser individually immediately after each questionnaire had been completed, in order to provide them with the opportunity to comment freely on the questionnaires (clarity of instructions, wording of definitions, the usability of the rating scales, etc.). Semi-structured interviews also lasting approximately half an hour were then held again with each reviser individually once the results had been compiled and compared. On this occasion, they were asked to explain and justify their answers to each question. The goal of these interviews was to determine the content validity of the questionnaires and to identify the information needed to contextualize and accurately interpret the findings.

### *3.2.10 Results of the first pilot of the impact questionnaire*

#### *3.2.10.1 The impact ratings*

The ratings awarded by the two respondents to items on the base list in the impact questionnaire are presented in Figure 3.1. Reviser A answered on the basis of the work of external translators. Reviser B answered on the basis of the work of in-house translators. The vast majority of items were rated by both revisers (29/41 and 38/41) the highest or second highest level on the scales. Ratings coincided in over half the skills and differed by only one step in nearly all the others. The Pearson correlation coefficient is 0.66, which is fairly close considering the respondents were referring to different groups of translators. "Inputting changes smoothly" was the only skill for which answers differed by two steps on the scale. Reviser B tended to rate skills higher than Reviser A did (17/41 times). Reviser B rated three skills higher than Reviser A did.

Figure 3.1. Impact ratings awarded for selected skills and knowledge types, by respondent\*



## \*Notes

The descriptors of the rating scale varied according to the questions asked:

- Knowledge ratings were awarded in response to the question “How much knowledge do translators at your organization need in the following areas?” Possible answers: none, minimal, some, sound, extensive.
- Skills ratings were awarded in response to the question “To what extent are the following skills important?” Possible answers: irrelevant, not very important, desirable, very important, essential.
- Error-avoidance skills ratings (accuracy, spelling, grammar, punctuation, unwarranted omissions, and additions, errors of fact, idiomaticity, word choice, style) awarded in response to the question “How important is it that translations at your organization do not contain... errors?” Possible answers: irrelevant, not very important, desirable, very important, essential.

For full descriptors of the items rated, see the text of the impact questionnaire in Appendix 1.

The relatively strong positive correlation between the ratings suggests a high degree of coincidence between the respondents on the relative impact of skills and knowledge types. The spread of the ratings suggests that it is possible to group the skills and knowledge types needed in IGO translation according to their relative impact.

### 3.2.10.2 Skills and knowledge types reported as missing from the base list

Reviser A said that “the ability to ensure the coherence or internal consistency of the target text” should be included in the list as “very important”. Otherwise, the revisers said that there



were no additional skills or types of knowledge that should have been included or ones that they felt should not have been included.

### *3.2.11 Analysis of the results of the first pilot of the impact questionnaire*

There was a clear tendency to assign less importance on the part of Reviser A, who awarded lower ratings than Reviser B on 17 out of the 41 occasions (41.5% of answers). This could reflect, among other possible influences, (1) a consistently different interpretation of the rating scale, (2) a different understanding of importance as revealed in interview, or (3) the different type of work given to external translators, which makes several skills less necessary at this IGO. Each of these possibilities is discussed below.

1. A consistently different appreciation of the scale would not matter that much, as the influence of one respondent would be relatively small in a larger sample population. For what it may be worth, both respondents said they had no difficulty using the scales.
2. A completely different understanding of “importance” would undermine the validity of the findings, however. In the interviews, the respondents said they had had no problems understanding the questions or which skills and knowledge were being referred to in the questionnaire. They reported that they had based their answers mainly on two considerations, sometimes separately, sometimes simultaneously: (i) how devastating the impact was of an error caused by the absence of that skill (e.g. a failure to detect errors in the original), and (ii) how often the skill was required (e.g. specialized knowledge). Their thinking went along the lines of “we really use that skill a lot” or “that’s important because that kind of error is a major source of embarrassment”. Although the two revisers in the pilot had not thought it a problem, clearer instructions would be needed to try to make sure that respondents took both aspects of importance into account. As it stood, the wording of the questionnaire meant that “importance” could have been interpreted in many different ways, including as referring solely to the frequency with which skills are required, which would not factor in the devastating impact of certain error-prevention skills that are possibly hardly ever called into play, but are vital when they are needed. That danger would have to be averted with careful rewording. Another point that came up was that, although respondents had been asked “Which type of translators will you base your work on? In-house or external?”, they had both responded in terms of the work itself and what “we” or “I” did. Not once did they refer to “they”. The impact questionnaire seemed to produce self-report data.

3. The last factor, the in-house/external variable, explains the largest discrepancy in the two revisers' answers. The one skill in which there was a difference of more than one step in rating was "in-putting changes", which is a direct reflection of the fact that external translators (who Reviser A accounted for) do not have to perform this task at the organization. She explained in the interview that she had rated the skill as "desirable" but not "irrelevant", however, because temporary in-house staff are usually recruited from among external translators and it was therefore "somewhat important" that they have the skill in case they are invited to work in-house for a while. This showed how useful it is to have access to information that makes it possible to contextualize findings. It was also a reminder of the need to have a large enough sample to minimize the distorting influence of slightly unusual answers like this one and to carefully qualify any claims about the representative nature of findings.

In other regards as well, the tasks performed by translators were found to determine the need for certain skills. Target texts are formatted by text-processing experts at this IGO. Translators and editors therefore do not need to worry about font size, paragraph indentation, bulleting, or any other text formatting issue. They are, however, responsible for capitalization and punctuation because the rules change from one language to another. Hence the high ratings for spelling and punctuation skills and the low rating for formatting abilities.

### *3.2.12 Ways in which to improve the reliability of the impact questionnaire as a measuring instrument*

The ratings were possibly distorted not only by the failure to define "importance" at the beginning of the questionnaire but also by the different phrasing of the questions on knowledge, errors and skills, and the different scales used. Any attempt to draw up a hierarchy of items that had been rated using different scales would be essentially flawed. The questions would therefore need to be reformulated and the scales redesigned. Specifically, the following changes appeared necessary:

- Define "importance" for respondents. Impact is a less ambiguous concept than importance, especially if the target is named, and it can be rated quantitatively.
- Rephrase the questions on error types to refer to skills.
- Convert the scales to five-point Likert scales, which offer the advantage of a neutral mid-point, with descriptors only for the lowest and highest levels.

- Standardize the wording of the questions and the rating scales for all the skills and knowledge types.
- Incorporate skills and knowledge types mentioned by respondents as missing from the profile.
- Make sure context information on the responsibilities of translators in the document production process are obtained.
- Make sure data can be separated by in-house/external type of work.

### *3.2.13 Results of the recruits questionnaire*

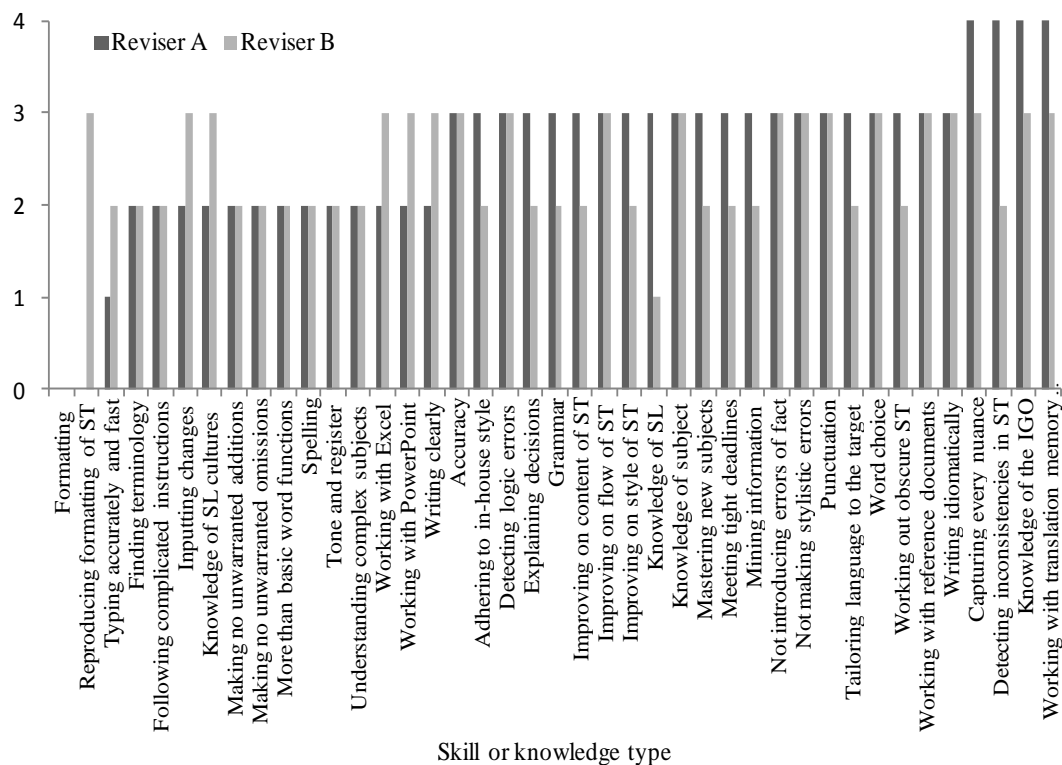
#### *3.2.13.1 The frequency ratings*

The ratings awarded by the two respondents to items on the base list in the recruits questionnaire are presented in Figure 3.2. The vast majority of skills and knowledge types (85% in the case of Reviser A and 95% in the case of reviser B) were rated by both revisers as either rarely lacking or sometimes lacking. Ratings differed by only one step on the scale in the case of almost half the skills (19/41). The Pearson correlation coefficient is 0.38, which is a weak correlation, compared with 0.66 in the impact ratings, showing that the difference between the respondents' answers is greater in the recruits questionnaire than in the impact questionnaire.

There were only three skills in the recruits questionnaire for which answers differed by two steps on the scale, however: SL knowledge, the ability to detect inconsistencies, and the ability to reproduce the formatting of the ST. In 38 out of 41 cases (93%), their answers either wholly coincided or only differed by one step on the five-point scale. There is also no clear tendency to rate more severely: Reviser A rated 15 skills as more frequently lacking than Reviser B did. Reviser B rated 7 skills as more frequently lacking than Reviser A did.

The spread of the ratings suggests that it is possible to group the skills and knowledge needed in IGO translation according to their relative scarcity among new recruits. The lower correlation of the ratings in comparison with the impact ratings suggests that frequency ratings can be expected to vary more, which is not surprising considering they are based on observations of translators' performances rather than perceptions of the nature of the work.

Figure 3.2. Frequency ratings awarded for the shortage of selected skills and knowledge types among new recruits, by respondent\*



\*Notes:

- Error-avoidance skills ratings (accuracy, spelling, grammar, punctuation, unwarranted omissions, and additions, errors of fact, idiomaticity, word choice, style) awarded in response to the question “How often do the translations of new recruits contain... errors?”
- All other ratings of skills and knowledge types awarded in response to the question “How often do new recruits lack ...”. Possible answers to all questions: 1=Never 2=Rarely 3=Sometimes 4=Often 5=Always.
- “N/A (not applicable)” answers are reported as a zero rating in the figure.
- For full descriptors of the items rated, see the text of the recruits questionnaire in Appendix 2.

The relatively low ranking of “finding the correct terminology” was also slightly surprising, but in the interviews it was explained that all source texts at this IGO are referenced prior to translation by professional terminologists. Similarly, the relatively low ranking of “understanding complex subjects” and “mastering new subjects” was explained as a reflection of the fact that new recruits do not use these skills because they are not given “difficult” work. This suggested that findings of the recruits questionnaire need to be interpreted in the light of information on the kind of work that new recruits do and how it differs from the work of the more experienced translators in the organization.

### *3.2.13.2 Skills and knowledge types reported as missing from the base list*

The revisers said that there were no additional skills or types of knowledge that should have been included, or ones that they felt should not have been included.

### *3.2.14 Analysis of the results of the recruits questionnaire*

As with the impact questionnaire, in the follow-up interviews the revisers said they had had no problems understanding which skills and knowledge were being referred to in the questions and that the questions were clear and the ratings scales easy to use. Various factors may have influenced their answers, however, as discussed below.

#### *3.2.14.1 The understanding of “new” in “new recruit”*

Reviser A answered on the basis of external translators (7 people) and temporary short-term, in-house staff (2 people). One person fell into both categories. She did not take new permanent in-house translators into account. Reviser B answered mostly on the basis of temporary and (relatively) recently appointed permanent in-house staff (3 people). This affected each reviser’s interpretation of “new”: externals at the organization are not hired again if they have not reached a certain standard in the first six months. After six months, they are not really considered new: they are considered tried and tested. Permanent staff, on the other hand, start with a two-year probationary appointment and often have their work revised for years, even decades in some cases. For Reviser B, therefore, “new recruits” included translators with up to two years’ in-house experience with the organization, while Reviser A only considered translators with less than six months with the organization. This could explain why Reviser A found skills and knowledge more often lacking. It also suggested that “new” should be defined in the instructions for respondents.

#### *3.2.14.2 The in-house/external work factor*

The difference in the recruitment procedures for in-house and external translators and hence their profiles could also have influenced ratings. Interviews with the head of service and other translators and revisers in the unit revealed that external translators are actively recruited on an ad hoc basis among bilingual experts in the subject covered by the IGO. Their technical knowledge is deliberately sought. Permanent in-house translators, on the other hand, are recruited through competitive examinations that tend to select candidates who have highly developed linguistic skills, but are not expected to be experts in the subject. This was

corroborated by an analysis of recent test papers used in the competitive examinations for permanent staff and the ad hoc tests for external freelance staff (provided under strict conditions of confidentiality and consequently not reproduced here). This could have influenced the ratings and might perhaps explain the larger difference in the ratings of SL knowledge.

#### *3.2.14.3 The revision style factor*

Another possible factor was that, according to the translators and the head of department, the two respondents revise in different ways. Reviser A is apparently very thorough and picks up every content-related mistake (distortions of meaning, terminology, specialized knowledge) and in-house style error, while Reviser B picks up fewer content errors, is a stickler for grammatical correctness, in-house style and usage, and makes more stylistic changes, usually by reworking sentences "to make them sound wonderful", as one translator put it. Reviser B does not seem to revise against the ST closely. The translators and the head of service therefore seem to hold both revisers in high esteem, but recognize different approaches in each one. Reviser A is more content-oriented; Reviser B is more style-oriented. Reviser A rated SL knowledge as something that is *sometimes* lacking; Reviser B reported that translators *never* lack knowledge of the SL. This may reflect the fact that Reviser B is known for not revising carefully against the ST (or that Reviser A's answers are based on the work of externals and temporary staff who only have to pass ad hoc exams that are not as demanding in terms of SL knowledge as the formal examinations passed by the translators on whom Reviser B's answers are based). Reviser A's higher scarcity rating for the ability to detect inconsistencies in the ST could also reflect the tendency of Reviser B to not revise closely against the ST. The influence of idiosyncratic revision style can only be counteracted by maximizing the sample size.

#### *3.2.14.4 The influence of anecdotal evidence*

As could be expected, the respondents did not answer consistently on the basis of the same set of new recruits. Sometimes they considered the work of just a few translators; sometimes a more "average" rating was given for all the translators they had in mind. In other words, the revisers answered some questions on the basis of one recruit and others on the basis of several. This can also be counteracted by maximizing the sample size and possibly by rewording instructions regarding what respondents should base their answers on (and then hoping they will do so).

### *3.2.14.5 The answers to the context questions*

With regard to the expectations of new recruits, answers coincided just over 80% of the time. Reviser A listed seven more skills as expected to be acquired after recruitment than Reviser B did: mastering new subjects, working out obscure STs, detecting inconsistencies in STs, detecting logic errors (eye for detail), improving on the style of the ST, improving on the content of the ST, and following complicated instructions. Reviser A and Reviser B also disagreed about one other item, knowledge of the subject, which was reported as expected to be acquired prior to recruitment Reviser A and after recruitment Reviser B. Overall, Reviser A and Reviser B coincided in 33/41 items. The discrepancies can be explained either by the varying interpretations given to the question or by the type of translation work on which the revisers based their answers and the screening procedures used at the IGO. In the follow-up interviews, it was revealed that Reviser A answered mainly on the basis of external translators, who are recruited from among experts in the subject, while Reviser B answered on the basis of permanent in-house translators, who are subjected to more rigorous language testing but are not expected to have as much subject knowledge.

With regard to in-house training, both revisers reported that no formal training is offered to new recruits. Informal training is provided in working with the translation memory software the IGO uses. A workshop is also given to new recruits on in-house style. New in-house recruits are closely monitored and mentored by the more experienced translators and the revisers until they show they no longer require it. This can be years after recruitment. Feedback is given verbally and through written feedback sheets that explain corrections made by the reviser. No training is provided to external translators.

### *3.2.15 Reactions to the questions on expectations and training*

In the interviews, it became apparent that the questions about the expectations and training of new recruits had aroused negative feelings among the respondents because they highlighted the organization's shortcomings in terms of the training it provides and the lack of proficiency of the translators it finds through its recruitment process. The respondents were worried that their answers would make the organization look bad, a reflection of the loyalty to the institution that IGOs try to foster as commented by Koskinen (2000: 68). This undermined the assumption we made at the outset that respondents would answer honestly because the shortcomings of new recruits would not reflect on them personally. The

respondents in the pilot said they had answered candidly in the pilot, however, because they knew the researcher already had reliable information from other sources and they had been assured anonymity. Nevertheless, the negative reaction to the questionnaire and the possibility of the context questions producing distorted answers or frequency ratings was a concern and would need to be addressed.

### *3.2.16 Ways in which to improve the reliability of the recruits questionnaire*

No measures can be taken to counteract different revision styles of respondents, just as none can be taken to reduce the influence of anecdotal data (the influence of the most memorable experiences), beyond maximizing the sample size. Measures could nevertheless be taken to neutralize the influence of other variables and to improve the affective response to the recruits questionnaire. Those that seemed feasible included:

- Defining “new” in “new recruit”;
- Making sure data can be separated by in-house/external type of work;
- Obtaining information on whether skills are to be acquired before or after recruitment and/or on in-house training, in a separate section from the ratings questions or in a different questionnaire in order to prevent any interference of negative feelings associated with inadequacy in that regard;
- Obtaining information on document processing to identify the need for certain skills;
- Rephrasing the questions on error types to refer to skills;
- Converting the scales to 5-point Likert scales, with descriptors only for the lowest and highest levels to minimize the influence of verbal level descriptors and not force respondents to choose negatively or positively. Items would thus be scored, making the data easier to handle in subsequent analysis;
- Standardizing the wording of the questions and the rating scales for all the skills and knowledge types.

### *3.2.17 Conclusions drawn from the first pilot*

Despite flaws in the wording of instructions and questions and in the design of the rating scales, both questionnaires could, with some adjustments and the support of adequate context data, be converted into instruments for drawing up hierarchies of the skills and knowledge used in IGO translation: one based on their relative impact on the communicative aims and



reputation of the organization, and the other on their relative shortage among new recruits. It was therefore decided to improve the design of the questionnaires and pilot them again to further investigate their usefulness.

### **3.3 The second pilot**

The basic objective in the second pilot was the same as in the first: to determine whether the impact and recruits questionnaires, either separately or together, could provide valid input for recruitment test design.

#### *3.3.1 The revised questionnaires*

In light of the findings of the first pilot, a number of changes were made to the questionnaires. These are described in detail below. It should be noted that I am unaware of any empirical way to test that the skills and knowledge types are actually separate items and doing so was not the prime aim of the research. The changes to the definitions of the items on the base list were made to ensure that it would be understood by respondents, with a view to proceeding on a trial-and-error basis.

- The questions referring to error types were rephrased to refer to specific text-production skills.
- In the recruits questionnaire, the instructions for rating the frequency with which skills and knowledge are found lacking were rephrased and standardized as: “Please think about the mistakes you usually correct when going over translations by new recruits. How often do you think those mistakes are due to a lack of the following things? 1 = Almost never; 5 = Almost always; N/A = Not applicable because the skill or knowledge is not required of new recruits.” New recruits were defined as “translators who have been working with the organization for less than 12 months”.
- The instructions for rating the importance of skills and knowledge in the impact questionnaire were rephrased and standardized as: “For the purposes of this survey, effective translations are those that achieve the communicative aims and uphold the reputation of the organization. How large is the impact of the following skills and knowledge types on the effectiveness of translations at your organization? 1=

minimal; 5 = enormous; N/A = not applicable because the skill or knowledge is not required.”

- To improve the construct validity of the questionnaires, the rating scale was adjusted in both cases to a single five-point Likert scale that allowed the option of a neutral mid-point value (3). The “not applicable” option was retained.
- New questions were included in both questionnaires to identify the SLs and TLs of the translation work under consideration. This variable had not been present in the first pilot, but was thought to possibly have a bearing on the rating of certain language- and culture-related skills and knowledge.
- The question on in-house training modalities was removed as it was likely to arouse negative feelings about shortcomings, and because the information, if required, could probably be obtained by another means, such as asking the head of service.
- The questions on the expectations of new recruits were removed from the recruits questionnaire, where they had aroused negative feelings, and re-phrased as one open-ended question in the impact questionnaire. The question was thought worth asking because the answers might not be as clear-cut as they should be about on-the job training.
- Questions on document processing at the organization were inserted into the recruits questionnaire.
- Questions on how long the respondent had been working with the organization were removed, as the questionnaires were becoming rather long and the information was considered nonessential.
- The base list was modified as outlined below. Many of these changes, unless otherwise stated, were based on a closer examination of the list, clearer thinking (the benefit of distance and hindsight), and the greater understanding gained in the interviews after the first pilot of how some of the definitions might be understood by respondents:
  - “Mining information” was split into three: “tracking down sources of information to check facts”, “tracking down sources to obtain a better grasp of the thematic aspects of a text (understand the topic)”; and “judging the reliability of information sources”. This breakdown was motivated by the direct observation of the work of the translators at the IGO used in the pilot.
  - “Working with reference documents” was rephrased as “mining background documents for accepted phrasing and terminology” to describe the activity in

more specific terms, as the original version could have been understood to refer just to checking the number of the next assignment, as pointed out by a translator interviewed in the first pilot.

- “Working with Excel” and “working with PowerPoint” were combined since both referred to working with text in formats other than Word documents.
- “Formatting” and “reproducing the formatting of the ST” were considered included in the ability to “handle more than basic Word functions” and therefore removed.
- “Inputting changes” was considered to be already covered by “following complicated instructions about what to do with a text” and was therefore removed.
- “Not making stylistic errors” and “improving on the style of the original” were merged into “writing elegantly regardless of the style of the original”.
- “Detecting logic errors” was considered covered by “detecting inconsistencies” and therefore removed.
- “Finding terminology” was narrowed down to “working with electronic terminology tools”, as non-electronic searches are covered in the other research skills.
- All the features listed as examples of what was involved in “improving on the content of the ST” (redundancies, unintended ambiguities, contradictions, unexplained acronyms, misleading headings, etc.) were ones that would affect the coherence of the TT, and the skills definition was rephrased accordingly.
- The negative references to introducing “unwarranted additions” and “errors of fact (e.g. names that have not been checked)” were removed as they were considered covered by the new, positive reference to ensuring the coherence of the TT.
- “Producing accurate translations” was removed because it involves a wide range and varying combination of skills hopefully included in the list or identified in the survey. In the context of a skills survey, it seemed to make more sense to approach accuracy through the skills that might contribute to it, such as the ability to find the most appropriate word or word combination and the ability to capture every nuance of the ST in the translation.
- “The ability to recast sentences (say the same thing in several different ways)” was added as being an identified skill (Pym 2003). It had been covered by the ability to input changes in the first list but it was thought better to isolate the skill since it is required in all translation activities, not just when inputting changes.

- “Word choice” was rephrased as the ability to “find the most appropriate word or word combination” to ensure that collocation and phrasing skills were also covered.
- “Knowledge of TL culture” was added because if SC knowledge is important for comprehension of a message in the SL, TC knowledge may be important for transmitting a message in the TL. It should have been included in the first version of the questionnaires.
- “Knowledge of SL and TL varieties” were added in response to comments made in interview by the Spanish-English translators in the first pilot about the challenges posed by texts written by authors from different Latin American countries and about the need to produce TTs that were not written too obviously in British or American English, a challenge also pointed out by Sekel (2008).
- “The ability to detect mathematical errors” was separated from the other logic skills, since it is a non-linguistic ability, yet it was retained to see if it was important or required. The translators in the pilot IGO, which handled mainly economic texts, said that translators were expected to notice if the sectorial distribution of investment, expressed as percentages, for example, did not add up to 100.
- The questionnaires were prepared in electronic format so that they could be answered online and the results more easily filtered and analyzed. The links were sent out in a formal e-mail.

The revised impact questionnaire is presented in Appendix 4 and the revised recruits questionnaire is presented in Appendix 5.

### 3.3.2 *The respondents*

The respondents were all employed by the same organization that served as the sample base for the first pilot. This offered two main advantages: (1) convenience, insofar as the researcher had ready access to the respondents and already had much of the information needed to contextualize findings, such as information on testing practice, training possibilities, expectations of new recruits, revision styles and document processing norms, and (2) the opportunity to compare findings with those of the first pilot.

This time the survey sample for the impact questionnaire was expanded to include not only more revisers, but also translators. The revisers in the first pilot had seemed to report on

their experience as translators as much as on their experience of revising the work of other translators. The idea was to explore any difference in the answers. Being able to include translators would furthermore make it possible to expand the sample base considerably. Also in the second pilot, participants included not only members of the English unit, but also of the Spanish unit, in order to explore the influence of the target language of the translation work, which had not been taken into consideration in the first pilot. Participation was as follows:

- Six revisers (including the two who participated in the first pilot) completed the recruits questionnaire
- Four of the six revisers (including the two who had participated in the first pilot) also completed the impact questionnaire.
- Additionally three translators completed the impact questionnaire.

### *3.3.3 Assumptions*

It was assumed that respondents would answer honestly, read and follow instructions about what to base their answers on, understand the questions and know how to use a five-point Likert scale.

### *3.3.4 Procedure*

The electronic links to the questionnaires and their covering letter were sent out by e-mail on 9 November 2009. Participants were given four days to complete the questionnaire. A brief reminder was sent out on the last day, and all results were in by the next morning. Brief follow-up interviews lasting approximately five minutes were held with each participant to ascertain whether they had had any difficulties completing the questionnaire and to collect comments. Participants were also consulted by e-mail, telephone and in person during the course of the analysis.

### *3.3.5 Results of the second pilot of the impact questionnaire*

#### *3.3.5.1 Profile of the respondents*

Seven people in total answered the questionnaire: three revisers and four translators. Four respondents were from the English unit and three from the Spanish unit. Five people answered on the basis of in-house work, and two on the basis of external work.

### *3.3.5.2 The impact ratings*

The ratings awarded to the items on the base list presented in the second pilot of the impact questionnaire are presented in Table 3.1. For the sake of space, abbreviated labels for the skills and knowledge types are used in tables and figures and sometimes in the text of this section. For the exact definitions used in the questionnaires in the second pilot and the corresponding abbreviations, see Appendix 11.

Table 3.1. Impact ratings awarded for each skill and knowledge type: distribution of responses, average ratings and variance, ordered by mean rating\*

Skill or knowledge type	Rating						Mean rating	No. of responses	Variance
	1	2	3	4	5	N/A			
Knowledge of the SL	0	0	0	0	7	0	5.00	7	0.00
Knowledge of TL grammar	0	0	0	0	7	0	5.00	7	0.00
Convey the ST message clearly	0	0	0	0	7	0	5.00	7	0.00
Adhere to in-house style conventions	0	0	0	0	7	0	5.00	7	0.00
Ensure the completeness of the TT	0	0	0	0	7	0	5.00	7	0.00
Mine reference material for phrasing	0	0	0	0	7	0	5.00	7	0.00
An extensive TL vocabulary	0	0	0	1	6	0	4.86	7	0.12
Knowledge of TL spelling rules	0	0	0	1	6	0	4.86	7	0.12
Produce idiomatic translations	0	0	0	1	6	0	4.86	7	0.12
Write elegantly regardless of the ST	0	0	0	1	6	0	4.86	7	0.12
Ensure the coherence of the TT	0	0	0	1	6	0	4.86	7	0.12
Track down sources to check facts	0	0	0	1	6	0	4.86	7	0.12
Achieve the right tone and register	0	0	1	0	6	0	4.71	7	0.49
Understand complex topics	0	0	0	2	5	0	4.71	7	0.20
Work out the meaning of obscure ST	0	0	0	2	5	0	4.71	7	0.20
Detect inconsistencies, etc.	0	0	0	2	5	0	4.71	7	0.20
Knowledge of TL punctuation rules	0	0	0	2	5	0	4.71	7	0.20
Produce translations that flow smoothly	0	0	0	2	5	0	4.71	7	0.20
Find the best words	0	0	0	2	5	0	4.71	7	0.20
Recast sentences in the TL	0	0	0	2	5	0	4.71	7	0.20
Track down sources to understand topic	0	0	0	2	5	0	4.71	7	0.20
Explain translation decisions or problems	0	0	0	2	5	0	4.71	7	0.20
Capture every nuance of the ST	0	0	1	1	5	0	4.57	7	0.53
Work with electronic terminology tools	0	0	1	1	5	0	4.57	7	0.53
Knowledge of SL culture(s)	0	0	0	3	4	0	4.57	7	0.24
Translate fast and under pressure	0	0	1	2	4	0	4.43	7	0.53
Tailor language to the readers' needs	0	0	0	4	3	0	4.43	7	0.24
Judge the reliability of sources	0	0	0	4	3	0	4.43	7	0.24
Master new subjects quickly	0	0	2	1	4	0	4.29	7	0.78
Knowledge of TL varieties	0	1	1	1	4	0	4.14	7	1.27
Knowledge of the subject	0	0	2	2	3	0	4.14	7	0.69
Knowledge of TL cultures	0	0	1	4	2	0	4.14	7	0.41
Type accurately and fast	0	1	1	2	3	0	4.00	7	1.14
Handle more than basic Word functions	0	1	1	3	2	0	3.86	7	0.98
Knowledge of the organization	0	0	3	2	2	0	3.86	7	0.69
Detect mathematical errors in the ST	0	0	3	2	2	0	3.86	7	0.69
Follow complicated instructions	0	1	3	0	3	0	3.71	7	1.35
Work with Excel and/or PowerPoint	1	0	0	5	1	0	3.71	7	1.35
Knowledge of SL varieties	1	0	1	4	1	0	3.57	7	1.39
Work with translation memory software	1	1	1	4	0	0	3.14	7	1.27

\*Ratings awarded in response to the question "For the purposes of this survey, effective translations are those that achieve the communicative aims and uphold the reputation of the organization. How large is the impact of the following skills and knowledge types on the effectiveness of translations at your organization? 1= minimal; 5 = enormous; N/A = not applicable because the skill or knowledge is not required."

No skill or knowledge type was reported as not required (N/A). All the skills and knowledge types were rated 3 or above, suggesting none of them was considered even remotely unimportant. The items rated most highly were: knowledge of the SL, knowledge of the finer points of TL grammar, writing and conveying the ST message clearly, adhering to in-house style conventions, ensuring the completeness of the TT and mining reference

material for phrasing and terminology. The least important knowledge or skill was making effective use of translation memory software (probably because its impact is very indirect), followed by knowledge of the different SL varieties and then, together, following complicated instructions about a text and working with Excel and/or PowerPoint.

For over half the skills and knowledge types (24/40), the variance was less than 0.25, which is low. Generally speaking, respondents agreed more on what was important than on what was not important. Average ratings range from 3.14 to 5.00, with the top 50% scoring 4.71 or above. Only one item in the top twenty, the ability to achieve the right tone and register, has a variance of over 0.25 (0.49).

### *3.3.5.3 Analysis of differences between the revisers' and the translators' answers*

The difference between the ratings awarded by the revisers and the translators are shown in Figure 3.3. The Pearson correlation coefficient between the two sets of answers is 0.57. Differences of more than 1 point on the scale only occur at the two extremes: in the case of working with Excel and/or PowerPoint and regarding knowledge of the SL – the translators think the former are far less important than revisers do, and vice-versa. This could be to do with generation-related or idiosyncratic attitudes to working with technology. There is no real tendency to rate the items higher among one group or the other. Performing a Wilcoxon rank test (as set out in Castellan and Siegel 2005) shows that there is no statistically significant difference between the two sets of ratings, with a 95% confidence level (test value of 1.34, p-value of 0.18).<sup>10</sup>

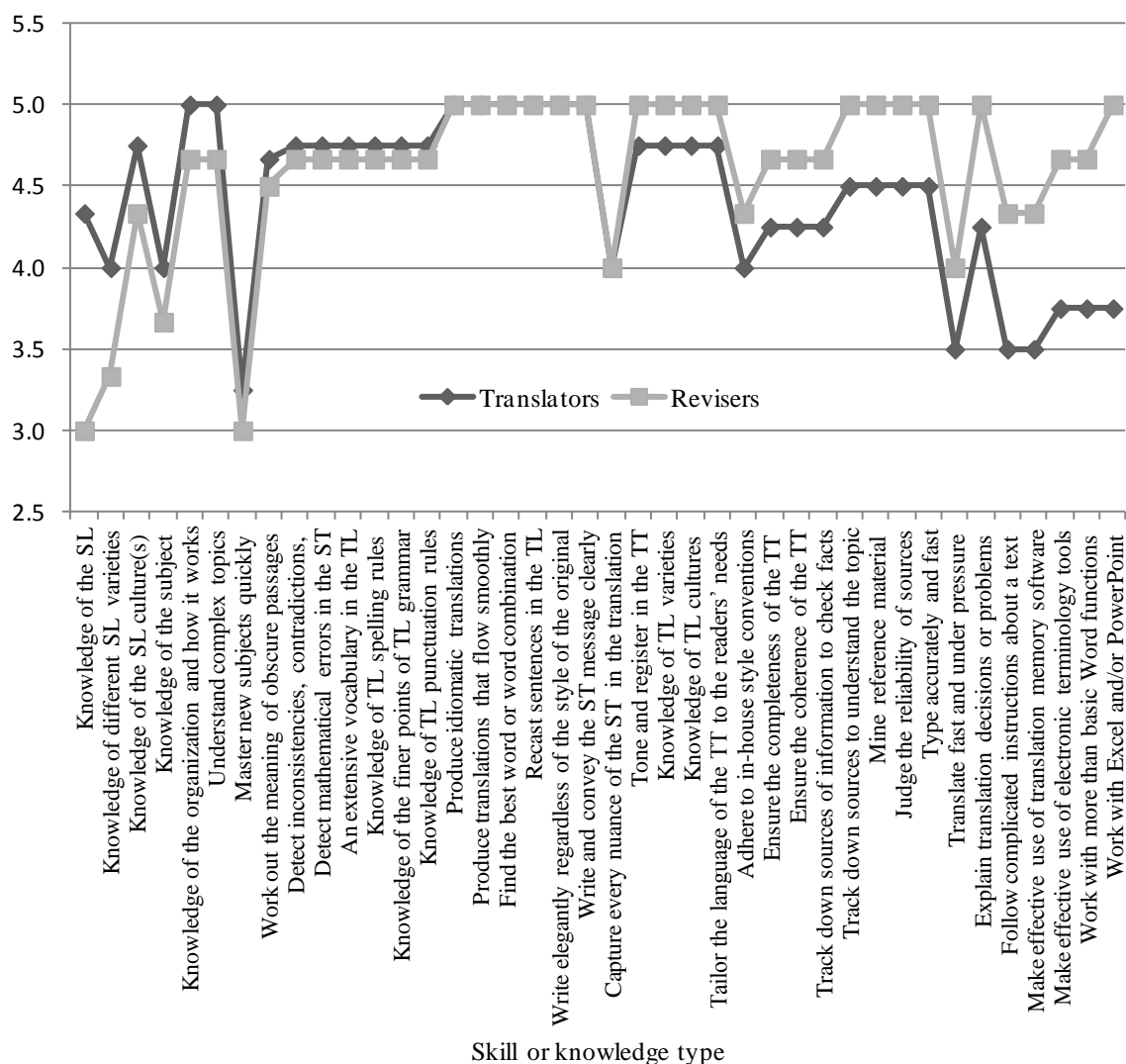
The difference between the responses of each language section and the responses in each pilot is analyzed in section 3.3.7.

---

<sup>10</sup> The Wilcoxon rank tests in the analysis described in this chapter were performed applying the criterion that a difference of one point or more on a five-point scale is indeed a difference. The Wilcoxon rank test was selected as the most appropriate method for determining the statistical significance of the difference between two sets of average ratings obtained in the second pilot, given the small size of the sample. I would like to thank Dr. Humberto Soto, Statistician of the United Nations Economic Commission for Latin America and the Caribbean, for his guidance in this regard.



Figure 3.3. Impact ratings by position in the language service\*



\*Ratings awarded in response to the question “For the purposes of this survey, effective translations are those that achieve the communicative aims and uphold the reputation of the organization. How large is the impact of the following skills and knowledge types on the effectiveness of translations at your organization? 1= minimal; 5 = enormous; N/A = not applicable because the skill or knowledge is not required.”

#### 3.3.5.4 Mention of other skills and knowledge required

One respondent listed “emotional and psychological resilience; the ability to tolerate unreasonable demands and incompetent administration” as being important; another reported “being tough-skinned”. No rating was given in either case. These psychological skills or traits are beyond the scope of this project, as discussed in the introduction. Both “psychological resilience” and possibly “the ability to tolerate unreasonable demands” are factors in timed tests, however, even if they are not measured, and many translation tests are indeed timed.

Whether this is appropriate would depend on the importance of the ability to translate fast and under pressure, something the research hoped to discover.

### *3.3.5.5 Expectations regarding when skills and knowledge should be acquired*

When asked to specify which knowledge or skills are expected to be acquired after recruitment, two of the seven respondents did not report any. When interviewed, however, they said they had not given the matter much consideration and just thought in very general terms that the organization recruited only fully competent translators. When shown the answers given by the five respondents who had reported that some skills and knowledge could be acquired on the job, they wholeheartedly agreed with what their colleagues had put. The responses of those five are listed below:

Respondent 1: “Knowledge of the organization, technical knowledge, the ability to adhere to in-house style conventions.”

Respondent 2: “In-house conventions, spelling rules and specialized computer skills can be learned after recruitment; knowledge of specialized topics (economics) can be picked up as needed.”

Respondent 3: “To adhere to in-house style conventions.”

Respondent 4: “It’s a question of degree, not of yes/no. These skills improve greatly in the course of our work for the [organization].”

Respondent 5: “Recruits, even seasoned translators, may not have a thorough knowledge of all the technical subject areas covered by the texts they are asked to translate, since they cannot be experts in all fields and, indeed, many subject areas (climate change is one example) are evolving constantly. What translators do need are excellent research skills. They are not expected to have experience in the use of translation memory software, although this would be considered an asset. In-house style conventions may be difficult to absorb all at once, but some evidence of sensitivity to these rules is expected from the start. Similarly, knowledge of Excel and PowerPoint can be learned on the job.”

This suggested that, as far as this question is concerned: (1) not everyone thinks carefully when answering, (2) the answers provided by the few who do respond may need to be taken as representative of the opinions of their less conscientious colleagues, and (3) the answers provided to this question should be combined and considered as a whole.

### *3.3.5.6 General comments on the questionnaire*

The following general comments were made:

1. “Nothing comes to mind right now, but I will comment privately if I later have any useful input.”
2. “I think the questionnaire was very comprehensive.”
3. “At my organization, it is especially important for translators to work together with other translators and with the authors of the different texts.”

It would seem that on the whole, respondents had a favorable reaction to the questionnaire. The third comment touches on the types of skills not covered by this research project, such as teamwork and other collaboration skills. It also explains the high rating of the ability to “explain translation decisions or translation problems posed by the ST (to authors, users or revisers)”, which was (but probably should not have been) on the list presented in the questionnaire. A comparison of the responses of the Spanish and English units is presented, together with the analysis of other variables, in section 3.3.7, after the presentation of the results of the recruits questionnaire

### *3.3.6 Results of the recruits questionnaire*

#### *3.3.6.1 Profile of the respondents*

Six people completed the questionnaire: three revisers from the Spanish Unit and three revisers from the English Unit. One of the Spanish revisers and one of the English revisers based their answers on external translators, the others on in-house staff. The influence of the in-house/external variable is examined in section 3.3.7.2.1.

#### *3.3.6.2 The frequency ratings*

The ratings awarded in the recruits questionnaire are presented in Table 3.2. No skills or knowledge types were left unrated, and none were reported as not required (N/A). The table shows that, on average, a lack of the following skills and knowledge most frequently accounts for the mistakes that revisers have to correct: adhering to in-house style conventions; knowledge of the subject; capturing every nuance of the ST in the translation; detecting inconsistencies, contradictions, nonsense, etc. in the ST; producing translations that flow smoothly even when the ST does not; and working out the meaning of obscure ST passages. All these items were rated above 3 on the five-point frequency scale. No skills or

knowledge types were rated 4 or 5 (almost always lacking). Typing and spelling mistakes are very rare (rated close to 1), as are mistakes caused by the second-lowest rated skills: inadequate knowledge of language varieties and TL vocabulary, and the ability to use electronic terminology tools, which were each awarded 1.5 points on the scale.

The variance values show that respondents rated many skills and knowledge types very differently. Given that the ratings reflect the different experiences of different revisers with different new recruits working on different types of work, this was not totally unexpected. A large sample base would be needed to obtain mean values that would represent tendencies better. The influence of the TL and the in-house/external work factor would also need to be taken into account. These are analyzed in section 3.3.7 below.

Table 3.2. Frequency ratings awarded for each skill and knowledge type: distribution of responses, average ratings and variance, ordered by mean rating\*

Skill or knowledge rated	Rating awarded					N/A	Rating average	Response count	Variance
	1	2	3	4	5				
Adhere to in-house style conventions	0	1	1	3	1	0	3.67	6	0.89
Knowledge of the subject	0	0	4	1	1	0	3.50	6	0.58
Capture every nuance of the ST	0	2	1	2	1	0	3.33	6	1.22
Detect inconsistencies, etc.	1	1	1	1	2	0	3.33	6	2.22
Produce translations that flow smoothly	0	2	1	3	0	0	3.17	6	0.81
Work out the meaning of obscure ST	1	1	1	2	1	0	3.17	6	1.81
Write elegantly regardless of the ST	0	2	2	2	0	0	3.00	6	0.67
Recast sentences in the TL	0	2	3	0	1	0	3.00	6	1.00
Ensure the coherence of the TT	0	3	1	1	1	0	3.00	6	1.33
Find the best words	0	3	1	1	1	0	3.00	6	1.33
Convey the ST message clearly	1	2	0	2	1	0	3.00	6	2.00
Produce idiomatic translations	1	1	2	2	0	0	2.83	6	1.14
Knowledge of the organization	1	1	3	1	0	0	2.67	6	0.89
Translate fast and under pressure	1	2	1	2	0	0	2.67	6	1.22
Knowledge of the SL	1	3	0	1	1	0	2.67	6	1.89
Judge the reliability of sources	0	4	1	1	0	0	2.50	6	0.58
Explain translation decisions or problems	1	2	2	1	0	0	2.50	6	0.92
Knowledge of TL punctuation rules	1	2	2	1	0	0	2.50	6	0.92
Knowledge of TL grammar	1	2	2	1	0	0	2.50	6	0.92
Understand complex topics	2	1	2	0	1	0	2.50	6	1.92
Track down sources to understand topic	1	2	3	0	0	0	2.33	6	0.56
Detect mathematical errors in the ST	0	5	0	1	0	0	2.33	6	0.56
Track down sources to check facts	2	1	2	1	0	0	2.33	6	1.22
Master new subjects quickly	2	2	0	2	0	0	2.33	6	1.56
Ensure the coherence of the TT	2	2	0	2	0	0	2.33	6	1.56
Mine reference material for phrasing	1	4	1	0	0	0	2.00	6	0.33
Knowledge of SL culture(s)	1	4	1	0	0	0	2.00	6	0.33
Work with translation memory software	2	3	0	1	0	0	2.00	6	1.00
Achieve the right tone and register	3	0	3	0	0	0	2.00	6	1.00
Work with Excel and/or PowerPoint	3	1	2	0	0	0	1.83	6	0.81
Follow complicated instructions	3	1	2	0	0	0	1.83	6	0.81
Handle more than basic Word functions	4	0	1	1	0	0	1.83	6	1.47
Tailor language to the readers' needs	2	4	0	0	0	0	1.67	6	0.22
Knowledge of TL cultures	4	0	2	0	0	0	1.67	6	0.89
Work with electronic terminology tools	3	3	0	0	0	0	1.50	6	0.25
Knowledge of TL varieties	3	3	0	0	0	0	1.50	6	0.25
Knowledge of SL varieties	3	3	0	0	0	0	1.50	6	0.25
An extensive TL vocabulary	5	0	0	1	0	0	1.50	6	1.25
Knowledge of TL spelling rules	4	2	0	0	0	0	1.33	6	0.22
Type accurately and fast	5	0	1	0	0	0	1.33	6	0.56

\*Ratings awarded in response to the question "Please think about the mistakes you usually correct when going over translations by new recruits. How often do you think those mistakes are due to a lack of the following things? 1 = Almost never; 5 = Almost always; N/A = Not applicable because the skill or knowledge is not required of new recruits."

### 3.3.6.3 Other skills reported as lacking and needed

Two participants answered the question "Are there other skills or types of knowledge that new recruits lack and should have at your organization?" Neither had answered the impact questionnaire, which is why they were not mentioned in section 3.3.5.4. One stated "basic numeracy" and rated it 2 on the scale. The other stated "the ability to read between the lines

and realize (from the overall context and history of the organization) the actual meaning of a complex sentence (although this cannot really be expected of new recruits)” and awarded it a frequency rating of 4. In other words, the lack of this skill was thought by this reviser to often account for the mistakes made. The first would seem to justify the inclusion of the question about detecting mathematical errors, although it should be borne in mind that the translators at this IGO mainly translate texts on economics and financial issues. The second ability is possibly covered by the nuance item.

#### *3.3.6.4 Document processing and translators' responsibilities*

At this organization, texts are sometimes or often edited prior to being translated; they are rarely pre-translated, nearly always referenced, sometimes revised fully and always sent to the author for approval. Translators are always responsible for adhering to in-house style conventions, nearly always responsible for spelling and hardly ever responsible for the formatting of their translations.

Two comments were made in this section: “Re the question on how often texts are fully revised: if this is about the work of relative newcomers to [the organization], the answer should be ‘Always’, but since we’re ridiculously understaffed, it often doesn't get done”; and “Our bilingual text-processing unit takes care of the formatting.”

#### *3.3.6.5 Comments made in the recruits questionnaire*

In the space provided for observations about the skills and knowledge that new recruits lack, one reviser commented:

Negotiating through the editorial conventions of [the organization] and becoming familiar with the host of bodies and organizations is like wandering through a veritable labyrinth and I think it takes a long time to become comfortable with all the rules. It's an ongoing learning process. So new recruits are up against a formidable task, especially since standards are very high in the English service.

This confirmed that to contextualize the findings and identify their implications for recruitment testing, it is necessary to identify the skills and knowledge that are expected to be acquired after recruitment, as is done in the impact questionnaire.

### *3.3.6.6 Results of the semi-formal follow-up interviews with respondents*

All participants said that they had no problem following the instructions and understanding the questions, although some felt that a few of the skills descriptions could be improved to make them less open to interpretation. Specifically, two respondents suggested that some reference to the need to maintain quality standards be included when asking about “the ability to translate fast and under pressure (e.g. to meet tight deadlines)”. Two other participants commented that they had been unsure whether “the ability to find the most appropriate word or word combination” referred to a research skill or a language production skill. This ambiguity may have distorted the results for this particular ability. One respondent suggested clarifying or using a synonym for “nuance” as the word was not that common and might not be known by all non-native English-speaking respondents. On the whole, participants felt the questionnaire was thorough but not too long.

### *3.3.7 Analysis of the results of the second pilot*

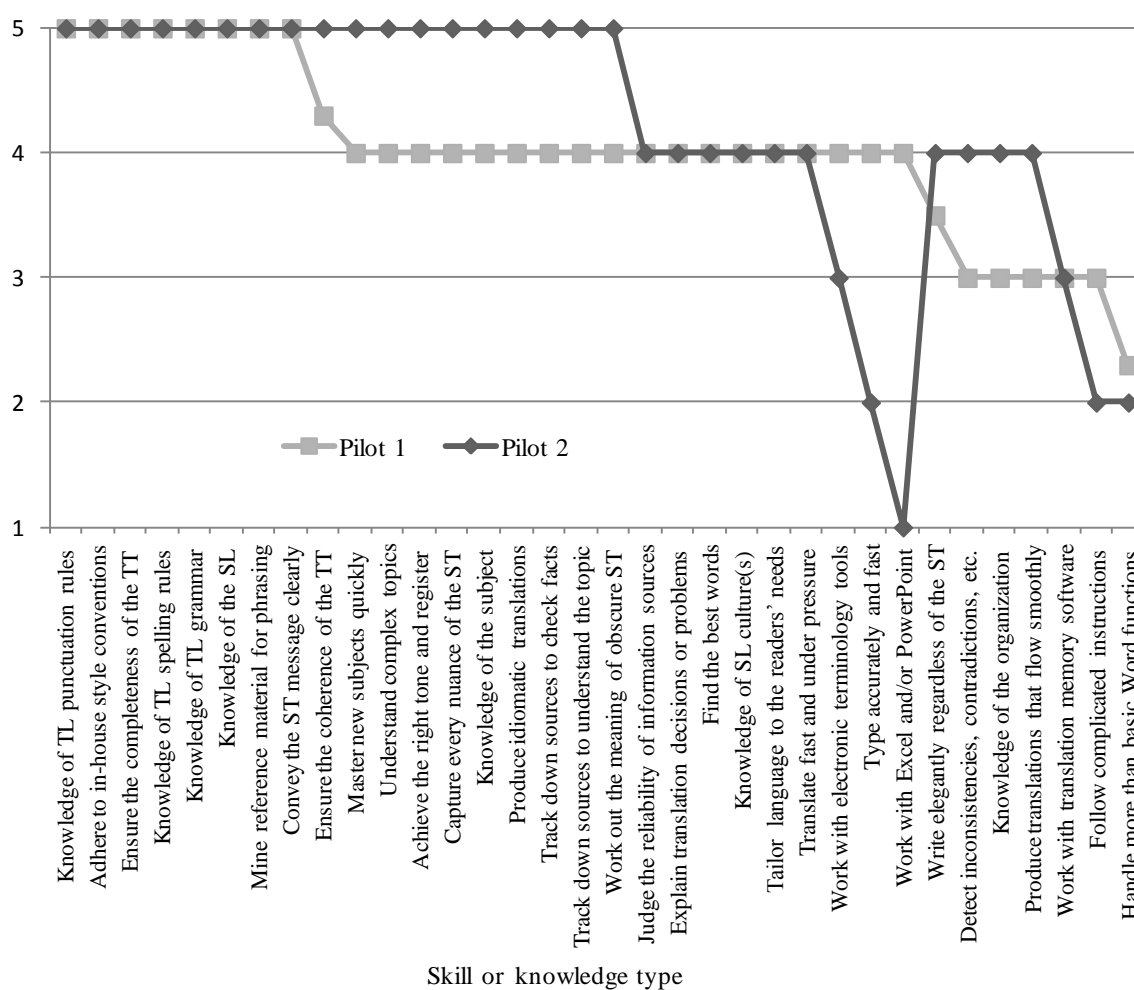
#### *3.3.7.1 Comparison of the ratings awarded by the revisers who participated in both pilots*

The smallest (one-point) shifts between the findings of each pilot can possibly be explained by the modifications made to improve the construct validity of the questionnaires (i.e. the rephrasing of instructions and the introduction of single five-point scales). The analysis therefore focuses on larger differences. Only 29 of the ratings are in fact wholly comparable inasmuch as the definition of the skills or knowledge was practically identical in both pilots. Another five sets of ratings were made comparable by taking averages for the skills that were combined as explained in section 3.3.1. Skills without equivalents were left out of the comparison (e.g. recasting sentences, accuracy, knowledge of TL cultures and language varieties). The scales used in both pilots have been converted to numerical ones for the sake of comparison, although there is no guarantee that the scales used in the first pilot were understood the same way as in the second.

In the impact questionnaire, ratings fully coincided between the two pilots 14/34 times in the case of Reviser A, and 18/34 times in the case of Reviser B, as shown in Figures 3.4 and 3.5. The only two differences of over one point were in Reviser A’s results and refer to working with Excel and PowerPoint and typing accurately and fast, which were rated two points lower in the second pilot. When consulted, Reviser A explained that she had lowered the rating of typing skills the second time around because, although they were important insofar as they were often required, they did not have that much impact on the effectiveness

of the organization's translations as defined in the second pilot. The Excel and PowerPoint skills had been lowered because, since the first pilot, she had discovered that the text-processing unit was now working with a new software program that meant it was in fact more convenient if translated text was prepared in Word. Two tentative conclusions were drawn from this: some respondents pay close attention to the wording of instructions, and narrowing down the definition of importance to "impact on effectiveness" seemed to help ensure that respondents would award the ratings on the same basis (which would make them more comparable). Based on the conversation with Reviser A, it seemed that the questionnaires would provide a fair reflection of the situation in the organization. This augured well for the validity of the survey findings.

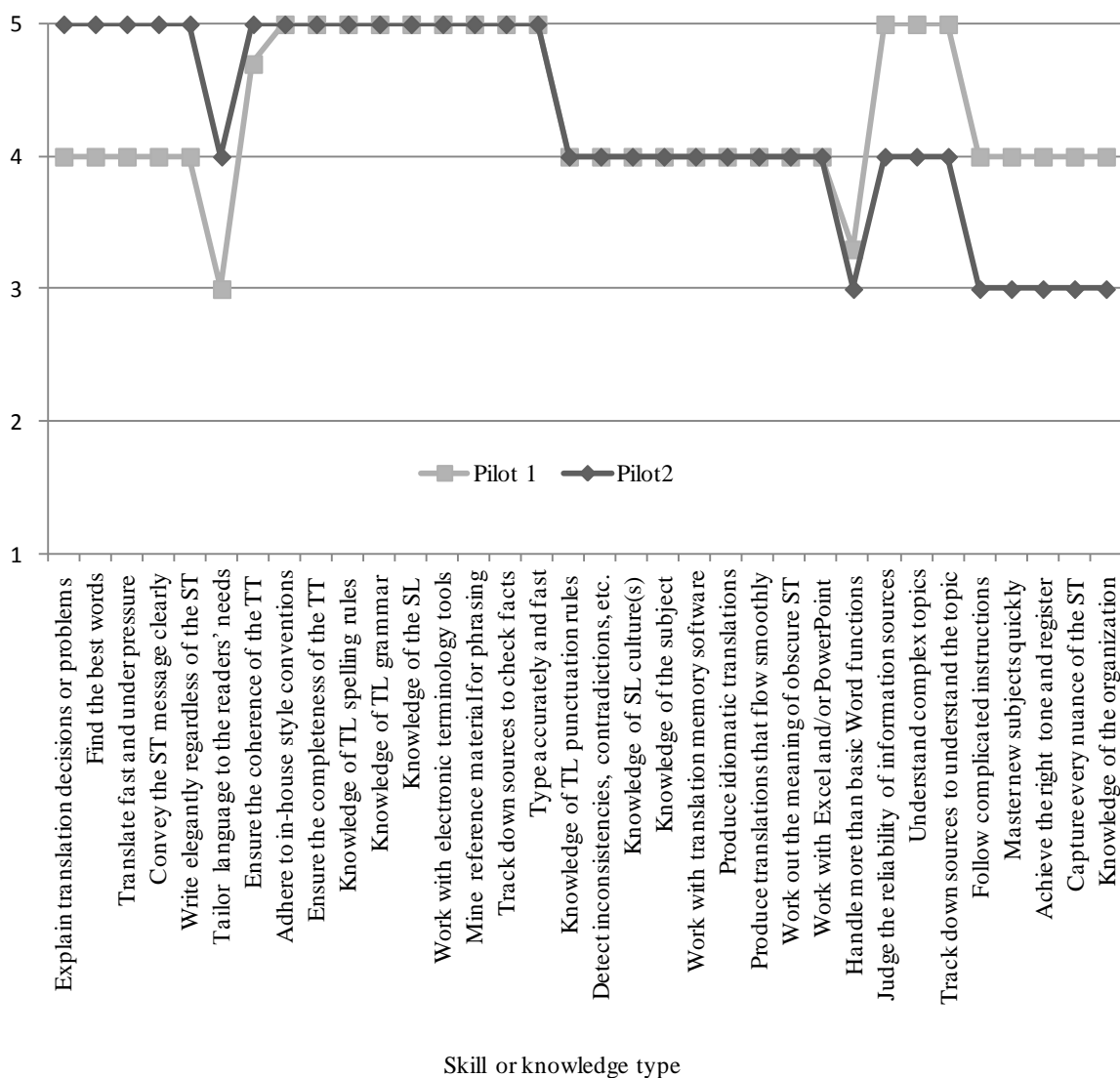
Figure 3.4. Reviser A: impact ratings in pilot 1 and pilot 2\*



\* The ratings that are not whole numbers correspond to skills that were not identically defined in the two pilots. They therefore reflect the averages calculated according to the adjustments carried out to make the skills comparable, as described above.



Figure 3.5. Reviser B: impact ratings in pilot 1 and pilot 2\*

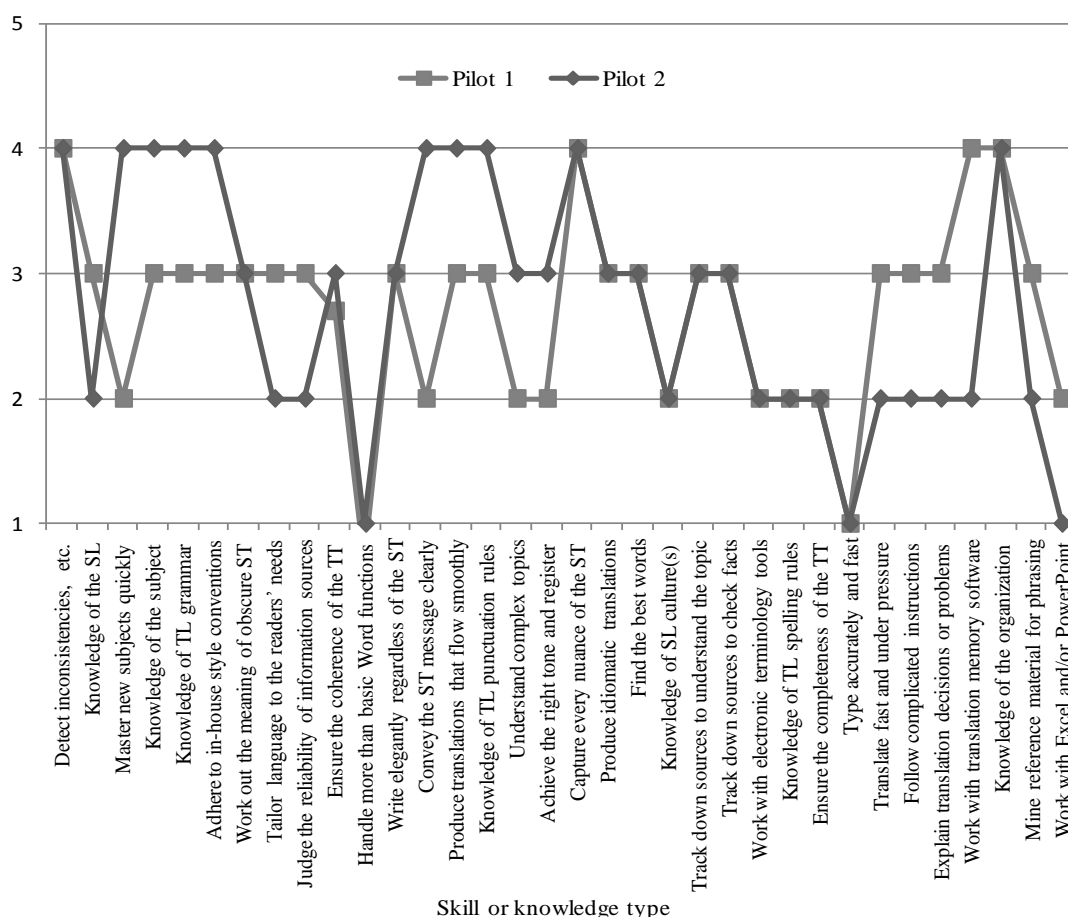


\* The ratings that are not whole numbers correspond to skills that were not identically defined in the two pilots. They therefore reflect the averages calculated according to the adjustments carried out to make the skills comparable, as described above.

The differences between the ratings awarded by Reviser A and Reviser B in the recruits questionnaires in the two pilots was slightly larger than in the impact questionnaire. Ratings fully coincided 10/31 times in the case of Reviser A and 11/31 times in the case of Reviser B, as shown in Figures 3.6 and 3.7 respectively. There were differences of more than one point with reference to three skills in the case of Reviser A (mastering new subjects, writing elegantly and working with translation memory software, which were two points lower in each case); and with reference to knowledge of the organization, mining reference material

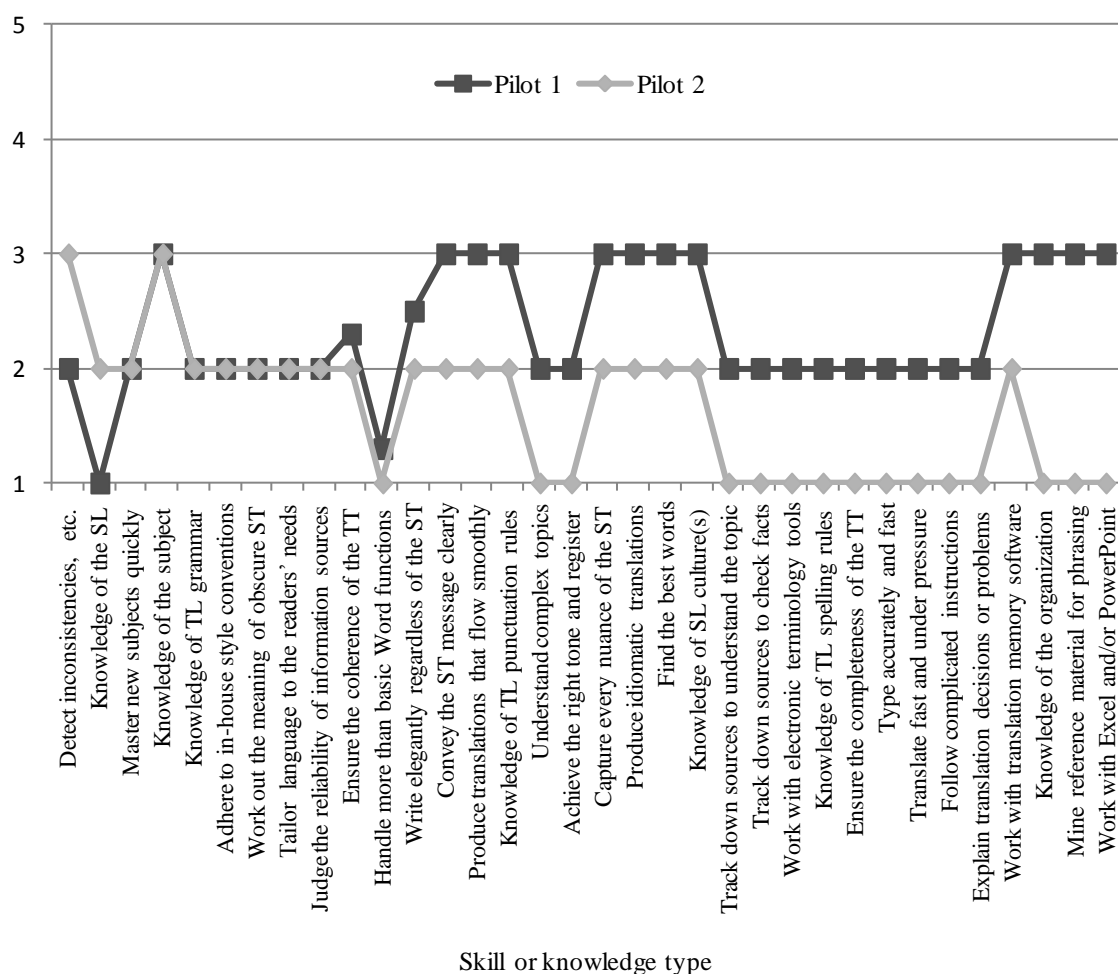
and working with Excel and PowerPoint in the case of Reviser B. Reviser A explained that the differences in her case reflected the influence of having hired a new recruit since the first pilot who had strong computer skills (especially the ability to take full advantage of translation memory software) but problems writing clearly and coming to grips with new subjects. Reviser B said that the differences in his case stemmed from the narrowing down of the definition of “new recruit” (to translators who have been working with the organization for one year) because it excluded several of the translators on whose work he had based his ratings in the first pilot. Here again we had evidence that respondents followed instructions and that narrowing down definitions of terms as had been done in the second pilot had enhanced the validity and the comparability of answers.

Figure 3.6. Reviser A: frequency ratings in pilot 1 and pilot 2\*



\* The ratings that are not whole numbers correspond to skills that were not identically defined in the two pilots. They therefore reflect the averages calculated according to the adjustments carried out to make the skills comparable, as described above.

Figure 3.7. Reviser B: frequency ratings in pilot 1 and pilot 2\*



\* The ratings that are not whole numbers correspond to skills that were not identically defined in the two pilots. They therefore reflect the averages calculated according to the adjustments carried out to make the skills comparable, as described above.

The fact that the changes in Reviser A's and Reviser B's answers between one pilot and the next are explicable and respond either to real changes in circumstances or to the structural changes made to the questionnaires suggest that the findings obtained by these measuring instruments are not completely arbitrary and seem to reflect real situations. The results are based, however, on just two individuals. For meaningful observations to be made about the ways in which new IGO translators meet requirements in general, the survey base would have to be expanded considerably. This would mean arranging to have as many revisers from as many IGOs as possible participate in the study.

### 3.3.7.2 The influence of different variables on the ratings awarded in the two questionnaires

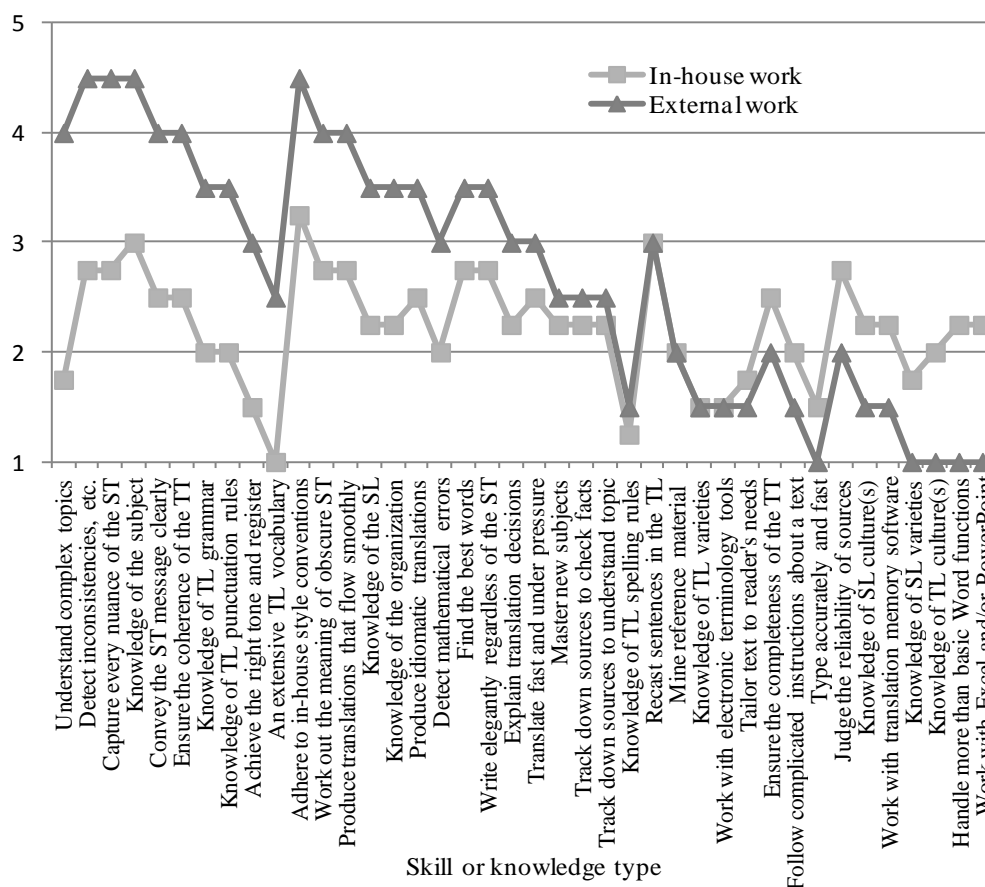
The findings of the first pilot had suggested that the in-house/external variable had a direct bearing on ratings. Another variable was introduced in the second pilot: the target language of

the translators' work. The influence of these variables on the ratings in each questionnaire is analyzed below.

*3.3.7.2.1 The influence of the in-house/external work variable.* In the recruits questionnaire, four revisers based their answers on in-house work and two on external work. The average ratings for each group are presented in Figure 3.8. Almost half (17/40) of the average ratings differ by more than one point on a five-point scale, and only four average ratings coincide exactly. Performing a Wilcoxon rank test and applying the criterion that a difference of one point or more on a five-point scale is indeed a difference, shows that the difference between the two sets of ratings is significantly different with a 95% confidence level (test value of 3.62; p-value of 0.0003).

The only three skills or knowledge types whose absence is found to account for mistakes more often in in-house work than in externally produced work are “knowledge of TL cultures”, “ability to work with more than basic Word functions” and “ability to work with PowerPoint and Excel”. In the case of the two computer skills, external translators are rarely asked to use them, which explains why the skills are rated as almost never accounting for mistakes (1 on the scale). The difference in the rating of knowledge of TL cultures could, in the opinions of the heads of unit, be attributed to the fact that only in-house staff translate the highly sensitive political documents, such as speeches and official letters, in which knowledge of the TL cultures most comes into play.

Figure 3.8. Frequency ratings for selected skills and knowledge, by in-house/external work\*



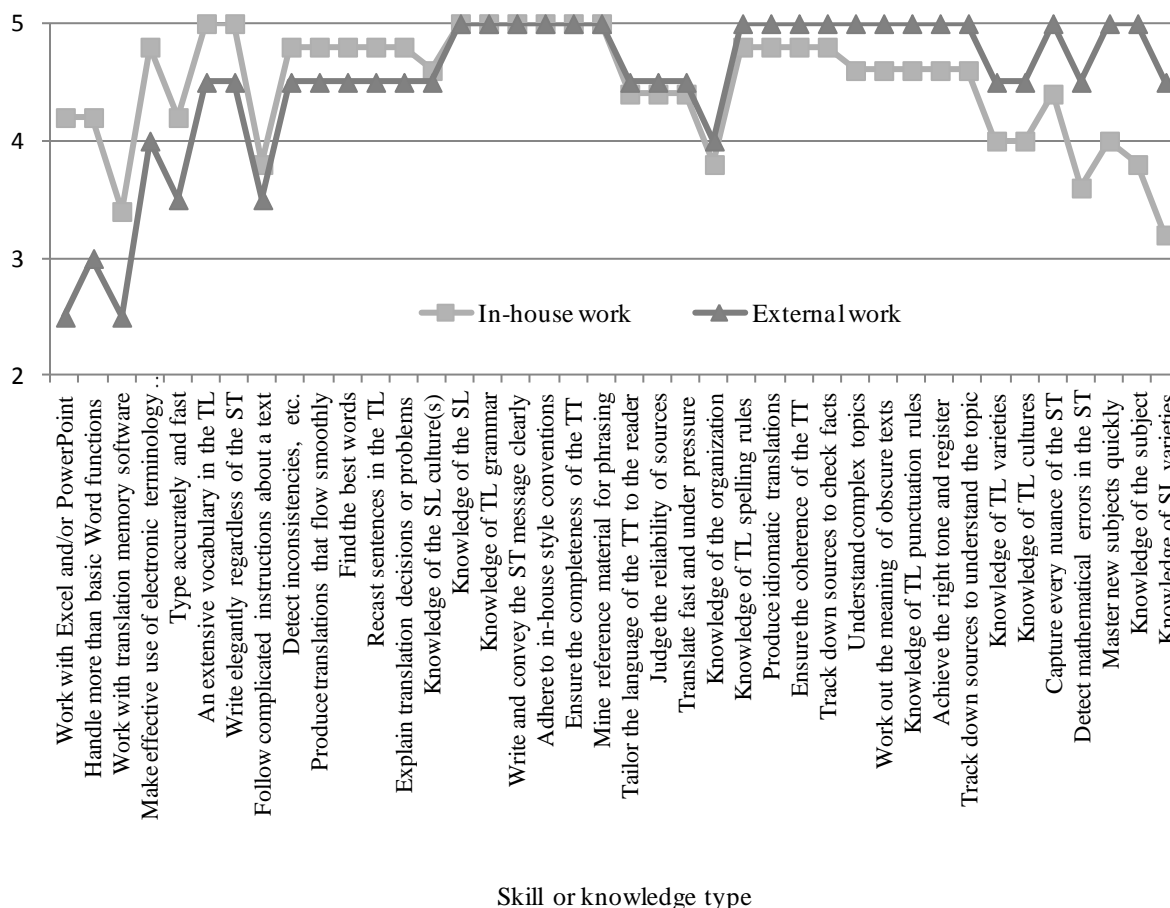
\* Simple averages of ratings awarded on a five-point scale.

Separating answers by in-house/external work shows that this variable also has a bearing on the impact ratings. Four revisers based their answers on in-house work and two on external work in the impact questionnaire, and the average ratings awarded by each group are shown in Figure 3.9. The difference is not only not as great as in the case of the frequency ratings, but not statistically significant for a 95% confidence level since the p-value is above 0.05, albeit only just (test value for the Wilcoxon rank test is 1.83;  $p=0.068$ ).

The average impact ratings differ by more than one point in the case of five skills and knowledge: “mastering new subjects and knowledge of the subject” and “SL varieties” are rated as having a larger impact on external work, and “working with Excel” and “working with more than basic Word functions” are rated as having a larger impact on in-house work. This mirrors the two major differences regarding the expectations of in-house and external translators in this organization: the IGO actively recruits external translators who already have knowledge of the subject handled by the organization. In-house translators are expected to acquire subject knowledge on the job, and externals are hardly ever given documents in

PowerPoint, which are usually rush jobs and therefore done in-house. Fiddly jobs, such as incorporating authors' changes or bringing translations into line with new versions, also belong exclusively to the in-house domain.

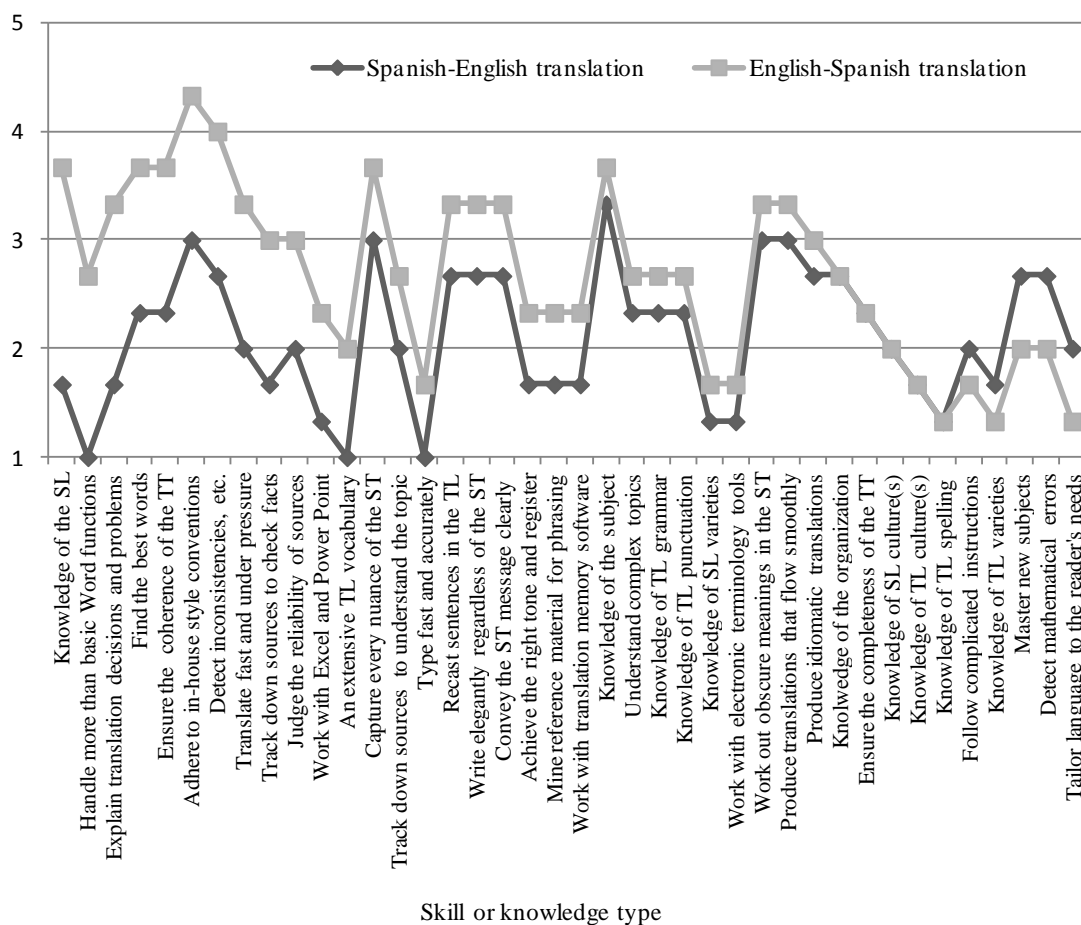
Figure 3.9. Impact ratings for selected skills and knowledge, by in-house/external work\*



\* Simple averages of ratings made on a scale of 1-5.

3.3.7.2.2 *The influence of the language service.* Expanding the survey sample in the second pilot introduced a new variable, the language service, which is determined by the target language of the translations. The differences in the average frequency ratings for the two language services surveyed are presented in Figure 3.10. A Wilcoxon rank test performed using the same criterion as in the analyses above shows that the difference between the two sets of findings is statistically significant but less so than in the case of the difference between ratings for in-house and external work (Wilcoxon test value of 2.66 versus 3.62, with p-values of 0.007 versus 0.0003, respectively). This suggested that results would need to be broken down by language service.

Figure 3.10. Frequency ratings for selected skills and knowledge, by target language\*

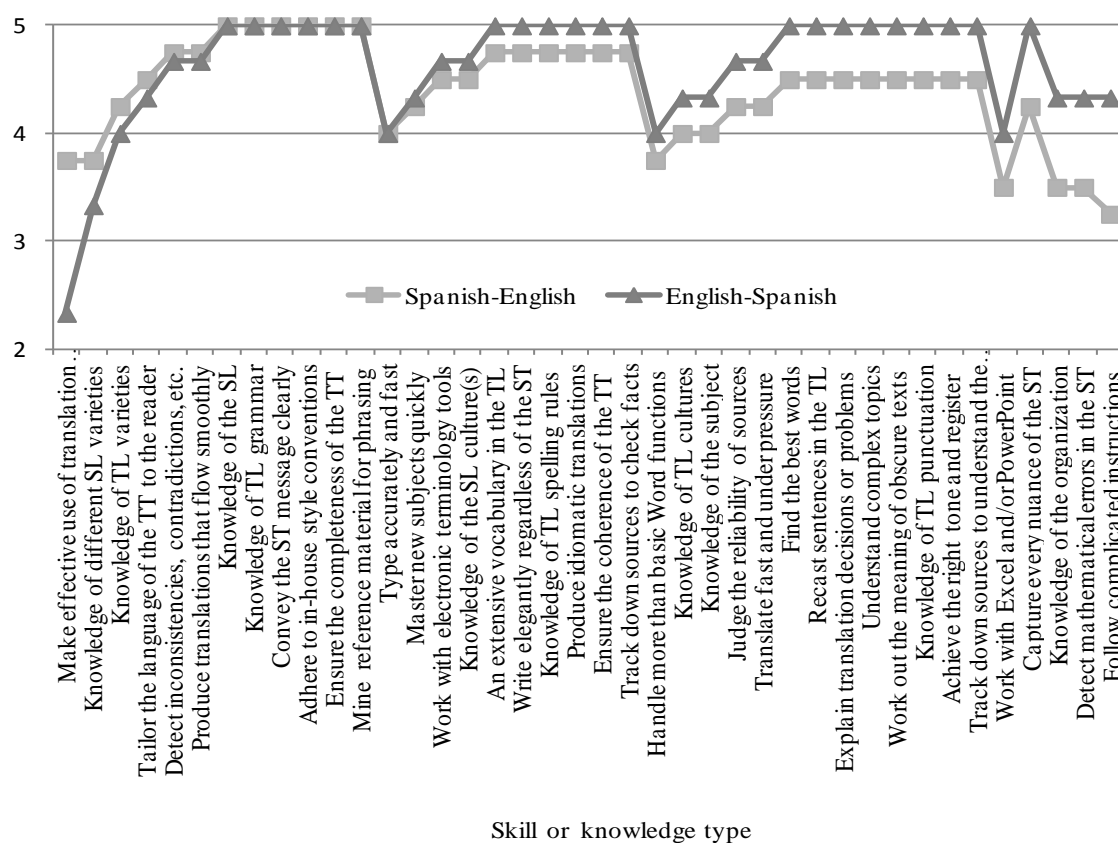


\* Frequency with which the mistakes made by new recruits can be attributed to a lack of the skill or knowledge. Average ratings awarded on a five-point scale, where 1=almost never and 5= almost always.

For one quarter of the skills and knowledge types (those on the left in Figure 3.10), the difference in the average rating is one point or more. For three quarters of the skills, the Spanish revisers awarded higher ratings than their counterparts in the English unit. These differences could be explained by the presence in such a small survey sample of one particularly weak translator or one particularly negative rater in the Spanish unit, or one generous rater in the English unit. Nevertheless, it should also be borne in mind that, in this organization, the Spanish translators do more editing work than translation per se, and most of the skills rated as more frequently missing among the Spanish translators are major editing skills: explaining decisions and problems posed by the ST to authors, finding the most appropriate word combinations, ensuring coherence, tracking down sources for facts, detecting inconsistencies, adhering to in-house style and having a broad TL vocabulary. This

feature may of course be unique to this IGO. A broader survey might reveal the importance of editing skills in IGO translation in general.

Figure 3.11. Impact ratings for selected skills and knowledge, by target language\*



\* Impact of the skill or knowledge on the effectiveness of translations. Average ratings awarded on a five-point scale, where 1=minimal impact and 5= enormous impact.

In the case of the impact ratings, however, performing the same Wilcoxon rank test reveals that the difference between the average ratings awarded by the Spanish unit and the English unit was not statistically significant (Wilcoxon test value of 1.34 and  $p=0.18$ ). The differences between the two units' impact ratings are shown in Figure 3.11. Only two mean ratings actually differ by more than one point: "making effective use of translation memory software", which is considered more important in the English unit, and "being able to follow complicated instructions about what to do with a text", which is considered more important in the Spanish unit. These discrepancies can possibly be explained by the slightly different work balance of the two units in this organization. Both make extensive use of electronic terminology tools, but only the English unit, which is larger and does more translation work as opposed to editing, takes full advantage of translation memory software. The English



translators produce all documents using the software, share databases, work extensively with the glossary function and make sure background documents are aligned. The Spanish translators, on the other hand, if they use translation memory software at all, do not share databases or use the glossary function. The greater importance given to the ability to follow complicated instructions in the Spanish unit may reflect the influence of the amount of editing work the unit does.

The in-house/external variable therefore seems to have a larger bearing on results than does the target language of the work, although both significantly influence the findings of the recruits questionnaire only. The sample is very small, but in general it can be concluded that the in-house/external factor has a stronger influence over how people answer the questionnaires than the target language, at least at this IGO.

### *3.3.8 Conclusions drawn from second pilot*

#### *3.3.8.1 The reliability and validity of the questionnaires as measuring instruments*

Comparative analysis of the ratings awarded between pilot 1 and pilot 2 shows that respondents answered both questionnaires with a certain degree of consistency, inasmuch as most inconsistencies could be explained. This suggested that the questionnaires were fairly reliable measuring instruments.

The results of the interviews held with all the respondents in the second pilot and the analysis of the changes in their answers from the first to the second pilot suggest that standardizing the rating scale and giving clearer instructions had improved the face and construct validity of the questionnaire since the first pilot. Respondents appeared to have a clearer idea of what they were rating and of what to base their ratings on. Minor adjustments still needed to be made to the wording of some of the item descriptions: (1) to remove the ambiguity of “finding” in “finding the right words or word combinations”, (2) to clarify what was meant by nuance, and (3) to incorporate the idea of maintaining quality despite having to work fast.

Another indicator of the validity of the questionnaires as measuring instruments was the fact that the base list seemed to be a valid point of departure for obtaining information on the subject. No items were reported as not required, and only a few skills and knowledge types were reported as missing. These included one analytical skill (reading between the lines) and basic numeracy, which were, as discussed earlier, possibly covered by others on the list, and two psycho-social skills beyond the scope of this project. This suggested that both

questionnaires would make it possible not only to rank the skills and knowledge included in the list, but also to identify all the skills and knowledge that IGO translators and revisers think are required in their work.

In short, the changes made to the rating scale, the rewording of instructions and the reformatting of the questionnaires seemed to have improved their structural, face and content validity and hence made them more reliable measuring instruments.

The in-house/external work variable, at this organization at least, seemed to influence answers sufficiently to warrant their separate analysis, especially when applying the recruits questionnaire. The different nature of the work of the two language services also seemed to have influenced answers to some extent, and the option of analyzing the answers by target language should thus be kept open in a broader survey.

The frequency ratings would be likely to vary more than the impact ratings, and as large a survey sample as possible should be used to neutralize the influence of the more idiosyncratic aspects of work at a particular organization or unusual attitudes or experiences of some respondents if trying to identify general trends, while retaining the option to break findings down by organization or language service. Data on the expectations of new recruits and the role of translators in document production should also be collected to contextualize findings.

### *3.3.8.2 The hierarchies produced by the questionnaires and their usefulness as input for identifying recruitment priorities*

Both questionnaires produced usable hierarchies. The recruits questionnaire discriminated more among the components of the skills-knowledge set, but the impact questionnaire produced a hierarchy that could equally well be converted into usable weightings for recruitment testing. And with a larger sample base, the discriminatory power of both could be expected to increase.

The hierarchy obtained from the impact ratings reflects the relative impact that, according to the persons with the most in-depth understanding of the translation service, the different skills and knowledge types have on the work of that service. This has implications for recruitment. The higher the impact of a skill or knowledge type, the more important it is for the organization to recruit translators with that skill or knowledge type. Adjusting selection procedures to identify candidates with the profile that emerges from the findings of the impact questionnaire should thus help organizations find translators who will have the most positive impact on the effectiveness of their translations.

The hierarchy obtained from the frequency ratings suggests the relative frequency with which the lack of the skills and knowledge accounts for revision work in the translation service. This also has implications for how to weight the different components of translation competence in recruitment. The more frequently lacking the skill or knowledge type, the more important it is for the organization to find new recruits with that skill or knowledge type. Adjusting selection procedures according to the findings of the recruits questionnaire should help reduce the revision workload and hence enhance the efficiency of the translation service.

The two questionnaires do not produce the same or even particularly similar hierarchies, however. Those obtained from the two questionnaires for in-house work, for which the sample base was largest, are presented in Table 3.3.

Some skills and knowledge types appear near the top of both lists, but using the findings of the impact questionnaire to identify recruitment needs would prioritize some skills and knowledge that the application of the findings of the recruits questionnaire would not, such as TL vocabulary, SL knowledge, TL grammar and mining reference material, for example. In fact, these items have low priorities according to the logic of the recruits questionnaire. Even allowing for the fact that their low ranking in the frequency questionnaire may reflect a successful aspect of current candidate selection procedures, the two hierarchies and their implications are quite different. According to the findings of the recruits questionnaire, for example, subject knowledge, working out the meaning of obscure passages and judging the reliability of sources of information should be top recruitment priorities, but they are not considered to have that much impact on the effectiveness of translations in terms of achieving the IGO's communicative aims in comparison with other components of the skills-knowledge set. They are generating a large proportion of the revision work, however, or they would not be ranked so highly in the recruits questionnaire, which means their absence has a negative impact on the productivity of the translation service, if not directly on its communicative aims. They therefore need to be taken into account.

At this point it became clear that not only were the weighted profiles emerging from the two questionnaires quite different, they were both relevant. It was not a matter of deciding which approach was more appropriate and which questionnaire should be used to rank the components of the skills-knowledge set, but of how to take into consideration the findings of both, and to do so in light of current testing practice and the opportunities for post-recruitment development.

Table 3.3. Comparison of the hierarchies obtained for in-house work from the impact and frequency questionnaires in the second pilot, both rated on scales of 1-5

Skills and knowledge types, by average impact ratings	
Rating	Skill or knowledge type
5.0	An extensive TL vocabulary Write elegantly regardless of the ST Knowledge of the SL Knowledge of TL grammar Convey the ST message clearly Adhere to in-house style conventions Ensure the completeness of the TT Mine reference material for phrasing
4.8	Produce translations that flow smoothly Find the best words Knowledge of TL spelling rules Produce idiomatic translations Ensure the coherence of the TT Work with electronic terminology tools Detect inconsistencies, contradictions, etc. Recast sentences in the TL Explain translation decisions or problems Track down sources to check facts
4.6	Knowledge of SL culture(s) Knowledge of TL punctuation rules Achieve the right tone and register Understand complex topics Work out the meaning of obscure ST Track down sources to understand the topic
4.4	Tailor language to the readers' needs Translate fast and under pressure Capture every nuance of the ST Judge the reliability of information sources
4.2	Work with Excel and/or PowerPoint Handle more than basic Word functions Type accurately and fast
4.0	Knowledge of target-language varieties Knowledge of target-language cultures Master new subjects quickly
3.8	Knowledge of the organization Knowledge of the subject Follow complicated instructions about a text
3.6	Detect mathematical errors in the source text
3.4	Work with translation memory software
3.2	Knowledge of the different varieties of the source language

Skills and knowledge types, by average frequency ratings	
Rating	Skill or knowledge type
4.00	Ability to read between the lines*
3.25	Adhere to in-house style conventions
3.00	Knowledge of the subject Recast sentences in the TL
2.75	Capture every nuance of the ST Produce translations that flow smoothly Find the best words Write elegantly regardless of the ST Detect inconsistencies, contradictions, etc. Work out the meaning of obscure ST Judge the reliability of information sources
2.50	Convey the ST message clearly Ensure the coherence of the TT Produce idiomatic translations Translate fast and under pressure Ensure the completeness of the TT
2.25	Knowledge of the SL Knowledge of the organization Work with translation memory software Handle more than basic Word functions Work with Excel and/or PowerPoint Explain translation decisions or problems Master new subjects quickly Track down sources to check facts Track down sources to understand the topic Knowledge of SL culture(s)
2.00	Knowledge of TL grammar Knowledge of TL punctuation rules Mine reference material for phrasing Follow complicated instructions Detect mathematical errors in the ST Knowledge of TL cultures
1.75	Tailor language to the readers' needs Understand complex topics Knowledge of SL varieties
1.50	Achieve the right tone and register Work with electronic terminology tools Knowledge of TL varieties Type accurately and fast
1.25	Knowledge of TL spelling rules
1.00	An extensive TL vocabulary

\* Suggested by one respondent

Determining recruitment priorities on the basis of the impact ratings alone would possibly award a disproportionately high priority to components of the skills-knowledge set that are actually commonly found among translators. The high-impact skills and knowledge therefore need to be appreciated in light of their relative scarcity. Determining recruitment priorities on the basis of frequency ratings alone, meanwhile, would possibly award a disproportionately high priority to skills and knowledge types that are only incidentally generating a lot of revision work as they are based on the latest batch of new recruits and hence reflect the effectiveness of current testing practice. The high-frequency components of the skills-knowledge set therefore need to be appreciated in light of their relative impact in relation to the other ones involved in producing translations to the perceived standards required and the extent to which they are assessed in recruitment. This would also compensate for the fact that new recruits are not always given the most challenging work. IGOs need to find translators who have these oft-lacking components *in addition to*, not instead of, the high-impact ones already successfully tested.

In both cases, the expectations of new recruits needed to be taken into account in order to determine what should be included in recruitment tests and what should be developed through post-recruitment training or experience. The method developed to do this is described in section 3.4 below.

### **3.4 Methodology for identifying recruitment priorities**

The method devised for cross-referring the findings of the two questionnaires in order to identify recruitment priorities consists of four simple steps:

- Step 1: Isolate the skills and knowledge that entry-level translators are expected to have.<sup>11</sup>
- Step 2: Map the frequency ratings and the impact ratings on a scatter chart.
- Step 3: Apply the logic of qualitative risk analysis, as described by Tusler (1996), for example, to group the skills and knowledge types into categories, as shown in Figure 3.12.
- Step 4: Classify the components of each category according to whether they are measured in current recruitment tests or not.

---

<sup>11</sup> To identify in-house *training* priorities, the procedure would be the exact opposite, i.e. isolate the components of the skills-knowledge set that new recruits are not expected to have.

Performing steps 1-3 produces the classification shown in Figure 3.12.

Figure 3.12. Model for the categorization of skills and knowledge types that entry-level translators at IGOs are expected to have, by impact and frequency<sup>ab</sup>

	High-frequency	Low-frequency
High-impact	Category A	Category B
Low-impact	Category C	Category D

<sup>a</sup> Impact refers to the impact of the skills and knowledge types on how effectively translations achieve the communicative aims of the organization and protect its image.

<sup>b</sup> Frequency refers to the frequency with which the skills and knowledge types are found lacking among new recruits.

The Category A items and knowledge types are the top recruitment priorities, followed by the Category B and C items and then the Category D items. The high-impact, high frequency components of the skills-knowledge set in Category A are important both for improving the effectiveness of translations and reducing the revision workload. The Category B components are more important for averting errors that might harm the reputation of the IGO, and the Category C components for enhancing productivity. The Category D components are required but are less important. The cut-off points between high and low impact and high and low frequency can be determined by the rating scales used in the questionnaires and how selective the exercise is meant to be.

Bearing in mind that the frequency ratings reflect how current testing procedures filter candidates, the categories need to be broken down further according to whether the skills and knowledge in them are assessed in the organization's current recruitment examinations (Step 4), as shown in Figure 3.13.

Figure 3.13. Groupings obtained by cross-referring the findings of the impact and frequency questionnaires and according to whether items are tested in current recruitment

	High-frequency	Low-frequency
High-impact	Tested ( i )	Tested (iii)
	Not tested ( ii )	Not tested (iv)
Low-impact	Tested (v)	Tested (vii)
	Not tested (vi)	Not tested (viii)

The implications of each category are as follows:

- (i) These skills are not being adequately tested and need to be weighted heavily.
- (ii) These skills need to be incorporated into testing and heavily weighted.
- (iii) These skills are being adequately assessed in recruitment.
- (iv) New recruits seem to have these skills or they are not required often. The organization should maybe check recruits have them, but there is no need to weight them as heavily as those in groups (i) and (ii).
- (v) These skills are not being adequately tested. They need testing but do not need to be weighted as heavily as those in group (i).
- (vi) These skills need to be incorporated into testing but do not need to be weighted as heavily as those in group (ii).
- (vii) These skills are being found. The organization should check they are not being weighted unnecessarily heavily in the grading scheme used. They should not be weighted more than the skills in group (iii).
- (viii) These skills do not need incorporating into testing.

The findings of both questionnaires can thus be cross-referred and classified according to current testing practice to identify the components of the skills-knowledge set that the IGO should prioritize in its selection of translators, as well as the adjustments it should make in its recruitment testing. This method is applied to several translation services in Part II below, using the data obtained from the survey conducted of IGOs, which are presented in Chapter 4 and analyzed in Chapter 5.

### **3.5 Conclusions on the methodology for identifying the skills and knowledge required in IGO translation**

The following conclusions were reached after testing, refining and retesting the instrument designed to identify and rank the skills and knowledge types used in IGO translation.

#### *On the impact questionnaire*

With improvements to the wording of a few of the items (regarding finding words, capturing nuances and maintaining quality under pressure) in the base list, the impact questionnaire would seem to be a reliable instrument for measuring the relative impact of the different components of translation competence on the effectiveness of translations at IGOs. By averaging the impact ratings, a weighted list of those components can be obtained that has implications for training and recruitment.

#### *On the recruits questionnaire*

With the same improvements to the wording of a few of the items in the base list as for the impact questionnaire, the recruits questionnaire would seem to be a reliable instrument for measuring the relative scarcity of the different components of translation competence among new recruits at IGOs. By averaging the frequency ratings, a weighted list of those components can be obtained that also has implications for training and recruitment.

#### *On the methodology devised for identifying the relative importance of the components of the skills-knowledge set*

The findings of the two questionnaires can be cross-referred to identify recruitment priorities. These can be further broken down to identify possible adjustments that IGOs should make to current testing practice to find translators with the profile that most closely matches their needs.

In light of these findings, it was decided to make the few phrasing changes found to be necessary in both questionnaires and then to circulate the questionnaires among a large a sample of IGO translators and revisers as possible to identify the profile of the ideal candidate for IGO translation work in general and determine whether that profile varied from one organization or translation service to another. The possibility of analyzing the influence of the TL, in-house/external work and reviser/translator variables would be kept open, even



though only the type of work, and to some extent the TL, seemed to affect ratings significantly in the pilot IGO. The context information obtained in the second pilot regarding expectations of new recruits and document-processing responsibilities would also be collected in the survey, as it had proven invaluable for interpreting the results and cross-referring the findings of the two questionnaires.

## **Chapter 4 Results of the survey of the skills and knowledge required in IGO translation**

Having obtained evidence in the second pilot that both the impact and the recruits questionnaires were reliable instruments for measuring the relative importance of components of translation competence, and having devised a method for cross-referring the resulting hierarchies, it was decided to survey as large a number of IGO translators and revisers as possible. The goal was to use the results to rank the set of skills and knowledge required at IGOs in general and then to analyze the implications for text-based recruitment testing at individual IGOs.

### **4.1 The revised questionnaires**

Only a few changes were made to the questionnaires in light of the findings of the second pilot. The description of the ability to translate fast was reworded as “the ability to maintain quality even when translating under time pressure”. The need to resolve the ambiguous wording regarding “finding the right word” and to define “nuances” more precisely was addressed by replacing the corresponding items with the following ones: “the ability to find the right word or word combination to capture the exact and detailed meanings (nuances) of the source text” and “the ability to convey the intended effect of the source text”. The latter item was a new addition thought to possibly be important in IGO translation work given the political nature of some texts and the concerns with upholding the reputation of the institution. The descriptors of the extremes of the impact scale were changed to “extremely small” and “extremely large”, which seemed more suitable than “minimal” and “enormous”. The revised impact questionnaire is presented in Appendix 6 and the revised recruits questionnaire in Appendix 7.

### **4.2 The survey sample**

To reach the largest possible number of translators and revisers at IGOs, the backing of the International Annual Meeting on Language Arrangements, Documentation and Publication

(IAMLADP) was sought and obtained. This is a forum and network of managers of international organizations employing conference and language service providers, mainly translators and interpreters. The sample population was relatively homogeneous inasmuch as the respondents worked in the same domain, had similar academic profiles and had all passed similar recruitment examinations. In terms of characteristics, it was highly representative of the population to which we wished to generalize the results. The final sample included 165 completed impact questionnaires, of which 163 were usable, and 157 completed recruits questionnaires, of which 153 were usable. The answers from all the usable questionnaires were included in the analysis.

### **4.3 Procedure**

The questionnaires were prepared in electronic format and were sent out in February 2010 by the Secretary of IAMLADP to the focal points in each member IGO (usually the head of language services) for forwarding to translators and revisers. This was an efficient way to reach a broad spectrum of the target population. IAMLADP backing lent credibility to the exercise, making it possible to obtain responses from a large number of professional translators and revisers. The cover e-mail ensured participants anonymity and invited them to complete the questionnaires online. Data were collected over a period of three months.

### **4.4 The results of the impact questionnaire**

#### *4.4.1 The respondents*

The impact questionnaire was completed by 163 language professionals (7 heads of unit, 27 revisers and 129 translators) from over 24 bodies of the European Union (EU), the United Nations (UN) and other international organizations.<sup>12</sup> The distribution of responses across

---

<sup>12</sup> The exact number of organizations is difficult to determine as the specificity of answers varied. The distribution of replies was as follows: European Union bodies: EESC and Committee of the Regions (65); EU Council (13); European Patent Office (2); unspecified European Union (2); United Nations system: UNOG (27), unspecified UN (12), ECLAC (8), UNHQ (4), FAO (3), UNOV (2), UNON (1), ICAO (1), ECA (1), ESCWA (1), ITU (1), ICC (1), ILO (1) Special tribunal for the Lebanon (1), and World Bank (1); ICRC (1); Non-IAMLADP members invited to participate by the researcher: IDB (1) IIC (4), SELA (1); freelancers

these three broad groups was 82:58:6, respectively. To protect the anonymity of individual organizations, they are identified only by which of these large umbrella groups they belong to.

Almost 91% of respondents reported on the skills and knowledge required in in-house work, and just over 9% on those required in external work.

Translators working in a wide range of language combinations participated in the survey, as shown in Figure 4.1 (source languages) and Figure 4.2 (target languages). English was a source language for the vast majority of respondents (79%), followed by French (70%) and Spanish (44%), and these three languages were also the most common target languages (29%, 18% and 15%, respectively). The average number of source languages was three, with translators from the European Union bodies accounting for nearly all those translating from more than two languages. The vast majority of translators (92%) work into only one target language, presumably their mother tongue (or L1). This is possibly one of the features that set translators at international organizations apart from translators in other fields. The exceptions in this survey population are the Arabic translators in the United Nations, who work both into and out of their L1, as do some of the Spanish-English translators in what are referred to in here as Inter-American (IA) organizations (IDB, IIC and SELA), and some translators in the EU working with languages spoken by relatively smaller populations.

---

working for more than one organization (8); and an unspecified Commission (1). A list of the acronyms used is presented on page 1.

Figure 4.1. Source languages of respondents to the impact questionnaire, by number of respondents

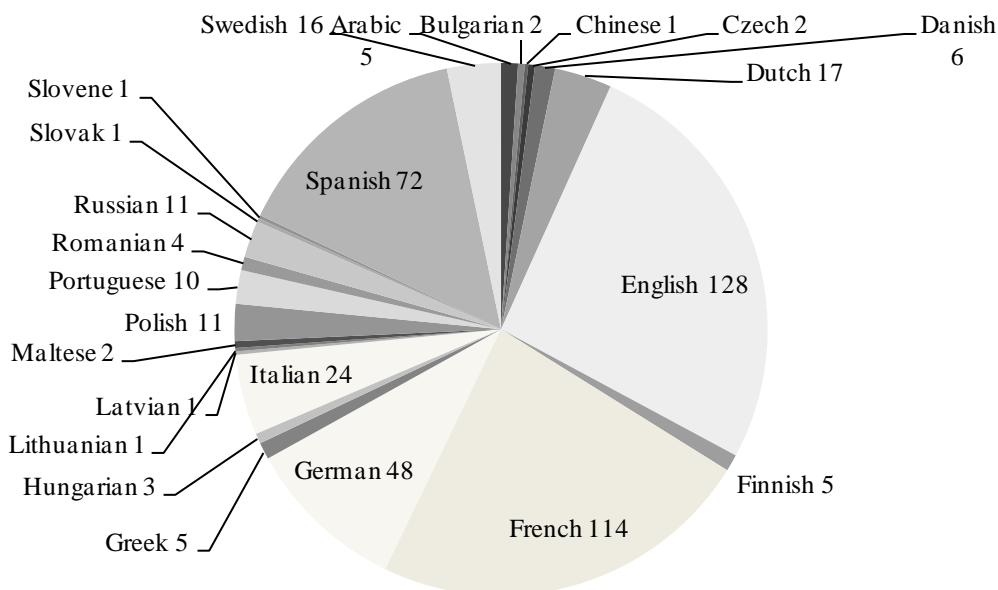
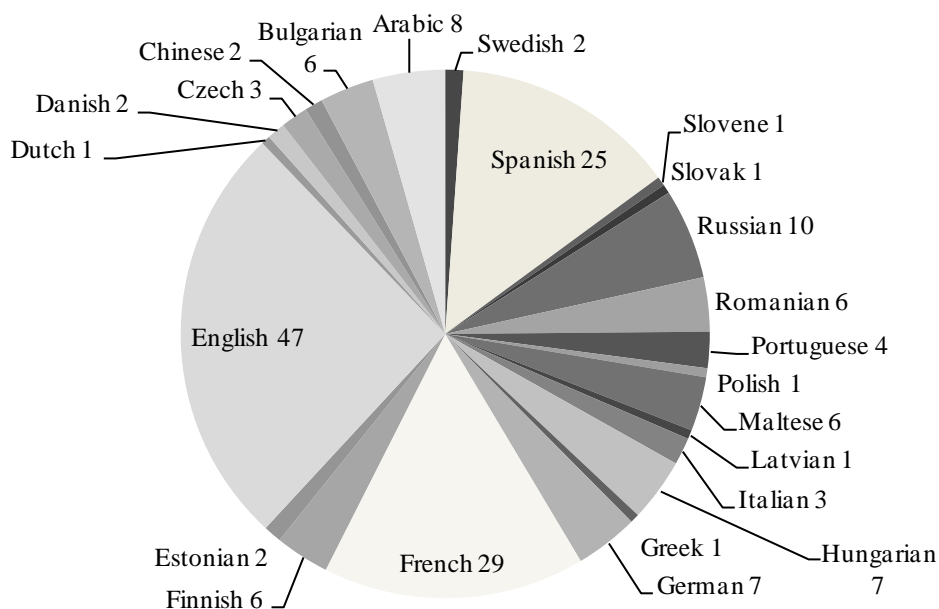


Figure 4.2. Target languages of respondents to the impact questionnaire, by number of respondents



#### 4.4.2. The impact ratings

The mean impact ratings for the skills and knowledge types presented in the impact questionnaire are shown in Table 4.1.

Table 4.1. Impact ratings of selected skills and knowledge types<sup>a</sup>*(Response count and mean rating)*

Rank	Skill or knowledge type	Rating Scale <sup>b</sup>						Mean Rating
		1	2	3	4	5	N/A	
1	Ensure the completeness of the TT	1	0	5	19	137	1	4.80
2	Convey the ST message clearly	1	0	7	20	135	0	4.77
3	Knowledge of SL	1	1	5	24	132	0	4.75
4	Ensure the coherence of the TT	1	1	2	29	130	0	4.75
5	Knowledge of TL spelling rules	1	1	10	22	129	0	4.70
6	Knowledge of TL grammar	2	1	6	29	125	0	4.68
7	Work out the meaning of obscure passages	1	2	6	37	117	0	4.64
8	An extensive TL vocabulary	2	2	11	24	124	0	4.63
9	Capture nuances of ST	1	5	8	29	120	0	4.61
10	Knowledge of TL punctuation rules	1	2	12	31	117	0	4.60
11	Produce idiomatic translations	3	2	9	28	120	1	4.60
12	Maintain quality even under time pressure	1	0	14	38	108	2	4.57
13	Convey the intended effect of the ST	3	3	12	28	117	0	4.55
14	Detect inconsistencies, contradictions, etc.	2	3	12	34	112	0	4.54
15	Adhere to in-house style conventions	1	2	9	46	104	1	4.54
16	Track down sources to check facts	3	3	11	34	112	0	4.53
17	Mine reference material for phrasing	2	2	13	37	109	0	4.53
18	Produce translations that flow smoothly	3	3	12	40	105	0	4.48
19	Achieve the right tone and register	3	5	13	37	105	0	4.45
20	Understand complex topics	1	1	18	51	92	0	4.42
21	Knowledge of the organization	1	3	16	57	85	1	4.37
22	Recast sentences in the TL	2	6	16	46	93	0	4.36
23	Judge the reliability of information sources	6	5	12	43	97	0	4.35
24	Track down sources to understand the topic	2	3	20	53	85	0	4.33
25	Work with electronic terminology tools	2	5	19	50	86	1	4.31
26	Follow complicated instructions	3	2	27	47	82	2	4.26
27	Master new subjects quickly	2	6	23	65	67	0	4.16
28	Write elegantly regardless of the ST	3	5	28	53	73	1	4.16
29	Subject knowledge	2	5	33	53	70	0	4.13
30	Knowledge of TL culture(s)	5	10	25	43	80	0	4.12
31	Tailor language to the readers' needs	4	9	24	55	69	2	4.09
32	Explain translation decisions and problems	4	10	25	63	60	1	4.02
33	Work with translation memory software	8	10	26	46	68	5	3.99
34	Knowledge of SL culture(s)	1	11	41	61	49	0	3.90
35	Knowledge of TL varieties	11	13	31	36	67	5	3.85
36	Handle more than basic Word functions	4	21	46	50	34	8	3.57
37	Knowledge of SL varieties	7	21	46	59	28	2	3.50
38	Type accurately and fast	9	18	51	52	27	6	3.45
39	Detect mathematical errors in the ST	25	22	49	26	21	20	2.97
40	Work with Excel and/or PowerPoint	29	31	41	35	8	19	2.74

<sup>a</sup> Where TT=target text; TL= target languages; SL=source language; ST= source text. Descriptions of skills and knowledge have been abbreviated for the sake of space. For the full descriptions, see the impact questionnaire in Appendix 6.

<sup>b</sup> Ratings awarded in response to the question "How large is the impact of the following skills and knowledge types on the effectiveness of translations at your organization?" using a scale of 1-5, where 1=extremely small and 5=extremely large.

Being able to make sure that all the content of the original is relayed in the translation is apparently the most important skill translators need to have at IGOs. Although a rather fundamental aspect of the translator's task in any context, it may assume particular importance in certain IGO settings such as committee meetings, in which participants are working from parallel language versions and any omissions (or unnecessary additions for that matter) may incur comment, confusion and delays and hence embarrassment and costs. Being able to express ideas clearly in the target language and having sound knowledge of the source language are also unsurprisingly rated as highly important assets. Ensuring coherence, i.e. internal consistency within each translation and all the documents produced by their translation services, is another understandable priority in IGO translation services. Translations are the written records of the institution's workings (Koskinen 2008:125). Possibly more unexpected is the fact that analytical skills are among those rated as being highly important (mean rating of 4.5 or above). The ability to work out obscure passages in the source text and the ability to detect inconsistencies, contradictions, nonsense, unintended ambiguities, misleading headings, etc., are ranked 7th (mean rating 4.64) and 14th (mean rating 4.54), respectively. This may be explained by the need to translate unedited source texts. The importance of being able to maintain quality even when working under time pressure (ranked 12<sup>th</sup>) reflects another common aspect of IGO translation work, especially when servicing meetings.

It should be noted that even the skills and knowledge types at the bottom end of the ranking produced by the impact questionnaire are not considered unimportant. On a scale of 1-5, a rating of 3 or more suggests the skills or knowledge type is at least of some importance. This applies to 38 of the 40 included in the list. The only two to which it does not apply are the ability to detect mathematical errors and the ability to work with Excel and PowerPoint, which were also reported as not required at all by about 12% of respondents. In other words, the vast majority of skills and knowledge types presented for rating were considered to have an impact of the effectiveness of IGO translation work.

#### *4.4.3 The additional skills and knowledge reported in the impact questionnaire as being needed on the job*

The optional open question in the impact questionnaire on other skills and knowledge required was answered by 48 people (29% of respondents), and general observations were made by 50 (31%). The responses provide several insights into the nature of translation work

at IGOs today. The skills and knowledge types that, according to the survey participants, are required in addition to the 40 included in the ranking can be loosely grouped as shown below. The number of times each one is mentioned is presented in parentheses, and the respondents' own wording is used as far as possible:

- Communication skills to be able to elicit assistance or answers from others in the organization, especially authors of source texts (11).
- Ability to work with revisers: openness to feedback and ability to learn from it (10).
- Teamwork skills, including for working on large translation projects and with shared translation memories (8).
- Language skills: ability to learn new languages, translate from several source languages, ability to create new terminology for new concepts, to adapt to the style used in the unit or department(8).
- Knowledge of translation theories and practices (6).
- Critical-thinking skills (ability to “reason through” the translation, a “second channel” in the brain to think beyond just choosing the right word) (5).
- Organization and time management skills (5).
- General knowledge, awareness of current and world events (5).
- Interpersonal skills: patience with authors, requesters and colleagues (4).
- Flexibility, adaptability to cope with unpredictable workloads, changes in procedures, etc. (4).
- Experience in other international organizations, ability to cope with life in another country, awareness of cultural differences (3).
- Ability to work independently (3).
- Ability to deal with a wide range of specialized subjects (3).
- Self-revision skills (1).
- Willingness to search for information (1).

As shown in the above list, interpersonal skills such as the ability to work as a member of a team and communicate with “clients” are mentioned more than any other type. Several comments on the need to be able to seek clarification from authors were accompanied by remarks on the poor quality of source texts and the ability to argue points diplomatically and “not to be rattled by criticisms of one’s English from people not really qualified to make them”, as one translator put it.



In addition to suggesting other skills and knowledge types, several respondents to the impact questionnaire took the opportunity to emphasize the importance of expert subject knowledge (of economics and legal systems for instance), which was ranked only 29th, however, by respondents across all organizations, possibly revealing the prime importance of this knowledge in some organizations (rated 5 by 70 respondents) but not in others.

A number of respondents said they had nothing to add and commented on the comprehensiveness of the list of skills and knowledge types presented for rating. Some took the opportunity to sum up what they thought translators at IGOs need. One EU translator wrote: “Basically, we need a thorough knowledge of source languages and the ability to efficiently find whatever info/terminology we need and convey the message in the target language clearly while adhering to in-house rules, often [translating] from source texts of poor linguistic quality.” What would appear to be a seasoned reviser at one UN body noted: “The skills required haven't changed much: thorough knowledge of source and target languages, highly developed analytical skills, and the ability to assimilate new knowledge quickly.” Both, incidentally, mention two not wholly unrelated aspects of IGO translation that are not usually given much attention in translator training programs or recruitment: the translation of poorly written source texts and the need for superior analytical skills. Other comments that have an indirect bearing only on the skills-knowledge set but offer insights into the nature of translation work at IGOs are included in Appendix 8.

#### *4.4.4 The skills and knowledge that can be acquired after recruitment*

Just over half the survey population (84/163) answered the optional question on whether translators are expected to have all the 40 skills and types of knowledge at the moment they start working for the organization. Expectations varied slightly, as shown in Table 4.2, in which answers have been broken down by broad organization group.

The skills and knowledge most often mentioned as not being expected prior to recruitment are (or may be) organization-specific. This is the case of translation technology usage, adherence to in-house style, knowledge of the organization and the ability to work with references. Most organizations have their own custom-designed document search and terminology tools, and, according to informal consultations made and as suggested by the analysis of the variables in the preceding section, the extent to which translation memory programs are used varies considerably from one organization to the next and even within the same organization from one language service to another.

Table 4.2. Skills and knowledge that translators are expected to acquire after recruitment, by group of organizations and proportion of answers (Percentages)<sup>a</sup>

Skill or knowledge type	European Union	United Nations	IA <sup>b</sup>
Work with translation memory software	68	21	80
Work with electronic terminology tools	60	38	60
Adhere to in-house style conventions	44	59	40
Knowledge of the organization	30	38	40
Handle more than basic Word functions	16	14	
Track down sources to check facts	14	10	20
Mine reference material for phrasing	14	21	
Follow complicated instructions	12	14	
Work with Excel and/or PowerPoint	12	10	
Subject knowledge	10	24	
Maintain quality even under time pressure	6	10	
Knowledge of TL varieties	4	3	
Judge the reliability of information sources	4	7	
Knowledge of TL spelling rules	2	3	
Explain translation decisions and problems	2	7	
Master new subjects quickly	4		
Capture nuances of ST	2	3	
Convey the ST message clearly	2		
Tailor language to the readers' needs	2		
Ensure the coherence of the TT	2		
Work out the meaning of obscure passages	2		
Detect inconsistencies, contradictions, etc.	2		
Knowledge of SL varieties	3		

<sup>a</sup> Figures correspond to the proportion of translators from each organization group that listed the skill or knowledge as not being required prior to recruitment.

<sup>b</sup> Based on the answers of 50 translators from the EU, 29 from the UN and 5 from the IA Group (SELA, IIC and IDB).

#### 4.4.5 *The profile(s) of the “ideal candidate” emerging from the findings of the impact questionnaire*

The findings confirm that translators at IGOs employ a wide range of skills and knowledge in their work. Together with those that might be expected to be ranked among the most important for all kinds of translation work (SL and TL knowledge and writing skills) are skills that might be more exclusively associated with translating at international organizations (such as adhering to in-house style and mining reference material for acceptable terminology and phrasing), as well as specialized knowledge (of the organization and of economics, politics, current affairs, etc.), analytical skills (understanding complex topics and mastering new ones), and editing skills, such as detecting inconsistencies in a text. Translators also need

research, interpersonal and organization skills; they must be willing to work with technology, and to work fast.

How the profile may vary according to different variables, such as target language, whose opinion the weightings are based on and the type of work, is analyzed in Chapter 5.

## 4.5 The results of the recruits questionnaire

### 4.5.1. *The respondents*

The recruits questionnaire was completed by 153 people in total (128 revisers and 25 heads of service) from over 20 bodies of the United Nations, the European Union and other international organizations.<sup>13</sup> The distribution of responses across these three broad groups was 83:65:3, respectively. Again, to protect anonymity, organizations and individual respondents are identified only by which of these large umbrella groups they belong to.

Just over 80% of respondents reported on the work of in-house translators, and slightly under 20% on that of external translators.

The results are based on new recruits working with a wide range of language combinations, as shown in Figures 4.3 and 4.4, with most work being from and into English, French and Spanish.

---

<sup>13</sup> The exact number of organizations is difficult to determine as the specificity of answers varied. The distribution of replies was as follows: UNOG (20), UNOV (8), UNON (2), unspecified UN (20), ICAO (2), ECLAC (6), ECA (1), ESCWA (1), ITU (1), ILO (4), ICC (2), IFAD (3), WFP (3), Special tribunal for the Lebanon (1), Special tribunal for the Ruanda (1), ITLOS (1), FAO (3), World Bank (3), IMF (1); Non-IAMLADP members invited to participate by the researcher (2 organizations, 4 respondents); IDB (2) IIC (2); 65 EU (EESC and Committee of the Regions (45), European Patent Office (3), EU Council (16), EBRD (1)); 1 freelance reviser working for more than one organization.

Figure 4.3. Source languages of recruits' work, by number of respondents

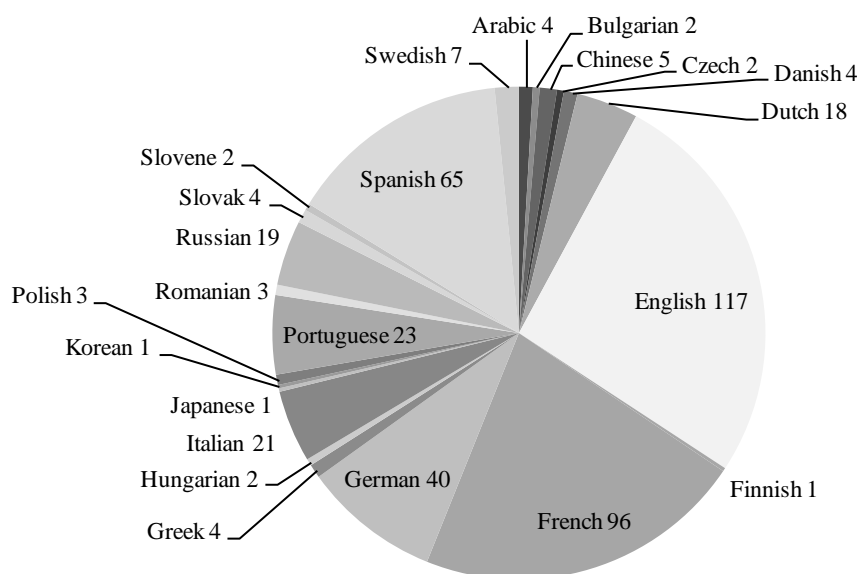
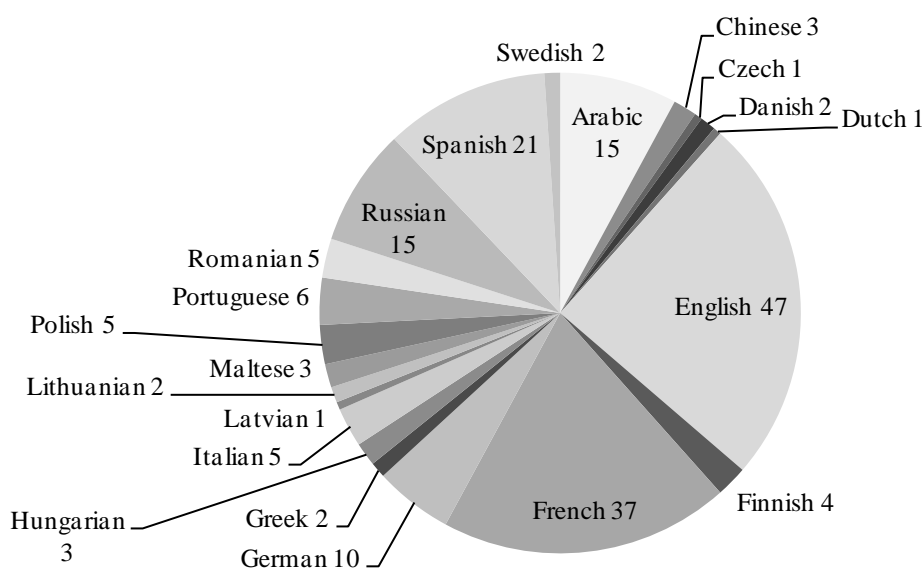


Figure 4.4. Target languages of recruits' work, by number of respondents



#### 4.5.2 The frequency ratings

The mean frequency with which each skill or knowledge type was found lacking among new recruits is presented in Table 4.3. New recruits at least sometimes are reported to lack 18 of the 40 skills and knowledge types that revisers were asked to rate on the five-point scale (these were all rated above 3). Of these, almost half (eight) are related to TL writing skills,

three are research skills, two refer to analytical skills (working out obscure passages and detecting inconsistencies), two to specialized knowledge (of subject areas and of the organization), one to capturing detailed and exact meanings (nuances), one to in-house style and one to productivity. In other words, in addition to linguistic skills traditionally associated with translation and deficient TL writing skills usually associated with revision work, analytical skills (the ability to detect inconsistencies in the source text and to judge the reliability of sources) are among those rated, on average, as more often than not lacking among new recruits. Indeed, the relative inability of new recruits to figure out meaning is generating almost as much revision work as their (expected) ignorance of in-house style. And this shortcoming is not due to their lack of language skills, since their SL knowledge is only relatively rarely found to be deficient (ranked 22<sup>nd</sup> with a mean frequency rating of only 2.78).

Table 4.3. Frequency with which selected skills and knowledge types are found lacking among new recruits<sup>a</sup>*(Number of responses and mean rating)*

Rank	Skill or knowledge type	Rating Scale <sup>b</sup>						Mean Rating
		1	2	3	4	5	N/A	
1	Produce translations that flow smoothly	4	13	41	60	34	1	3.70
2	Work out the meaning of obscure passages	4	22	37	63	26	1	3.56
3	Write elegantly regardless of the ST	4	21	43	54	29	2	3.55
4	Capture nuances of ST	2	18	54	54	23	2	3.52
5	Adhere to in-house style conventions	12	18	46	40	35	2	3.45
6	Recast sentences in the TL	4	25	45	55	22	2	3.44
7	Subject knowledge	3	22	54	55	18	1	3.41
8	Detect inconsistencies, contradictions, etc.	7	25	44	56	20	1	3.38
9	Produce idiomatic translations	9	28	45	51	18	2	3.27
10	Convey the intended effect of the ST	8	28	56	42	17	2	3.21
11	Judge the reliability of information sources	15	24	46	44	20	4	3.20
12	Maintain quality even under time pressure	6	31	59	33	20	4	3.20
13	Knowledge of the organization	12	29	47	40	21	4	3.19
14	Ensure the coherence of the TT	8	28	59	40	16	2	3.19
15	Convey the ST message clearly	7	32	52	50	11	1	3.17
16	Achieve the right tone and register	13	30	49	41	19	1	3.15
17	Track down sources to check facts	15	30	51	38	17	2	3.08
18	Track down sources to understand the topic	12	31	55	37	15	3	3.08
19	Mine reference material for phrasing	11	36	52	41	10	3	3.02
20	Tailor language to the readers' needs	12	42	48	31	14	6	2.95
21	Understand complex topics	17	46	45	31	12	2	2.83
22	Knowledge of SL	18	38	64	23	9	1	2.78
23	Master new subjects quickly	19	49	44	25	11	5	2.73
24	Explain translation decisions and problems	25	35	50	23	10	10	2.71
25	Knowledge of TL grammar	20	55	40	27	10	1	2.68
26	An extensive TL vocabulary	25	49	44	26	8	1	2.63
27	Knowledge of TL punctuation rules	28	49	39	27	9	1	2.61
28	Ensure the completeness of the TT	24	51	45	22	9	2	2.61
29	Knowledge of SL culture(s)	21	52	49	24	5	2	2.60
30	Follow complicated instructions	29	39	46	18	7	14	2.53
31	Work with Excel and/or PowerPoint	30	28	19	13	11	52	2.48
32	Work with translation memory software	37	37	33	20	9	17	2.46
33	Work with electronic terminology tools	38	40	39	23	6	7	2.45
34	Knowledge of TL varieties	39	45	40	15	7	7	2.36
35	Knowledge of TL culture(s)	47	43	36	18	7	2	2.30
36	Knowledge of SL varieties	38	50	42	7	8	8	2.29
37	Handle more than basic Word functions	40	41	25	15	7	25	2.28
38	Knowledge of TL spelling rules	55	51	24	13	8	2	2.13
39	Detect mathematical errors in the ST	50	39	22	7	5	30	2.01
40	Type accurately and fast	57	40	17	5	5	29	1.88

<sup>a</sup> Where TT=target text; TL= target languages; SL=source language; ST= source text. Descriptions of skills and knowledge have been abbreviated for the sake of space. For the full descriptions, see the recruits questionnaire in Appendix 7.

<sup>b</sup> Ratings were awarded in response to the question "Please think about the mistakes you usually correct when going over translations done by new recruits. How often do you think those mistakes are due to a lack of the following? 1=Almost never and 5=Almost always."

#### 4.5.3 *Comments on other skills and knowledge found lacking and general observations made about new recruits and training*

The optional question on other skills and knowledge that new recruits often lack and need was answered by 26 people (17% of respondents), and general observations were made by 48 (31%). The answers can be grouped as follows (the number of times each one is mentioned is presented in parenthesis and the respondents' own wording is used as far as possible):

- General knowledge, awareness of current affairs (10)
- Organization and time management skills, including ability to attain the right balance between speed and quality (7)
- Quality management skills: how and when to adjust criteria, ability to check for internal consistency and read for sense (7)
- Knowledge of translation and terminology theories, strategies and practices (6)
- Target-language skills: thorough grasp of own language, correct grammar; the ability to write idiomatically, avoid “apishly reproducing the structure of the source text”, and achieve the right tone and register (6)
- Ability to work with CAT tools and judge the results obtained (5)
- Client orientation, professional conduct (3)
- Awareness of the source-language structures that pose translation challenges (2)
- Sensitivity to the subtext and implications of what the text is saying (2)
- Ability to approach and work effectively with authors of source texts (2)
- Teamwork skills (2)
- Ability to work with poorly written source texts (1)
- Ability to work with revisers: openness to feedback and ability to learn from it (1)
- Willingness to take the time needed to search properly for information (2)

Several revisers also took the opportunity in the comments section to stress the shortage of certain skills and knowledge presented in the questionnaire: subject or technical knowledge (ranked 7<sup>th</sup> in the frequency ratings), for example, was mentioned again eight times; and knowledge of the organization and its activities (ranked 13<sup>th</sup>) was referred to again three times. There were in general more remarks about the lack of target-language writing skills than about deficient source-language comprehension skills.

The component most often mentioned as lacking among new recruits (in addition to those rated in the questionnaire), however, was general knowledge. This is considered vital

because, as one reviser put it, “[t]he more understanding you have of the general background, the better are your translations”. There were several comments on how younger translators in particular rarely read newspapers or news magazines to keep themselves informed of current affairs.

There were a number of observations about the inferior quality of the translations produced by new recruits in general: they “tend to translate words and not meanings” (and “[s]ince most documents are written by non-native speakers, the translations can be meaningless”); they lack the ability to “step back” and read their own work for “internal consistency”. Also, reaffirming the high ranking of the abilities to produce translations that flow smoothly and to produce idiomatic translations, according to some respondents, there is “too much tendency to stick to the [original language]”; and “translations are too literal”. A number of revisers also expressed concern over an apparent lack of translation strategies and techniques (mentioned six times in the comments), which is possibly reflected in the relatively high ranking of the ability to recast sentences.

Several comments were made about time and workload management and having to work to increasingly shorter deadlines. Some bemoaned new recruits’ inability to keep up; others lamented the lack of room afforded under the more “market-oriented approach” for “continuous learning by (new) translators, which is a vital prerequisite to staying abreast with new knowledge”.

The number of comments on the need for translators with advanced computer skills (5) seems to point to two views of training needs among revisers. There are those who appear to think translators should have mastered the relevant technology before recruitment, and others who think there is a need to “get ‘back to basics’, instead of worrying about how to use new technologies (which can usually be learned very quickly)”. The same reviser commented, “If graduates have a sound grasp of their working languages (and of the history and culture they represent), well-developed skills in their own language, and solid analytical skills (developed, for example, by a more traditional university course in modern languages), other skills can be learned fairly easily.”

The other comments made by revisers and translators in this and the impact questionnaire are included in Appendix 8. Although they do not have a direct bearing on the identification of the skills-knowledge set, they provide an insight into the world of IGO translation and the concerns of IGO translators and revisers. As such, they constitute valuable input for the analysis of recruitment test design presented in Part II below.



#### 4.5.4 *The expectations of translators in the document production process*

The figures in Table 4.4 are response counts, unweighted by organization, to questions on different aspects of document processing. Presenting them by organization is difficult because of the way in which revisers identified their employer and because, upon closer inspection, differences arise not just between organizations, but between the different language services within the same organization. Among the UN bodies, for example, editing services are provided rarely for the English translators, often for the French, sometimes for the Russian and always for the German translators. The frequency with which texts are pre-translated (i.e. previously translated text, “reprise”, is identified) in EU bodies ranges from never to always, with opinions divided even within the same language service: in the Spanish service, two revisers answered “never”, another two “rarely” and another two “sometimes”; in the English service, the spread was even greater (eight “never”, three “rarely”, five “sometimes” and three “often”); while in the Portuguese service, four revisers said the source texts are never pre-translated for them, and one said they always are.

Organizations are still making a huge effort to revise translations, with translation work being fully revised at least sometimes, if not often or always, against the source text in all cases except the following: the work of in-house Spanish translators at one UN body, freelance Spanish translators at another UN body, freelance French translators at yet another, and in-house Greek translators at an EU body. In these cases, translations are only rarely revised.

Few organizations send translations for approval by the author or requester. This is not surprising given that authors and requesters are not necessarily qualified to judge the quality of a translation. The only translation services where authors or requesters are always involved are the Arabic unit at one UN body (although another Arabic reviser said they never were) and those of smaller, highly specialized organizations: two regional economic commissions of the United Nations, and one of the IA bodies. At one of the UN regional commissions and at the IA body, nearly all translation work is from and into English and Spanish, and authors are usually bilingual experts in their field. This possibly explains the practice of systematically submitting translations for authors’ approval.

Table 4.4. Summary of responses to questions on document processing, unweighted percentages of total responses rounded to the nearest whole number

Question	Response options				
	Never	Rarely	Sometimes	Often	Always
1 How often are source texts edited (for style and/or content) before they are translated?	9	32	32	22	5
2 How often are source texts pre-translated before being sent to the translator (previously translated text is highlighted or provided)?	29	27	26	14	4
3 How often are source texts referenced before being sent to the translator (names are checked, official translations of terms given, background documents and sources identified, etc.)?	10	18	22	41	9
4 How often in practice are translations fully revised against the source text?	0	4	16	52	28
5 How often are translations (whether revised or not) sent to the author or requester for approval?	46	30	12	6	6

The larger organizations in particular are engaging in some upstream quality control by having texts edited before they are sent for translation. In the UN bodies, this usually occurs at least sometimes or often, but not always, and for non-English source texts, even rarely. A smaller proportion of source texts are edited prior to translation at the European Union, where, according to the revisers, texts are only rarely or sometimes subjected to this process. The only EU institution where texts are often edited is a highly technical body. This ties in with the number of complaints made by the EU translators in the impact questionnaire about the quality of the source texts they have to work with. It is, however, the smaller bodies within the UN system that seem to have systematically managed to incorporate source-text editing into the document production process: English texts are always edited prior to translation in three of them, and often at two others. Spanish texts are also often edited for the English translators at one of the regional commissions. At ten of the organizations, editing services are only rarely provided.

More translation services seem to be providing support to translators in the form of references (question 3). Names are checked, official translations of terms given, background documents and sources identified as follows: often or always in most UN bodies; sometimes in most of the EU (but rarely in a few EU bodies); sometimes or often in two of the IA institutions; and rarely or never in five relatively small organizations and certain EU services.

#### *4.5.5. The profile(s) of the “ideal candidate” emerging from the findings of the recruits questionnaire*

The results of the recruits questionnaire show which skills and knowledge types are most often found lacking among new recruits in IGOs. In general, new recruits are reported as mostly lacking writing skills (the ability to produce idiomatic, elegantly written translations that capture the nuances of the source text and to recast sentences) and analytical skills (the ability to work out obscure meanings and detect inconsistencies). Insufficient subject knowledge is also a common cause of error, as is the inability to maintain quality when working under time pressure. Deficient research skills account for revision work at least some times. Computer skills seem to be the least lacking. The answers to the open-ended questions about new recruits reveal frustration among revisers with the lack of general knowledge, critical thinking and target-language writing skills. To reduce the revision load at IGOs, therefore, candidates for translation posts would ideally be particularly proficient in these areas. How the profile may vary from one organization or language service to another or by type of work is analyzed in Chapter 5.

## Chapter 5 Discussion of the results of the skills and knowledge survey

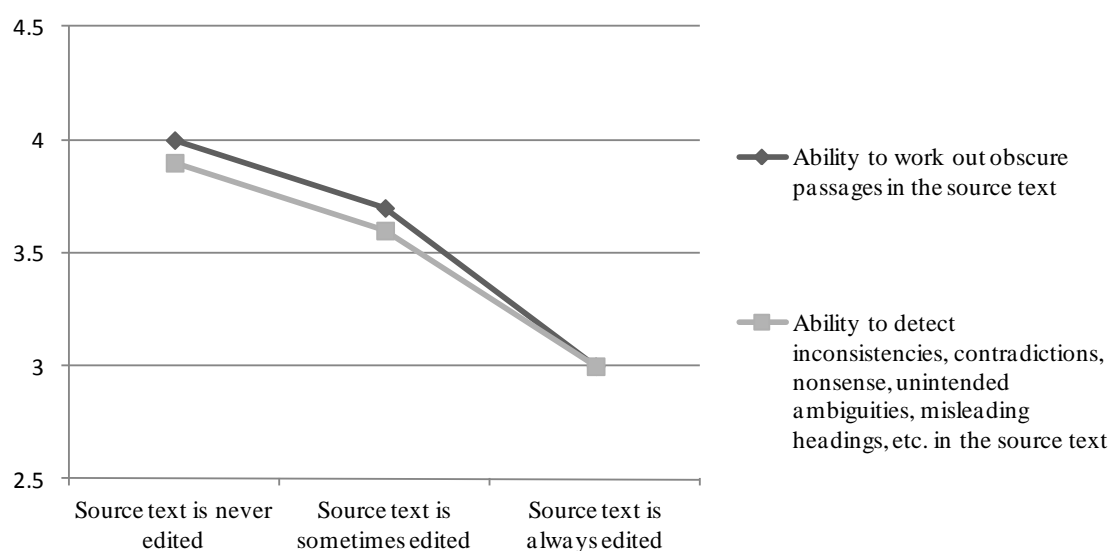
The aim of the impact and recruits questionnaires was to collect data for the identification of recruitment priorities. The response was good, with the participation of over 310 professional IGO translators, almost half of whom were revisers or heads of service, representing a broad range of IGOs and language combinations. Many of the comments made in the survey in fact offer glimpses of the challenges faced in IGO translation: the deadlines, the constraints on word choice, the concern for consistency, checking facts and usage precedents, the need to negotiate with “clients”, revisers and colleagues, the range of often technical subjects covered, the poor quality source texts, etc. These should of course be appreciated in light of the rewards of IGO translation work that have not been mentioned so far. In the pursuit of data on the skills and knowledge required, this research has highlighted the difficulties, not the benefits of being employed as a translator by an IGO, such as having decent salaries, job security, competent and motivated colleagues, a front seat on the stage of international relations, the opportunity to learn about all aspects of human activity and the satisfaction of feeling you are placing your talents at the service of a greater cause.

Apart from collecting data on the relative importance of the components of the skills-knowledge set, which we shall examine in detail in section 5.3, the survey was also used to gather information on other matters that are relevant to the text-based testing of translation competence. The data on the expectations of new recruits (section 4.4.4) revealed that candidates are rarely expected to have research and technological skills or knowledge of the organization and style guidelines prior to hiring, since those can be acquired on the job. This raises the question of how (and how efficiently) this occurs, especially as all of the components mentioned, except Excel and PowerPoint skills, were rated as being at least somewhat important (3 or above on the impact scale). That is another area for research, however; as far as we are here concerned, the implication for recruitment testing of these results is that these components do not need to be assessed during the selection process.

The data obtained on document processing at IGOs (section 4.5.4) exposed a dilemma their human resources departments face: whether to hire more editors to improve the quality of the source texts and more reference assistants, or more translators with editing (including content-editing) and research skills. The questions had been included to contextualize the findings on the relative importance of the various skills and types of knowledge required. The supposition was that, for example, if translators are not responsible for fact checking, then

research skills will be less important or, if the organization has an army of editors, then translators will not have to detect inconsistencies or nonsense in the source text. The question is whether the availability (and quality) of editing, reference and text-processing services affects the skills-knowledge set that translators need to have. Figure 5.1 shows how often the revision work attributed to a lack of abilities to work out the meaning of obscure passages and to detect inconsistencies, contradictions, nonsense, etc. varies according to the frequency with which source texts are edited.

Figure 5.1. The frequency with which the lack of editing skills among new recruits generates revision work, by the frequency with which source texts are edited\*



\*Frequency rated on a scale of 1-5, where 1 = never and 5 = always

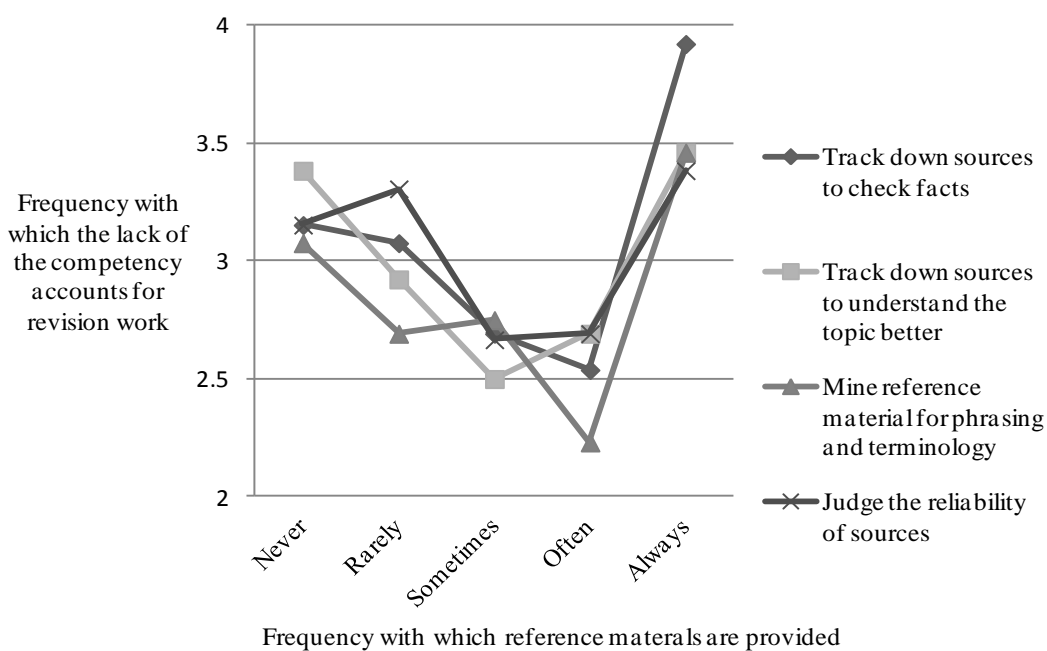
The implication is that there is a negative correlation between the provision of editing services and the amount of revision work that needs to be done. The Pearson correlation coefficient is 0.974 in the case of working out obscure meanings and 0.982 in the case of detecting inconsistencies. These editorial skills were ranked on average as the second and eighth competencies whose absence is most frequently responsible for revision work (see Table 4.3 in section 4.5.2), and organizations may (if they wish to boost the productivity and quality of their translation services) have to decide whether they are going to invest in editing services or in finding more translators with editing skills, or perhaps both. This could have major implications for training institutions.

The spread of the responses to question 2 on previously translated text (and, as revealed by closer inspection, the fact that the identification of pre-translated text is reported

as occurring both often and rarely and even always and never in the same translation service) possibly suggests that the question was interpreted in two ways: as referring to a preliminary version being produced by machine translation software or as referring to the use of translation memory software. Further research would need to be conducted to shed light on the subject.

Figure 5.2 shows how the frequency with which poor research skills account for revision work varies in relation to the provision of reference material. There is no linear correlation between the two variables, but the pattern is similar for each research skill. They all affect the volume of revision work most under the two extreme situations: when no reference material is provided, which means that the research is left to the translators; and when translators work with reference material all the time. These are the two circumstances under which weak research skills are most likely to become apparent. This shows how the frequency ratings obtained in the recruits questionnaire reflect both how much and how often skills and knowledge are required.

Figure 5.2. Average frequency with which deficient research skills generate revision work, by frequency with which reference material is provided



The fact that translators are generally not held responsible for formatting the translated text in IGOs (this is usually handled by a text processing unit) may in part explain why the ability to work with more than basic Word functions, Excel and Power Point was not ranked

highly in either the impact or the recruits questionnaires (see Table 5.1). The high ranking of spelling and adhering to in-house conventions in the impact questionnaire is similarly explained by the fact that these aspects of the translated text *are* generally the responsibility of the translator.

Table 5.1. Translators' responsibilities

<i>Are translators responsible for...</i>	<i>Yes</i>	<i>No</i>
Formatting?	21	131
Spelling?	143	9
Adhering to in-house style conventions?	137	15

For the purpose of this research, we will presume the status quo, i.e., that IGO translators at least sometimes have to work with poorly written source texts and to check references and carry out topical research themselves and they are rarely responsible for formatting. This analysis serves as a reminder, however, that the candidate profiles emerging from this research are not permanent: they are time-bound. They will vary as new technologies, new working arrangements, and more effective upstream and downstream quality control measures are introduced. The candidate profile targeted in recruitment testing must therefore be periodically updated to make sure it matches the one actually required on the job. Hence the importance of being able to perform a needs analysis and identify recruitment priorities. Awareness of the influence of different factors on the profile is essential for making the needs analysis as accurate as possible. The bearing of different variables (other than document processing and referencing responsibilities) on the candidate profile is discussed in sections 5.1 and 5.2 below.

The profiles that emerge from both questionnaires are of interest. The hierarchy obtained from the simple averages of the impact ratings reflects the opinions of 163 translators, revisers and heads of unit working in over 20 translation services. As such, it provides an indication of the relative importance that professional translators working for IGOs collectively attach to the skills and knowledge required on the job. As Kelly notes, professional considerations are one of the factors that should define the learning objectives of training programs (2005: 22-28). This information could therefore be of interest both to translators thinking of working for IGOs and to translator-training institutions. The fact that knowing how to spell correctly is apparently more important than being able to work with translation memory software, for example, may come as a slight surprise, as might the relatively low ranking of cultural knowledge. The hierarchy suggests ways in which translators could prepare themselves for work at IGOs. They might familiarize themselves,

for example, with style guides to see what aspects they cover, practice translating texts on complex topics, start gaining more than a layperson's knowledge of a certain subject or subjects, try out translation-memory software and electronic terminology tools, and work on building up speed without sacrificing accuracy. And translator trainers might wish to use the impact hierarchy to assess the content of their programs and to design materials and activities accordingly. The high rankings of coherence, grammar and spelling, for example, suggest criteria for evaluation. The need to be able to detect inconsistencies suggests new training activities that could be developed, as does the demand for editing skills.

From the recruitment perspective, however, it is necessary to know whether one size fits all or whether the profile needs to be personalized more at the organization or language-service level. The influence of these and other variables is analyzed in section 5.1 below.

The hierarchy of skills and knowledge types obtained from the average frequency ratings in the recruits questionnaire is less indicative of general needs since it reflects not only the outcome of pre-recruitment training, but of current screening procedures as well. It does give a picture of how 153 IGO revisers apparently spend their time (when revising) and, as such, suggests areas that should perhaps be targeted in the training of revisers. If individualized, i.e. made organization-specific, the hierarchy also highlights immediate in-house training needs or adjustments that need to be made to current selection processes. The influence of different variables on the frequency ratings is explored in section 5.2.

As discussed in section 3.3.8.2, in order to identify recruitment needs, the relative shortage of the different skills and knowledge types needs to be qualified by their impact on the organization's translations, just as the relative impact of the different skills and knowledge types needs to be qualified by the amount of revision work they account for among new recruits. The hierarchies are obtained by breaking down the data according to different variables and applying the method to cross-refer the frequency and impact ratings, described at the end of Chapter 3. The results are presented and discussed in section 5.3.

## **5.1 Analysis of the influence of different variables on the impact ratings**

### *5.1.1 The influence of the organization for which respondents answered*

The ways in which respondents identified the organizations they belong to (some referred to the umbrella organization, others to the specific body), the size of the sample and



confidentiality considerations precluded breaking down the survey population by more than broad organization group, i.e. into European Union bodies, the United Nations system, and what have been grouped as the Inter-American organizations (IIC, IDB and SELA). The mean impact ratings varied significantly only in seven cases, as shown in Table 5.2.

Table 5.2. Skills and knowledge for which impact ratings differed significantly<sup>14</sup>  
between organization groups  
(mean ratings awarded by group on 1-5 scale)

	European Union	Inter- American	United Nations
Knowledge of SL culture(s)	3.71		4.18
Subject knowledge	3.90	4.67*	4.42
Recast sentences in the TL	4.18		4.55
Explain translation decisions or problems	3.79		4.27
Work with translation memory software	4.22		3.76
Working with Excel and PowerPoint	2.53	3.85**	2.72
Detect mathematical errors in the ST	2.64		3.24

\* Significantly different only from EU value

\*\* Significantly different from both EU and UN values

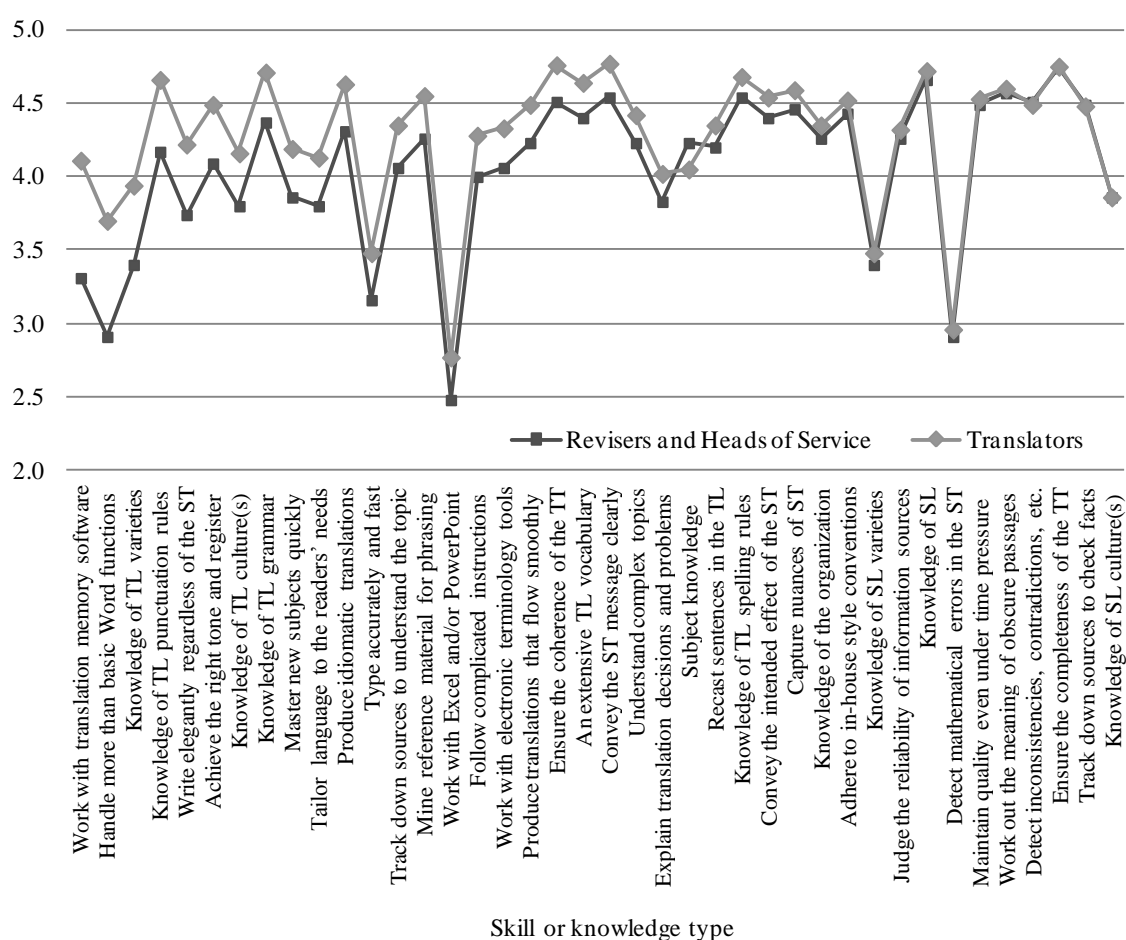
The differences in the ratings could, in the case of knowledge of the subject and source-language culture, be attributable to differences in the topics covered, and in the case of the ability to recast sentences, to the language combinations used. Translating from and into Arabic and Chinese, which, among the sample population, are solely used at the UN, probably requires far more reformulation, for example, than translating from French into Spanish. As to subject knowledge, the much higher rating by the Inter-American organizations probably reflects the fact that economic and financial texts constitute the bulk of their translation work. In the case of the ability to explain decisions, interdepartmental relations and working methods could explain the differences. How often Excel documents, PowerPoint and translation memory software are used probably underpins the impact ratings for the associated skills. This is mere speculation, however. For the purposes of this research, the differences are significant because they have an impact on the ranking of the skills and knowledge and should therefore be taken into account.

<sup>14</sup> Throughout this chapter, significant differences are calculated using t-tests for a confidence level of 95%.

### 5.1.2 The influence of position: the perspective of the reviser versus the perspective of the translator

Although the difference between the ratings awarded by translators and revisers was not significantly different in the pilot, in the full survey, translators rated the impact of all the skills and knowledge types on the list, except knowledge of the organization, higher than the revisers did. The tendency is shown in Figure 5.3.

Figure 5.3. Impact ratings, by position in service



The only statistically significant differences, however, were detected between the average ratings of the following items, which revisers rated lower in each case: knowledge of grammar (4.47 vs. 4.74), knowledge of punctuation (4.26 vs. 4.69), the ability to write elegantly (3.82 vs. 4.25) and clearly (4.65 vs. 4.80) and to achieve the right tone and register (4.18 vs. 4.52), and the ability to work with translation memory software (3.39 vs. 4.13) and

handle more than basic Word functions (2.97 vs. 3.73). The difference is hard to explain. Perhaps revisers, who spend many of their working hours correcting the grammar and punctuation and enhancing the elegance, clarity and tone of junior translators' work, have a different appreciation of the need for the skills and knowledge involved than those who strive to produce translations that meet those revisers' standards. Revisers may underestimate the importance of skills that now, as seasoned veterans, come more easily to them. They may also not view them as having such a high impact since their absence creates problems that can easily (from their viewpoint) be resolved. Revision costs money, however, and reducing the errors most commonly found in target texts is in the interests of IGOs (although possibly not in the interests of revisers?). Translators, meanwhile, may have the opposite attitude to the TL writing skills listed above: precisely because the associated errors are the ones that are constantly picked up on by revisers, translators attach more importance to them. As for the lesser importance attached to the two computer skills, could this reflect a generational divide, or the fact that revisers tend to use translation memory software less in their work and, in some organizations, still revise mainly on hard copy? One respondent commented on the "more sedate system" in place at one large IGO, "where we still use red ink on paper", saying that:

New translators in this organization need to be able to step into a time warp. We have some translators who have worked in the private sector and know all about timeliness and keeping the customer satisfied: they have trouble hiding their shock when they arrive.

The bearing of this variable on the hierarchies obtained by cross-referring the results of the questionnaires is examined to the extent possible in section 5.3 below, but the analysis of the causes of the differences between revisers' and translators' views are beyond the scope of this project and the implications are unclear. The purpose of this study is to identify the skills and knowledge that new recruits need to have. Whose opinion on the subject is worth more: that of those who are still, or have most recently been, new recruits or that of those who regularly assess translation work and translators?

### *5.1.3 The influence of the in-house/external work factor*

Only one significant difference was detected in the impact ratings awarded for in-house and external work: the ability to work with Excel and PowerPoint is much more important for external translators than for in-house translators (average rating of 4.08 vs. 2.60). This difference could have implications for test design, but its causes require further investigation first. Also, the sample size of those reporting on external translation work was quite small (16 people) and the IGOs they based their answers on were quite diverse (four different identified organizations and several unspecified ones), as could be expected.

### *5.1.4 The influence of the target language involved*

Only the differences attributable to TL (in other words, not to SL) were analyzed because the relative weight of the influence of each SL among those who work with more than one (which is almost all respondents) is unknown. For the same reason, those reporting bidirectionality (translating into and from the same language) were excluded from the analysis. Comparisons revealed only a handful of statistically significant differences: the French translators consider knowledge of other cultures far less important than the Russians do (average rating 3.5 vs. 5), and the Portuguese translators think tracking down sources to check facts was far less important than anyone else does (2 vs. 4.5-5). The members of the Portuguese service also attach far less importance to being able to judge the reliability of information sources (3.5 vs. 4-5) and to typing skills (2.75 vs. 3.14-4.25).

Other noticeable, but not statistically significant, differences were the lower ratings (possibly understandably) awarded by the Chinese translators to spelling (4 vs. 4.5-5) and punctuation (4 vs. 4.2-5). The Arabic, Chinese and Romanian translators share a relative disregard for Excel and PowerPoint skills (1.75, 1.5, 1.5 vs. 2.33-3.67), while the German translators consistently awarded higher ratings than anyone else to working with sophisticated Word functions (4.33 vs. 2.5-3.89). The Romanian translators worry least about producing idiomatic translations (4.0 vs. 4.25-5), the Finnish translators about needing to master new subjects (3.5 vs. 3.84-4.87), and the Bulgarians about writing clearly (4.5 vs. 4.63-5).

These differences are intriguing, but are not signs of major discrepancies. In fact, the results suggest that the TL is not a highly influential variable in the impact ratings. Apart from the statistically significant differences detected, answers varied most for knowledge of

SL varieties. This is not surprising considering the 24 different source languages reported in the survey sample. Not all of these exist in several varieties, as illustrated by the far lower impact rating awarded to knowledge of target-language varieties by the Hungarian, Maltese and Romanian translators (3.29, 3.83 and 3.25 vs. 4 for the Chinese, 4.2 for the Arabic, 4.63 for the Russian and 4.67 for the German translators).

## 5.2 Analysis of the influence of different variables on the frequency ratings

### 5.2.1 *The organization factor*

Respondents were grouped into EU, UN and Inter-American (IA) organizations for analysis. The IA group included IDB, IIC and SELA. The difference between the findings for the EU and IA was significant only in the case of knowledge of the subject (3.17 vs. 4), which was also rated much more highly in the impact questionnaire by these specialized institutions; and the difference between the findings for the UN and IA was only significant in the case of conveying the intended effect (higher in the UN), judging the reliability of sources (higher in the UN), and subject knowledge (higher in the IA group).

The average ratings awarded by the EU and the UN were noticeably different, however, as shown in Figure 5.4. In 15 of the cases (all the items from subject knowledge to the right of the figure), this difference was statistically significant. Figure 5.4 also shows that the UN revisers consistently rated the skills and knowledge as more frequently lacking. In fact, twice as many skills and knowledge types were reported as at least sometimes lacking (those in italics in Table 5.3) among new recruits in the UN (22 items) than in the EU (11 items). This suggests that the EU revisers are on the whole more satisfied with the performance of new recruits than their colleagues in the United Nations. This could reflect fundamental differences in the difficulty of the work, the languages involved, the quality standards set for translations, the effectiveness of the recruitment and selection process, or a combination of all these and other factors.

Figure 5.4. European Union and United Nations: average frequency with which selected skills and knowledge are found lacking

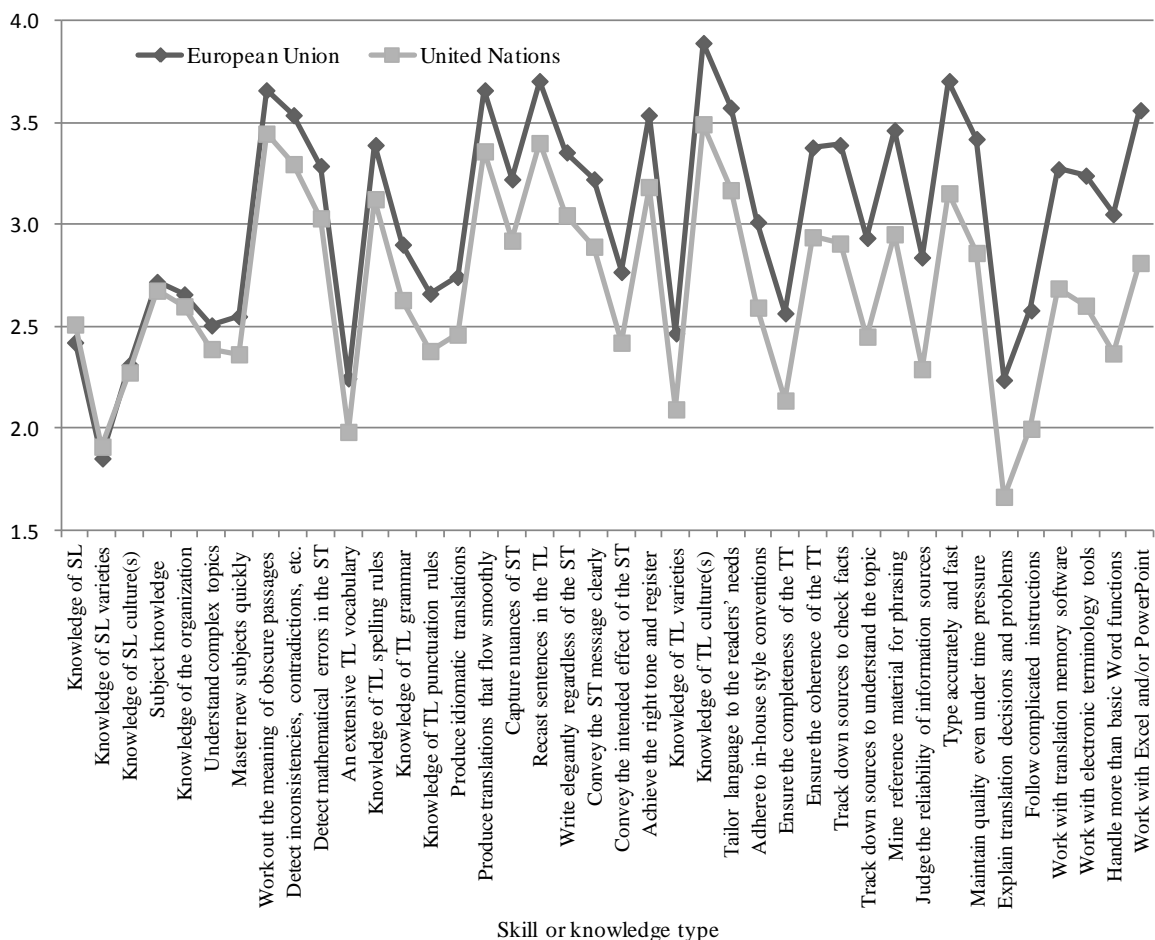


Table 5.3 shows how the difference in the ratings awarded by the two organization groups affects the ranking of the skills and knowledge types. The difference in the ranking of certain skills is considerable: note, for example, “conveying the intended effect of the source text” (ranked 18<sup>th</sup> in the EU, but 7<sup>th</sup> in the UN) and “knowledge of grammar” (20<sup>th</sup> and 28<sup>th</sup>, respectively) or “the need to be able to master new subjects quickly” (31<sup>st</sup> and 21<sup>st</sup>). The general and relative scarcity detected for each skill or type of knowledge has implications for recruitment. These findings also suggest that, to identify priorities accurately, data need to be broken down by organization.

Table 5.3. EU and UN: ranking of skills and knowledge types by frequency found lacking among new recruits (frequency ratings given in brackets)<sup>a</sup>

b	European Union	United Nations
1	<i>Produce translations that flow smoothly (3.49)</i>	<i>Produce translations that flow smoothly (3.89)</i>
2	<i>Work out the meaning of obscure passages (3.45)</i>	<i>Write elegantly regardless of the ST (3.70)</i>
3	<i>Write elegantly regardless of the ST (3.40)</i>	<i>Adhere to in-house style conventions (3.70)</i>
4	<i>Capture nuances of ST (3.36)</i>	<i>Work out the meaning of obscure passages (3.66)</i>
5	<i>Recast sentences in the TL (3.30)</i>	<i>Capture nuances of ST (3.66)</i>
6	<i>Detect inconsistencies, contradictions, etc.(3.18)</i>	<i>Subject knowledge (3.57)</i>
7	<i>Subject knowledge (3.17)</i>	<i>Convey the intended effect of the ST (3.56)</i>
8	<i>Adhere to in-house style conventions (3.15)</i>	<i>Recast sentences in the TL (3.54)</i>
9	<i>Produce idiomatic translations (3.13)</i>	<i>Detect inconsistencies, contradictions, etc.(3.54)</i>
10	<i>Maintain quality even under time pressure (3.05)</i>	<i>Judge the reliability of information sources (3.46)</i>
11	<i>Knowledge of the organization (3.03)</i>	<i>Ensure the coherence of the TT (3.42)</i>
12	<i>Judge the reliability of information sources (2.95)</i>	<i>Produce idiomatic translations (3.39)</i>
13	<i>Convey the ST message clearly (2.94)</i>	<i>Achieve the right tone and register (3.39)</i>
14	<i>Track down sources to understand the topic (2.92)</i>	<i>Convey the ST message clearly (3.38)</i>
15	<i>Achieve the right tone and register (2.91)</i>	<i>Maintain quality even under time pressure (3.35)</i>
16	<i>Track down sources to check facts (2.89)</i>	<i>Knowledge of the organization (3.29)</i>
17	<i>Ensure the coherence of the TT (2.86)</i>	<i>Mine reference material for phrasing (3.27)</i>
18	<i>Convey the intended effect of the ST (2.81)</i>	<i>Tailor language to the readers' needs (3.24)</i>
19	<i>Mine reference material for phrasing (2.69)</i>	<i>Track down sources to understand the topic(3.22)</i>
20	<i>Knowledge of TL grammar (2.68)</i>	<i>Track down sources to check facts (3.22)</i>
21	<i>Knowledge of SL (2.63)</i>	<i>Master new subjects quickly (3.05)</i>
22	<i>Tailor language to the readers' needs (2.60)</i>	<i>Understand complex topics(3.01)</i>
23	<i>An extensive TL vocabulary (2.60)</i>	<i>Explain translation decisions and problems (2.94)</i>
24	<i>Understand complex topics (2.59)</i>	<i>Knowledge of SL (2.90)</i>
25	<i>Work with Excel and/or PowerPoint (2.51)</i>	<i>Ensure the completeness of the TT (2.84)</i>
26	<i>Knowledge of TL punctuation rules (2.46)</i>	<i>Knowledge of SL culture(s) (2.77)</i>
27	<i>Explain translation decisions and problems (2.45)</i>	<i>Knowledge of TL punctuation rules (2.74)</i>
28	<i>Knowledge of SL culture(s) (2.42)</i>	<i>Knowledge of TL grammar (2.72)</i>
29	<i>Work with electronic terminology tools (2.39)</i>	<i>An extensive TL vocabulary (2.66)</i>
30	<i>Follow complicated instructions (2.38)</i>	<i>Follow complicated instructions (2.66)</i>
31	<i>Master new subjects quickly (2.37)</i>	<i>Knowledge of TL culture(s) (2.58)</i>
32	<i>Work with translation memory software (2.37)</i>	<i>Knowledge of TL varieties (2.57)</i>
33	<i>Ensure the completeness of the TT (2.29)</i>	<i>Work with translation memory software (2.55)</i>
34	<i>Handle more than basic Word functions (2.28)</i>	<i>Work with electronic terminology tools (2.51)</i>
35	<i>Knowledge of TL varieties (2.14)</i>	<i>Knowledge of SL varieties (2.47)</i>
36	<i>Knowledge of SL varieties (2.10)</i>	<i>Work with Excel and/or PowerPoint (2.42)</i>
37	<i>Knowledge of TL culture(s) (2.00)</i>	<i>Handle more than basic Word functions (2.31)</i>
38	<i>Knowledge of TL spelling rules (1.98)</i>	<i>Knowledge of TL spelling rules (2.25)</i>
39	<i>Type accurately and fast (1.91)</i>	<i>Detect mathematical errors in the ST (2.24)</i>
40	<i>Detect mathematical errors in the ST (1.67)</i>	<i>Type accurately and fast (1.85)</i>

<sup>a</sup> Where TT=target text; TL= target languages; SL=source language; ST= source text. Descriptions of skills and knowledge have been abbreviated for the sake of space. For the full descriptions, see the recruits questionnaire in Appendix 7.

<sup>b</sup> Ranking based on average ratings awarded in response to the question "Please think about the mistakes you usually correct when going over translations done by new recruits. How often do you think those mistakes are due to a lack of the following? using on a scale of 1-5, where 1=Almost never and 5=Almost always. Items in italics are those rated 3 or above.

### 5.2.2 *The skills and knowledge most often lacking, by language service*

The analysis of the distribution of responses by TL (performed on those sufficiently represented) revealed interesting but few differences between the language services. Considering all the permutations possible in the population under study (seven target languages and 40 skills and types of knowledge), the number of differences is small (26 in total, involving 15 skills). The differences described below were identified as statistically significant (using t-tests), but even then they only reveal distinctions that separate one language service from one or two others. In no skill or knowledge type is one set of new recruits significantly different from all the others.

Analytical skills were found most often lacking among the new recruits in the French services: the French translators recorded the highest frequency ratings for problems with understanding complex topics, working out obscure meanings and mastering new subjects (ratings of 3.52, 4.30 and 3.62, respectively). They also scored highest (i.e. the most often lacking) in achieving the right tone and register (3.67) and knowing different varieties of both their source and target languages (2.77 and 3.85). On the other hand, the English translators excelled in these three skills (frequency ratings of 2.81, 1.73 and 1.72). They were also reported to have the broadest knowledge of both source and target cultures (2.13 and 1.81) and to be the best at tailoring texts to the readers' needs (2.62), explaining and justifying their choices (3.00) and working with translation memory software (1.96). Those are apparently some of the comparative weaknesses of the Russian translators (3.69, 3.33 and 3.15, 3.46, 3.27, respectively). The new recruits in the Russian services also have the most difficulty conveying the intended effect of the source text (3.85), something that the Portuguese translators apparently do with relative ease (2.33). Meanwhile, the Arabic and German translators are the best at making sense of obscure passages (3.25 and 2.00 versus 4.30 for the French translators, for example). The German translators also excel at detecting inconsistencies and following complicated instructions about what to do with a text (average rating of 2.00 in both cases), unlike their counterparts in the Spanish services, who apparently have problems with the latter (frequency rating of 3.27).

These findings raise all kinds of questions. It is rather intriguing that new German and French translators seem to have complementary skills sets and that most differences arose between the English and the French and between the English and the Russian translators (in 8 and 7 skills, respectively). Further research might reveal whether the distinctions are the result of differences in the difficulty of the work given to new translators, the standards set in



each language service, technology usage, pre-recruitment training and experience, the biases of certain revisers, the real shortcomings of new recruits or the frequency with which skills and knowledge are actually required. Given the strong influence of the organization variable, these variations suggest it might make sense to break the data down more and compare language services within and across organizations. This is done in the exercise presented in section 5.3.

### 5.2.3 *The influence of the in-house/external work factor*

Unsurprisingly, external translators have more problems with adhering to in-house style, mining reference documents and mastering knowledge of the organization than do their in-house counterparts (frequency ratings of 3.48 vs. 3.35, 3.61 vs. 3.42, and 3.57 vs. 3.04, respectively). These are the skills for which statistically significant differences were found. As external translators tend to work in isolation, without day-to-day contact with the organization and its more experienced translators, and to work for several different organizations, it is much more difficult for them to acquire these skills. Less easily explained is the significantly higher frequency rating awarded to external translators for the ability to detect mathematical errors (2.68 vs. 2.55), which seems to generate more mistakes in their work than in that of their in-house colleagues. One possible reason is that over half the revisers reporting on external translators were from the more specialized organizations, many of which handle mostly technical, economic, financial and statistical texts: the IDB, ECLAC, WFP, FAO, IFAD, IMF, World Bank, EBRD, ICAO, European Patent Office. Translators at these organizations may therefore need these skills more often.

## 5.3 **The survey results as input for the identification of recruitment priorities at international organizations**

The impact ratings have been shown to vary by organization in key areas associated mainly with the topics covered and with document formatting, but not by whether the work is done in house or externally. Slight variations were found between the results for different TLs, and considerable differences of opinion were detected between revisers and translators. This suggests that when determining recruitment priorities on the basis of the impact ratings, the

IGO, TL and respondent-type variables, but not the in-house/external factor, need to be taken into account.

The frequency ratings have been shown to vary considerably by organization and slightly by TL, but only as regards skills and knowledge that are very difficult to acquire off-site in the case of the in-house/external work factor. From the staffing perspective, poor ability to adhere to in-house style and insufficient knowledge of the organization are probably better addressed through on-the-job training than testing. The IGO and TL variables are therefore those that perhaps need to be taken into account when using the frequency ratings to determine recruitment priorities.

The influence that the organization, reviser/translator and TL variables have on the hierarchies that can be obtained using the method outlined at the end of Chapter 3 is presented in Tables 5.4, 5.5 and 5.6 below. For comparison purposes, the same cut-off points of 4.5 on the impact scale and 3 on the frequency scale are used. On a scale of 1-5, it would seem reasonable to presume that an impact of 4.5 or over is high, and that a frequency rating of 3 or more suggests at least more often than not. These cut-off points make it possible to isolate the high-impact, high frequency (Category A) skills and knowledge types and classify the others as Category B (high-impact, low-frequency), Category C (low-impact, high frequency) and Category D (low-impact, low frequency) priorities for each set of data. They could be adjusted according to an individual organization's needs (to tackle productivity more, for example, an IGO could lower the cut-off point for the frequency ratings), or the results could be plotted on a scatter chart and clusters of components could be identified and weighted that way. Here the purpose, however, is to explore how different variables might affect the weightings of the components of the skills-knowledge set, and simple cut-off points are therefore used.

Table 5.4 makes it possible to compare the influence of the reviser/translator variable on the hierarchies obtained for IGOs in general and for separate organization groups.

Table 5.4. Weightings of the components of the skills-knowledge set identified for IGOs in general and individual IGO groups, on the basis of the impact ratings of revisers alone, translators alone and revisers and translators combined<sup>a</sup>

IGO, type of respondents to the impact questionnaire, number of respondents <sup>b</sup>	Category A High-impact, high-frequency	Category B High-impact, low-frequency	Category C Low-impact, high-frequency	Category D Low-impact, low- frequency
All IGOs, Revisers and translators IQ = 163 RQ= 153	8, 9, 15, 17, 20, 21, 26, 28, 29, 31, 34	1, 11, 12, 13, 14, 27	4, 5, 16, 18, 19, 22, 30, 32	2, 3, 6, 7, 10, 23, 24, 25, 33, 35, 36, 37, 38, 39, 40
All IGOs Revisers only IQ = 34 RQ= 153	8, 9, 17, 20, 21, 26, 28, 29, 34	1, 11, 12, 27	4, 5, 15, 16, 18, 19, 22, 30, 31, 32	2, 3, 6, 7, 10, 13, 14, 23, 24, 25, 33, 35, 36, 37, 38, 39, 40
All IGOs Translators only IQ = 129 RQ= 153	8, 9, 15, 16, 17, 20, 21, 22, 26, 28, 29, 31, 34	1, 11, 12, 13, 14, 27,	4, 5, 18, 19, 30, 32	2, 3, 6, 7, 10, 23, 24, 25, 33, 35, 36, 37, 38, 39, 40
United Nations Revisers and translators IQ = 68 RQ= 82	6, 8, 9, 15, 16, 17, 18, 20, 21, 22, 26, 28, 29, 31, 32, 34	1, 11, 12, 13, 14, 27	4, 5, 7, 19, 25, 30	2, 3, 10, 23, 24, 33, 35, 36, 37, 38, 39, 40
United Nations, Revisers only IQ = 20 revisers RQ= 82 revisers	4, 8, 9, 15, 16, 17, 18, 20, 21, 28, 29, 32, 34	1, 11, 12, 27	5, 6, 7, 19, 22, 25, 26, 30, 31	2, 3, 10, 13, 14, 23, 24, 33, 35, 36, 37, 38, 39, 40
United Nations, Translators only IQ = 45 RQ= 82	6, 8, 9, 15, 16, 17, 18, 20, 21, 22, 26, 28, 29, 30, 31, 32, 34	1, 11, 12, 13, 14, 27	4, 5, 7, 19, 25	2, 3, 10, 23, 24, 33, 35, 36, 37, 38, 39, 40
European Union Revisers and translators IQ = 83 RQ= 65	8, 9, 15, 17, 34	1, 11, 12, 13, 14, 20, 27, 28, 29, 31	4, 5, 16, 18, 19, 26	2, 3, 6, 7, 10, 21, 22, 23, 24, 25, 30, 32, 33, 35, 36, 37, 38, 39, 40
European Union Revisers only IQ = 9 RQ= 65	26, 34	27, 28, 29	4, 5, 8, 9, 15, 16, 17, 18, 19	1, 2, 3, 6, 7, 10, 11, 12, 13, 14, 20, 21, 22, 23, 24, 25, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40,
European Union Translators only IQ = 74 RQ= 65	8, 9, 15, 16, 17	1, 11, 12, 13, 14, 20, 21, 27, 28, 29, 31	4, 5, 18, 19, 26, 34	2, 3, 6, 7, 10, 22, 23, 24, 25, 30, 32, 33, 35, 36, 37, 38, 39, 40
Inter-American IGOs Revisers and translators IQ = 9 RQ= 5	4, 6, 8, 15, 16, 17, 18, 19, 26, 27, 28, 31	1, 11, 12, 13, 14, 20, 21, 22, 30, 34	5, 9, 29	2, 3, 7, 10, 23, 24, 25, 32, 33, 35, 36, 37, 38, 39, 40
Inter-American IGOs Revisers only IQ = 5 RQ= 5	5, 8, 9, 15, 16, 18, 26, 27, 28, 29	1, 11, 12, 13, 14, 20, 34	4, 6, 17, 19, 31	2, 3, 7, 10, 21, 22, 23, 24, 25, 30, 32, 33, 35, 36, 37, 38, 39, 40
Inter-American IGOs Translators only IQ = 4 RQ= 5	4, 6, 8, 15, 16, 17, 18, 19, 26, 27, 28, 31	1, 11, 12, 13, 14, 20, 21, 22, 25, 30, 34, 35, 36, 37, 40	5, 9, 29	2, 3, 7, 10, 23, 24, 32, 33, 38, 39

Notes:

<sup>a</sup> Numbers in the category columns refer to skills and knowledge types rated in the questionnaires, see key below. Items are listed in numerical order for ease of comparison.

<sup>b</sup> IQ= the number of respondents to the impact questionnaire; RQ= the number of respondents to the recruits questionnaire.

## Key to Tables 5.4, 5.5 and 5.6

- |  |   |
|--|---|
| 1 Knowledge of SL                              | 21 Convey the intended effect of the ST         |
| 2 Knowledge of SL varieties                    | 22 Achieve the right tone and register          |
| 3 Knowledge of SL culture(s)                   | 23 Knowledge of TL varieties                    |
| 4 Subject knowledge                            | 24 Knowledge of TL culture(s)                   |
| 5 Knowledge of the organization                | 25 Tailor language to the readers' needs        |
| 6 Understand complex topics                    | 26 Adhere to in-house style conventions         |
| 7 Master new subjects quickly                  | 27 Ensure the completeness of the TT            |
| 8 Work out the meaning of obscure passages     | 28 Ensure the coherence of the TT               |
| 9 Detect inconsistencies, contradictions, etc. | 29 Track down sources to check facts            |
| 10 Detect mathematical errors in the ST        | 30 Track down sources to understand the topic   |
| 11 An extensive TL vocabulary                  | 31 Mine reference material for phrasing         |
| 12 Knowledge of TL spelling rules              | 32 Judge the reliability of information sources |
| 13 Knowledge of TL grammar                     | 33 Type accurately and fast                     |
| 14 Knowledge of TL punctuation rules           | 34 Maintain quality even under time pressure    |
| 15 Produce idiomatic translations              | 35 Explain translation decisions and problems   |
| 16 Produce translations that flow smoothly     | 36 Follow complicated instructions              |
| 17 Capture nuances of ST                       | 37 Work with translation memory software        |
| 18 Recast sentences in the TL                  | 38 Work with electronic terminology tools       |
| 19 Write elegantly regardless of the ST        | 39 Handle more than basic Word functions        |
| 20 Convey the ST message clearly               | 40 Work with Excel and/or PowerPoint            |

The data presented in Table 5.4 show how recruitment priorities change from organization to organization and according to whether the opinions of translators or revisers or both are taken into account. The components with the most implications for recruitment testing are the Category A components, as they are important skills or knowledge types whose absence is also generating a high proportion of revision work. They are the ones that most urgently need to be found in greater abundance among new recruits, which means adjusting the focus of selection procedures or in-house training accordingly. Variations also occur between the other categories, but our analysis here will focus on the Category A items.

It is immediately clear that adjusting recruitment criteria according to individualized hierarchies rather than the all-IGOs hierarchy will change the components sought: the number will increase in the case of the UN (from 11 to 16), decrease dramatically in the case of the EU (from 11 to 5), and increase by 1 and change 6 in the case of the IA organizations.

The matter of whose opinion is taken into account also affects the weighting of recruitment priorities. If only the revisers' views are considered in the UN, instead of theirs and those of the translators together, then being able to understand complex topics, achieve the right tone and register, adhere to in-house style conventions and mine reference documents for past usage would no longer be Category A components, while subject knowledge would. It is interesting that revisers do not think that adhering to style guidelines and mining reference documents have as much impact as translators do. Maybe translators who possibly have yet to master in-house conventions still spend so much time wading

through manuals and reference material that they feel it is an important part of their work, while revisers see it as something easy to do. Such different perspectives could explain why translators rate being able to understand complex subjects highly, while revisers report a shortage of subject knowledge. The former implies being engaged in difficult tasks, the latter merely refers to specialization. If only translators' opinions are taken into account at the UN, subject knowledge would be replaced with the ability to track down sources to understand the topic: two sides in effect of the same coin. Revisers see a need for specialized knowledge; translators see a need for the ability to find the knowledge required for a particular translation. Again, it may be that translators are more focused on the process, being, as it were, in the thick of it, while the reviser view needs from a loftier end-product perspective.

The difference between the revisers' and the translators' perspective is even more striking in the case of the EU. Average opinion places in Category A two analytical skills (working out obscure meanings and detecting inconsistencies), two TL writing skills (producing idiomatic translations that capture all nuances of the original) and the ability to work fast. Adjusting recruiting priorities according to the revisers' opinions would reduce the number to two: adhering to in-house style conventions and speed. This suggests that the only changes the EU needs to make to make its revisers happy are (1) to shorten the turnaround time on the recruitment test to identify candidates who can maintain quality while under time pressure, and (2) put new recruits through a style-conventions boot camp upon arrival. Of course, some shifts in weightings in light of the other divisions may be needed (Category D skills and knowledge types might be weighted unnecessarily heavily in current testing practice, for example), but as far as Category A components and the revisers are concerned, those two measures suffice. EU translators, however, see a need for superior analytical skills and fundamental (from the IGO viewpoint) writing skills (they add "being able to produce translations that flow" to the list of high-impact skills required). Revisers see these components as missing more often than not, but not as having a particularly large impact. Translators meanwhile seem unaware of how slow they are. Of course, the number of EU revisers and heads of service who answered the impact questionnaire is quite small (9), but nevertheless possibly large enough to presume that the influence of any individual misinterpretations of the scale would be minimal. Moreover, the way in which each reviser awarded the ratings suggests that they were not unthinking in their answers. They just rate impact much lower than the translators do.

The difference between reviser and translator opinion in the IA group is much smaller. The translators in fact coincide 100 per cent with the combined opinions of both types of

respondent. Revisers do not rate the impact of “capturing nuances”, “writing elegantly regardless of the quality of the ST” and “mining reference documents for acceptable phrasing” as highly as translators do, either because they find these tasks easier or do not perform them in their job as often as translators. They attach more importance to “knowledge of the organization” and “detecting inconsistencies”, however, which generate mistakes that translators may not even be aware of, especially in the case of freelance external translators, on the basis of whom two of the five revisers in this group based their answers.

Pinning down the causes of the difference detected in the opinions of revisers and translators would require cognitive and psychological research far beyond the scope of this thesis. The difference is intriguing, but inexplicable for now. It will be up to those who make use of the hierarchies to decide whose opinion they wish to take into account or not. Most translation services would probably opt for using the viewpoints of both revisers and translators in their needs analysis or just the opinions of the revisers, but not just the opinion of translators. The first two options are therefore applied in the breakdown of the hierarchies by target language for the English and French translation services at the United Nations presented in Table 5.5. The size of the subgroups precluded making the same comparisons for the other language services.

The data in Table 5.5 suggest that the situation varies across the language services of the United Nations, and that the results of the questionnaires, if being used as input for recruitment testing, should be broken down by TL. The differences of opinion between revisers and translators involve subject knowledge, achieving tone and register, adhering to in-house style and the ability to recast sentences in the case of the English service, and TL vocabulary, grammar, writing elegantly regardless of the style of the ST, in-house style and ensuring the completeness of the TT, in the case of the French service. Whether to take these differences into account, as stated above, is a decision that will depend on the purpose at hand. The most striking implication of the data for the UN language services is that the Spanish service has far fewer adjustments to make to its selection and in-house training procedures than all the others. It only has five components in Category A, compared with 13 for English (revisers and translators combined), 15 for Arabic, 22 for French (revisers and translators combined) and 24 for Russian. Of course the distribution of components in the other categories also has implications for recruitment but, most significantly, the number of high-impact, oft-lacking skills and knowledge types varies considerably across the services, with the Russian and French revisers clearly finding more shortfalls among their new recruits than their colleagues in other services.

Table 5.5. Weightings of the components of the skills-knowledge set identified for selected language services of the United Nations

Language service, type of respondents to the impact questionnaire, number of respondents <sup>b</sup>	Category A High-impact, high-frequency	Category B High-impact, low-frequency	Category C Low-impact, high-frequency	Category D Low-impact, low-frequency
TL=English Revisers and translators IQ= 22 RQ= 22	8, 9, 15, 16, 17, 18, 20, 21, 22, 26, 28, 31, 32	1, 11, 12, 13, 14, 27, 29	4, 5, 19, 34	2, 3, 6, 7, 10, 23, 24, 25, 30, 33, 35, 36, 37, 38, 39, 40
TL=English Revisers only IQ=9 RQ= 22	4, 8, 9, 15, 16, 17, 20, 21, 28, 31, 32	1, 11, 12, 13, 14, 27, 29	5, 18, 19, 22, 26, 34	2, 3, 6, 7, 10, 23, 24, 25, 30, 33, 35, 36, 37, 38, 39, 40
TL=French Revisers and translators IQ= 16 RQ= 23	1, 4, 6, 8, 9, 11, 13, 15, 16, 17, 18, 20, 21, 22, 26, 27, 28, 29, 30, 31, 32, 34	12, 14, 38	3, 5, 7, 19, 25	2, 10, 23, 24, 33, 35, 36, 37, 39, 40
TL=French Revisers only IQ= 5 RQ= 23	1, 4, 6, 8, 9, 15, 16, 17, 18, 19, 20, 21, 22, 27, 28, 29, 30, 31, 32, 34	12, 23, 24	3, 5, 7, 11, 13, 25, 26	2, 10, 14, 33, 35, 36, 37, 38, 39, 40
TL=Spanish Revisers and translators IQ= 14 RQ= 12	8, 20, 26, 28, 34	1, 6, 12, 14, 27, 36	4, 5, 9, 15, 16, 17, 18, 19, 21, 22, 29, 30, 31, 32, 35	2, 3, 7, 10, 11, 13, 23, 24, 25, 33, 37, 38, 39, 40
TL=Arabic Revisers and translators IQ= 7 RQ= 13	6, 9, 15, 16, 17, 18, 19, 20, 21, 22, 28, 29, 30, 34, 35	1, 11, 12, 13, 24, 27, 31, 32, 36, 38	4, 8, 25, 26, 40	2, 3, 5, 7, 10, 14, 23, 33, 37, 39
TL=Russian Revisers and translators IQ= 10 RQ= 14	3, 4, 5, 6, 7, 8, 14, 15, 17, 18, 20, 21, 22, 24, 25, 26, 28, 29, 30, 31, 32, 34, 35, 37	1, 11, 12, 13, 23, 27, 36, 38	9, 16, 19, 39, 40	2, 10, 33

Notes:

<sup>a</sup> Numbers in the category columns refer to skills and knowledge types rated in the questionnaires; see key for Table 5.4 above. Items are listed in numerical order for ease of comparison.

<sup>b</sup> IQ= the number of respondents to the impact questionnaire; RQ= the number of respondents to the recruits questionnaire.

The difference between general translator and general reviser opinion in the EU was highly marked, but unfortunately the data for the EU cannot be broken down by reviser/translator opinion when exploring the influence of the TL variable because the nine revisers who answered the impact questionnaire by coincidence each work with a different TL (English, French, Spanish, German, Portuguese, Swedish, Romanian, Polish, Italian and Finnish). The breakdown by TL in the EU, shown in Table 5.6, does reveal different recruitment priorities for each language service in the organization, however, suggesting that the TL variable should be taken into account.

Table 5.6. Weightings of the components of the skills-knowledge set identified for selected language services of the European Union<sup>a</sup>

Language service and number of respondents <sup>b</sup>	Category A High-impact, high-frequency	Category B High-impact, low-frequency	Category C Low-impact, high-frequency	Category D Low-impact, low-frequency
TL=English IQ= 17 RQ= 19	8, 9, 16, 17, 18, 26, 28, 29	1, 11, 12, 13, 14, 15, 20, 21, 22, 27, 31, 32, 34	4, 5, 19	2, 3, 6, 7, 10, 23, 24, 25, 30, 33, 35, 36, 37, 38, 39, 40
TL=French IQ= 10 RQ= 14	5, 6, 7, 8, 9, 11, 13, 15, 16, 17, 20, 26, 28, 29, 30, 34	1, 12, 14, 27, 31	4, 18, 19, 21, 22, 32	2, 3, 10, 23, 24, 25, 33, 35, 36, 37, 38, 39, 40
TL=Spanish IQ= 5 RQ= 6	9, 20, 26, 27, 34	12, 37, 38	4, 5, 8, 11, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 28, 30, 36	1, 2, 3, 6, 7, 10, 13, 14, 29, 31, 32, 33, 35, 39, 40
TL=Portuguese IQ= 2 RQ= 6	4, 5, 8, 9, 17, 24, 26, 27, 28, 34	1, 20, 25	15, 16, 18, 19, 29, 30, 31, 32, 35, 37, 38, 40	2, 3, 6, 7, 10, 11, 12, 13, 14, 21, 22, 23, 33, 36, 39
TL=German IQ= 6 RQ= 9	1, 4, 5, 6, 8, 15, 16, 17, 18, 19, 20, 22, 26, 28, 29, 30, 31, 34	7, 9, 11, 12, 13, 14, 21, 23, 24, 25, 27, 32, 35, 36		2, 3, 10, 33, 37, 38, 39, 40

Notes:

<sup>a</sup> Numbers in the category columns refer to skills and knowledge types rated in the questionnaires; see key for Table 5.4 above. Items are listed in numerical order for ease of comparison.

<sup>b</sup> IQ= the number of respondents to the impact questionnaire; RQ= the number of respondents to the recruits questionnaire.

The priorities vary by number and type. Intriguingly, the Spanish section reported the smallest number of high-impact skills and knowledge types that are more often than not found missing among new recruits, as did the Spanish section in the UN. The German translation service at the EU seems to have the greatest problem finding recruits who can handle its work and, interestingly, are the only service to not find any low-impact components as more often than not lacking. Revision work in the German service thus consists mainly of making up for the lack of high-impact skills.

By the time the survey results have been broken down by organization and then by TL, the samples are quite small, but given the differences detected, such a narrow focus may indeed be justified. Recruitment decisions, unlike university training priorities, for example, should be based on particular, not general needs.

Of course, the analysis performed on the basis of the categories presented in Tables 5.4, 5.5 and 5.6 above fails to take into account significant differences in the weighting of components in the same category and treats those close to the cut-off scores in the same way as those that are major outliers. For showing the influence of the reviser/translator viewpoint



variable and the TL factor for certain cut-off points, the analysis has nevertheless been useful. To identify recruitment priorities more accurately, it would be more appropriate to start by plotting the data for a given language service on a scatter chart. This is done in Part II below, which explores the implications of the hierarchies obtained from the survey data for recruitment testing at different IGOs.

#### **5.4 Summary of the main findings regarding the skills-knowledge set required in IGO translation**

The main findings of the investigation of the skills and knowledge required in IGO translation are as follows:

1. The survey results show that translators at IGOs need far more than language skills: they also need research, computer, analytical and interpersonal skills, as well as extensive general knowledge and, possibly, specialized subject knowledge.
2. The range in the mean ratings awarded in the impact questionnaires shows that some skills and knowledge types are considered more important than others in terms of their impact on the communicative aims of the organization, thus confirming the hypothesis that:

*H<sub>1</sub>: Certain skills and knowledge types are more important than others in the context of IGO translation.*

3. Opinion on the relative importance of different components of the skills-knowledge set differs significantly between revisers and translators.
4. The range in the mean ratings awarded in the recruits questionnaires shows that some skills and knowledge types are more often lacking than others, thus confirming the hypothesis that:

*H<sub>2</sub>: Certain desired skills and knowledge types are more often lacking than others among new recruits at IGOs.*

5. Certain analytical skills were rated highly in both questionnaires suggesting that they are at least as important to find as, for example, rather fundamental components of translation competence, namely knowledge of the source language and target-language writing skills.
6. The importance of certain analytical skills (e.g. detecting inconsistencies and working out obscure passages), research skills (e.g. tracking down resources to check fact) and text-

formatting skills (e.g. working with more than basic Word functions) depends on the upstream and downstream quality-control procedures in place in the document production chain.

7. The profile that emerges from the weightings obtained from the cross-referral of the results of the two questionnaires varies from organization to organization.
8. The profile that emerges from the weightings obtained from the cross-referral of the results of the two questionnaires also varies from one language service to another within the same organization.
9. The profile that emerges from the weightings obtained from the cross-referral of the results of the two questionnaires is a dynamic entity that changes with time as the responsibilities of translators, the tools they work with, and the training they receive also change.

These findings have several implications, mainly for the training and recruitment of IGO translators. The weighted lists obtained from the two questionnaires suggest which skills and knowledge types students in translator training programs should try to develop, and which ones are particularly relevant. The weightings obtained from the recruits questionnaire are of greater interest, especially if broken down by organization, to individual IGOs for identifying in-house training priorities for new recruits and/or revisers. The rankings obtained by cross-referring the findings of the two questionnaires, meanwhile, can be used to identify possible recruitment priorities provided they are analyzed in the light of testing practice, expectations of new recruits, opportunities for post-hiring skills acquisition and the document-processing responsibilities of translators in the organization.

Since the ideal candidate profile is likely to change over time according to the competence of each generation of new recruits or with innovations in the workplace, possibly of most value to IGOs is the two-questionnaire survey method devised for identifying those recruitment priorities. Human resources officers or heads of service could apply the method to perform a needs analysis before they embark on each recruitment drive. This brings us back to the questions that initially motivated the research. How effective are current translation tests as instruments of selection? Do they screen out the translators who are best suited to the translation work at the IGO in question? The data collected in the survey make it possible to rank the components of the skills-knowledge set according to their impact on the quality and productivity of the translation service, and the resulting hierarchies can be used as a yardstick for assessing current testing practice. This line of enquiry is explored in Part II.

The method also enables IGOs to identify possible in-house training needs and is probably applicable to other specialized areas of translation work, as well as to interpreting and other fields in which the work of new recruits is closely observed. The limitations of the research presented here are discussed in the final conclusions presented in Chapter 12, after the exploration of the implications of the skills-knowledge set for recruitment testing.

## **PART II**

### **THE IMPLICATIONS OF THE SKILLS-KNOWLEDGE HIERARCHIES FOR RECRUITMENT TESTING AT INTER-GOVERNMENTAL ORGANIZATIONS**

*“In any international organization, the most important decisions are recruitment decisions. Vacant posts must be filled by the best qualified candidates possible. No organization can afford to squander a vacant post by hiring the wrong person.”<sup>15</sup>*

---

<sup>15</sup> Comment made by the head of and IGO translation service in the impact survey.



## Chapter 6 Introduction

From the recruitment perspective, determining the ideal profile is a crucial initial step in the selection process; the next is to identify candidates who match that profile. The question is whether the text-based tests currently used by IGOs are effective at filtering out the candidates who do not match that profile. In other words, do the priorities the tests reflect match those identified in Part 1 above? One way to explore the matter would be to analyze recent test papers and grading schemes. It proved difficult, however, to obtain papers and schemes from more than a handful of organizations, two of which are the object of the analysis presented in Chapter 8. There seemed to be some reluctance among many IGOs to allow their testing practice to be scrutinized too closely, possibly more from fear of flaws being exposed in their text selection and assessment methods than out of any real concerns regarding confidentiality. Informal interviews with heads of translation services at several IGOs revealed that, in most cases, there are no written or standardized test specifications or scoring schemes. The exchanges nevertheless revealed considerable interest in learning more about ways to improve translation testing and a willingness to describe their own approach. A survey was therefore chosen as the tool for collecting data on the components of the skills-knowledge set that are covered in current examinations at IGOs, in order to identify any gaps or shortcomings in that coverage. The design, application and results of the testing survey are presented in Chapter 7. The survey was followed by a detailed analysis of the text types and grading schemes used at two organizations in particular to explore if and how the weighting of components in their recruitment testing differed from that of the candidate profiles drawn up on the basis of the data obtained in Part I. The methodology used in that analysis and the ensuing results are presented in Chapter 8. To explore the implications of the findings of Part I further, the results of the testing survey and the more detailed analysis of testing at two organizations were used to design an experiment in which the performance of a group of translators on a typical IGO test was compared with their performance on a profile-adapted one, i.e. one designed specifically to detect candidates with the skills-knowledge set identified using the method devised in Part I. The preparation of the experiment is described in Chapter 9 and the results are presented in Chapter 10 and discussed in Chapter 11. Final conclusions and considerations are presented in Chapter 12.

The three main hypotheses tested in Part II are:

*H<sub>3</sub>: Not all the important components of the skills-knowledge set required for IGO translation work are tested in current recruitment examinations*

*H<sub>4</sub>: The weighting of skills and knowledge types in current recruitment examinations does not correspond to that of the profile required in the IGO in question*

and

*H<sub>5</sub>: A profile-adapted test will produce scores and rankings that are significantly different from those produced by current testing practice*

Where:

“Important” is defined as having a positive impact on the effectiveness of translations and a positive influence on productivity at the IGO in question; and

“Profile” refers to the set of skills and knowledge that translators need to have, and a “profile-adapted” test refers to one designed specifically to identify candidates with a particular profile.

## Chapter 7 The testing survey

### 7.1 Method

The testing survey had two objectives: (1) to identify to what extent the components of the skills-knowledge set identified for IGO translators in Part I are assessed in current testing practice and, if possible, to determine how they are weighted; and (2) to identify “typical” testing practice as a benchmark for a test trial.

The survey was split into two questionnaires, one on ad hoc testing and one on formal competitive examinations. This was because several organizations hold both types of examinations and they can be quite different in purpose, scale and content. As defined in the survey, ad hoc tests are those arranged whenever prospective translators present themselves for future work; and formal competitive examinations are those held for a number of candidates simultaneously, at more or less regular intervals or as the need arises.

The questions in each questionnaire were the same and covered the following topics: pre-test screening, test content, testing arrangements and test grading. All these aspects of testing would provide indications of what skills and knowledge were and were not tested. To ascertain the relative importance awarded to certain skills and knowledge in the assessment process, respondents were asked to rate the seriousness of different error types. This would make answers more comparable and provide input for the analysis of how closely the weighting of skills and knowledge types in their recruitment examination matched that of the profile emerging from the needs analysis performed in Part I.

The questionnaires were prepared in electronic format and piloted on two heads of service at one organization in March 2010. No flaws in the design of the questionnaires were identified during the pilot. The participants said they had understood and had been able to answer all the questions.

The questions asked were identical in both questionnaires and are presented in Appendix 9 for reference.



## 7.2 Results of the testing survey

### 7.2.1 Procedure

The questionnaires were sent out in April 2010 to IAMLADP focal points and the heads of translation services at other organizations.

### 7.2.2 The respondents

Replies to the questionnaire on competitive examinations were received from seven organizations, and replies to the questionnaire on ad hoc testing were received from eleven translation services. The findings of the informal conversations, consultations and interviews on the survey and testing in general, held with heads of language services and test graders from eight organizations between October 2009 and October 2010, are included as and when they shed light on the survey results.

### 7.2.3 Findings relevant to identifying the coverage and weighting of the components of the skills-knowledge set

The detailed findings of the testing survey were published in a confidential report submitted to IAMLADP in October 2010. The only findings not included here are: the answers to questions on measures taken to prevent cheating, which were not considered relevant; and the severity ratings awarded by each organization to certain error types, which were omitted to preserve the anonymity of the organizations included in the analysis in Chapter 8. The mean severity ratings are presented in section 7.2.3.10.

#### 7.2.3.1. Purpose of the examinations (Question 2)

Competitive examinations are used mainly to fill permanent in-house posts or to build a roster of suitable candidates for such posts, while ad hoc examinations are used mostly to contract translators for short-term or freelance work.

### 7.2.3.2 Pre-test screening (Question 3)

In most cases, candidates are invited to sit competitive and ad hoc examinations on the basis of their qualifications and their experience, as shown in their CVs. Some organizations use additional filters, such as general knowledge tests, to screen out unsuitable candidates prior to the translation test per se. In some cases candidates have to show adequate knowledge of the organization or current affairs or that they can write well in their mother tongue before they are allowed to take the translation test. Some components of the skills-knowledge set are thus assessed by means other than text-based testing.

### 7.2.3.3 Test content (Questions 4 - 6)

In competitive examinations, all the IGOs surveyed test translation into L1. None test translation out of L1, although one IGO has candidates demonstrate their ability to communicate in L2 by having them write a short memo or e-mail in that language. Applicants for reviser posts at two IGOs also have to revise a translation. Some UNHQ translation candidates (those recruited as translators/précis-writers) have to summarize in L1 a transcript of a speech made in L2 or L3. Applicants at another IGO have to answer general knowledge questions. The World Bank (WB) reported that on occasions it conducts two rounds of testing, the first for translation skills and the second for editing skills, even though successful applicants would be hired solely for translation work.

In the ad hoc examinations, all the IGOs surveyed test translation into L1. Two also test translation out of L1, and at the OAS, candidates can take tests into any of the four official languages (Spanish, French, Portuguese and English) as long as they can prove they have experience translating into that language. At UNOG, where freelance translators also work as précis-writers, they are asked to do a summary. At both ECLAC services, IFAD and IIC, candidates are also asked to edit a text. This was an interesting finding because, as in the WB, the translators are not being recruited to work as editors. Organizations reported having noticed, however, that translators who are also good editors are the kind of translators they want to hire. As one ECLAC head of service put it, “[w]e have found that the editing tests reveal which candidates have a keen eye for detail, sound knowledge of the finer points of grammar and punctuation, and the thoroughness to check things in our style manual, glossaries or websites. These are all skills that, when lacking, generate a considerable amount of revision work.” IGOs that include an editing task in the examination are the minority, however, and the task in question consists of editing in L1, which possibly does not require

the same skills as detecting inconsistencies when translating. This could be an area for further research.

#### 7.2.3.4 Test length (Question 7)

The typical lengths of the source texts and the durations of the examinations are presented in Table 7.1.

Table 7.1. Translation papers: number, length and time allowed for completion

Organization	Number of papers	Test paper length and minutes allowed for completion	Average total number of words	Total minutes allowed	Average words per minute
<b>Competitive examinations</b>					
EU	2	Paper 1: 45 lines (approx. 450 words) in 120 minutes Paper 2: 45 lines (approx. 450 words) in 120 minutes	900	240	3.75
UN	3	Varies from one language service to another but on average: Paper 1: 600 in 150 minutes Paper 2: 450 words in 90 minutes Paper 3: 450 words in 90 minutes	1500	330	4.55
NATO	2	Paper 1: 600 words in 180 minutes Paper 2: 390 words in 150 minutes	990	330	3.0
IMO	2	Paper 1: 600 words in 120 minutes Paper 2: 300 words in 60 minutes	900	180	5.0
WTO	3	Paper 1: 500 words in 180 minutes Paper 2: 300 words in 90 minutes Paper 3: 300 words in 90 minutes	1100	360	3.05
EPO	2	Paper 1: 300 words in 90 minutes Paper 2: 300 words in 90 minutes	600	180	3.33
WB	3	Length and time allowed not indicated			
<b>Ad hoc examinations</b>					
ECLAC(En)	1	650 words	650	unlimited	n.a
ECLAC(Sp)	1	950 words	950	unlimited	n.a
IFAD	1	600 words		unlimited	n.a
IMO	1	Paper 1: 600 words Paper 2: 300 words	900	unlimited	n.a
OAS	1	500-700 words in 90 minutes	600	90	6.7
SELA	1	250 words in 45 minutes	250	45	5.6
UNHQ	2	Paper 1: 500 to 600 words in 90 minutes Paper 2: 350 to 400 words in 60 minutes	925	150	6.2
UNOG	2	Paper 1: 400 - 450 words in 60 minutes Paper 2: 400 - 450 words in 60 minutes	850	120	7.1
UNOV	2	500 words in 90 minutes	500	90	5.6

In the shortest test, candidates translate 250 words in 45 minutes for each language offered. The United Nations competitive examination is the longest test: candidates translate 1,500 words in 330 minutes and, in some cases, write a summary or answer multiple-choice questions. UNOG and OAS seem to be looking for fast translators or rather, as one head of service put it, “translators who can, if required, come up with good solutions under pressure”.

NATO is the most generous in terms of time allowed for completing the translation. This shows that some IGOs, but not all, test the ability to maintain quality when translating under pressure, and that pressure varies considerably from one IGO test to the next. Almost half the ad hoc tests are done online and without any time constraints.

### 7.2.3.5 Text types used in examinations (Question 9)

Most organizations select texts that are typical of their translation work in terms of content and style. Only two organizations, OAS and UNHQ, do not do so in their ad hoc examinations. Nearly all organizations use general test-texts that do not require specialized knowledge of the organization or of a particular subject. ECLAC, SELA and the World Bank use only technical texts to identify translators who have specialized knowledge from the outset. They also mainly recruit translators for freelance or short-term work. In most examinations (67%) candidates are asked to translate a specialized text in addition to the general one. The organizations that use only general texts said they do so on the grounds that any necessary specialized knowledge can be obtained in-house.

### 7.2.3.6 Test-text features (Question 10)

The answers to the questions on test-text features, which consisted of rating how often test-texts contained particular elements, varied considerably from organization to organization, as shown in Tables 7.2 to 7.7 below.

Table 7.2. The typicality of test papers

	<i>content</i>				<i>style and register</i>			
	Rarely	Sometimes	Often	Always	Rarely	Sometimes	Often	Always
Competitive exams		NATO	IMO UN WB WTO	EPO EU			EU IMO NATO UN WB WTO	EPO
Ad hoc exams	OAS	UNHQ	ECL(Sp) EPO IMO UNOG UNOV	ECL(En) IFAD IIC SELA WB		IFAD OAS UNHQ	EPO IMO UNOG UNOV	ECL(En) ECL(Sp) IIC SELA WB

Table 7.2 shows that most IGOs seem to try to select texts that are typical in terms of both content and style. NATO, OAS and UNHQ are the only organizations whose translation

test papers are not even often typical of the documents translated by the translation service in terms of content. In the case of OAS and UNHQ, test-texts are not often typical in terms of style either. This could reflect differences in expectations of new recruits (for example, UNHQ may not expect candidates to have experience working for international organizations) or in what recruiters want the test to measure (general translation skills rather than the ability to handle certain types of texts). It could also reflect practical issues and the influence of test method. Most ad hoc tests are taken online and with access to the Internet, making it easier to reproduce working conditions and possibly easier to ask candidates to tackle more authentic material.

Table 7.3. The quality of writing of the source texts

*The text is ...*

	<i>well written</i>					<i>poorly written</i>				
	Never	Rarely	Sometimes	Often	Always	Never	Rarely	Sometimes	Often	Always
Compe- -titive exams			NATO WB	EPO UN WTO	EU IMO	EU	EPO IMO UN	NATO WB	WTO	
Ad hoc exams			SELA WB	ECL(Sp) EPO IFAD IIC OAS UNOV	ECL(En) IMO UNHQ UNOG	IIC UNHQ UNOG	ECL(En) EPO IFAD IMO	ECL(Sp) OAS SELA WB		

Table 7.3 shows that NATO, ECLAC (Spanish), SELA, OAS and WB sometimes use texts that are poorly written as translation test papers. WTO seems to use both types of texts or possibly mixed texts since it reported texts were often well written and often poorly written. It can be presumed that doing so is a deliberate attempt to see how candidates handle what is a common challenge in many international organizations.

Table 7.4. The style of test papers

*The text has...*

	<i>a particular rhetorical style (e.g. speeches, letters, laws)</i>					<i>stylistic features unique to the source language or culture</i>				
	Never	Rarely	Sometimes	Often	Always	Never	Rarely	Sometimes	Often	Always
Compe- -titive exams		IMO UN	EU WB WTO	EPO NATO			IMO	EPO UN WB WTO	NATO EU	
Ad hoc exams	ECL(En) IMO	ECL(Sp) EPO IFAD	IIC OAS UNOG WB	SELA UNHQ UNOV			ECL(En) SELA	ECL(Sp) EPO UNOV	IIC IMO OAS UNOG WB	IFAD UNHQ

Table 7.4 shows that over half these organizations at least sometimes select texts that have particular rhetorical features, presumably because these pose a greater challenge to the translator by introducing an additional element that candidates have to take into account. The respondent for UNOV commented, “I find that speeches provide a good test of the ability to manipulate the language and strike a particular register.” All but ECLAC (English), IMO and SELA at least sometimes select texts with source-language or source-culture-specific stylistic features. UNHQ deliberately does so, presumably to fulfill its stated aim of gauging candidates’ “culture générale”.

Table 7.5. The vocabulary of test papers

*The text contains vocabulary that is....*

		<i>straight-forward</i>	<i>difficult</i>	<i>regional</i>	<i>organization-specific</i>
Competitive exams	EPO	Rarely	Often	Rarely	Rarely
	EU	Rarely	Often	Never	Sometimes
	IMO	Often	Rarely	Never	Sometimes
	NATO	Often	Rarely	Never	Never
	UN	Sometimes	Often	Never	Never
	WB	Rarely	Sometimes	Rarely	Rarely
	WTO	Rarely	Sometimes	Never	-
Ad hoc exams	ECLAC (En)	Never	Often	Never	Often
	ECLAC (Sp)	Sometimes	Always	Never	Often
	EPO	Rarely	Often	Rarely	Never
	IFAD	Never	Always	Never	Sometimes
	IIC	Rarely	Often	Rarely	Often
	IMO	Rarely	Rarely	Never	Rarely
	OAS	Rarely	Sometimes	Rarely	Rarely
	SELA	Sometimes	Sometimes	Never	Always
	UNHQ	Never	Often	Never	Rarely
	UNOG	Often	Sometimes	Never	Never
	UNOV	Rarely	Often	Never	Rarely
	WB	Rarely	Often	Rarely	Sometimes

Table 7.5 shows that organizations adopt different tactics when it comes to the lexical difficulty of the source texts they use. IMO, NATO and UNOG stand out as the only organizations that often choose texts containing only straightforward vocabulary. Their tests are done without access to dictionaries or other lexical tools, however. The other organizations seem to want to test candidates’ handling of difficult vocabulary. IIC and ECLAC would seem to look for translators who can handle both difficult and organization-specific vocabulary. The test papers of IFAD and WB sometimes contain both challenges as well. IMO seems to try for a middle ground, having reported that its test papers rarely contain only straightforward vocabulary, yet rarely contain difficult words either. SELA is the only organization that always tests whether candidates can find and adhere to its norms of usage.

Few organizations test knowledge of regional varieties and then only rarely. WTO did not answer this question but did report that it sometimes modifies texts to incorporate vocabulary that WTO translators are expected to have.

Table 7.6. The comprehension challenges included in test papers

<i>The text contains...</i>		<i>logic errors, contradictions or inconsistencies</i>	<i>passages in which the meaning is obscure</i>	<i>subtleties that could easily be missed</i>
Competitive exams	EPO	Sometimes	Often	Often
	EU	Never	Rarely	Often
	IMO	Never	Rarely	Rarely
	NATO	Sometimes	Rarely	Sometimes
	UN	Rarely	Sometimes	Often
	WB	Sometimes	Sometimes	Rarely
	WTO	Never	Sometimes	Sometimes
Ad hoc exams	ECLAC (En)	Never	Never	Often
	ECLAC (Sp)	Rarely	Sometimes	Rarely
	EPO	Never	Often	Often
	IFAD	Sometimes	Often	Often
	IIC	Never	Never	Often
	IMO	Rarely	Rarely	Rarely
	OAS	Sometimes	Sometimes	Sometimes
	SELA	Rarely	Rarely	Rarely
	UNHQ	Never	Rarely	Often
	UNOG	Rarely	Rarely	Often
	UNOV	Never	Never	Often
WB	Sometimes	Sometimes	Sometimes	

Table 7.6 shows that EPO, IFAD, NATO, OAS and WB are the only organizations that sometimes check candidates' eye for detail and analytical skills by using source texts that contain contradictions or inconsistencies. UNHQ test papers never contain logic errors, contradictions or inconsistencies, but according to the respondent, "[m]ore often than not, the papers are aimed at testing the candidate's judgement and sense of logic", presumably on the basis of internal consistency. EPO OAS and WB test-texts also often include obscure passages, something NATO texts do not. EU test-texts contain neither inconsistencies nor obscure passages. ECLAC (English), IIC and UNOV also seem to prefer straightforward texts in which the meaning is clear. All the language services except IMO, ECLAC (Spanish) and SELA use texts that contain subtleties that might be missed by candidates who do not have a keen eye for detail.

Table 7.7. The drafting challenges included in test papers

<i>The text contains ...</i>		<i>sections that are difficult to translate into gender-neutral or politically correct language</i>	<i>culture-specific concepts that require explanation in the target language</i>	<i>sections that require considerable reformulation in the target language</i>
Competitive exams	EPO	Sometimes	Rarely	Often
	EU	Rarely	Sometimes	Sometimes
	IMO	Rarely	Rarely	Rarely
	NATO	Never	Never	Sometimes
	UN	Sometimes	Rarely	Sometimes
	WB	Sometimes	Rarely	Rarely
	WTO	Rarely	Never	Sometimes
Ad hoc exams	ECLAC (En)	Sometimes	Rarely	Always
	ECLAC (Sp)	Rarely	Rarely	Often
	EPO	Sometimes	Never	Often
	IFAD	Sometimes	Sometimes	Often
	IIC	Rarely	Never	Sometimes
	IMO	Rarely	Rarely	Rarely
	OAS	Rarely	Rarely	Rarely
	SELA	Rarely	Rarely	Rarely
	UNHQ	Sometimes	Often	Often
	UNOG	Rarely	Sometimes	Sometimes
	UNOV	Never	Never	Always
	WB	Rarely	Sometimes	Sometimes

Table 7.7 shows that it would seem that EPO, IFAD, UNHQ and WB occasionally try to measure candidates' political sensitivity, while NATO does not look for this in the recruitment process. EU sometimes tries to gauge candidates' cultural knowledge and awareness by including culture-laden concepts in the test-text. UNHQ often selects test-texts containing culture-specific concepts that require explanation. All but IMO, OAS and SELA at least sometimes use texts that contain passages that defy literal translation. UNOV does this deliberately. The respondent for UNOV said, "I look for texts that will test the translator's ability to reformulate but where the meaning is clear to anyone with a good knowledge of the language."

The answers to the questions on text type show that practice varies considerably. The texts in the competitive examinations of one organization, the EU, for instance, are always well written, sometimes contain organization-specific terms and often contain difficult vocabulary, while those of NATO are sometimes not well written, often contain only straightforward vocabulary and never contain in-house acronyms. Ad hoc test-texts at OAS, on the other hand, are rarely typical, sometimes contain obscure passages but are not particularly difficult to translate, while test-texts at IIC are always similar to those translated by the language service and are usually clearly written, but sometimes contain sections that



require considerable reformulation in the target language. Such variations could reflect different approaches to test construction, the pursuit of different skill-sets or a lack of solid criteria for translation test design.

Moreover, the vast majority (80%) of the answers about text features were “rarely”, “sometimes” or “often”. Only 7.6% of answers were “always” (these referred to texts being typical in content, register and style or well written), while 12.4% were “never” (these referred to texts being written in a particular regional variety or containing logic errors, other inconsistencies, in-house terminology or only straightforward vocabulary). This suggests that practice fluctuates not only across organizations, but also within organizations, from one language service to another in the case of those with several services, and from one exam to the next in the case of those reporting on behalf of just one language service. This could reflect either deliberate efforts to improve test content or to adjust to changes in the profile sought or a rather unsystematic, if not haphazard and arbitrary, selection of test-texts.

#### *7.2.3.7 Test-text selection and preparation (Questions 8 and 11)*

Most respondents reported that test-texts are drawn from both external sources, such as newspapers and specialized journals, and internal sources. Few organizations use solely internal documents (ECLAC, EPO, SELA, WB), and only EU and UNOG use solely external sources. Only three services, SELA, WB and NATO, occasionally provide specific translation instructions and background information (i.e. information on where the text came from, the intended reader, the purpose of the translation, etc.), a finding that may dismay advocates of approaches based on *Skopos* theory.

Most of the 11 services that modify test-texts, do so to adjust the length. Only a few alter the texts to insert particular challenges or to remove difficult parts, such as organization-specific terms. The practice is therefore not that widespread (found in 5 out of 18 services), and the comments suggest that changes are kept to a minimum and only introduced if found to be absolutely necessary.

#### *7.2.3.8 Test conditions (use of computers, provision of reference material, access to the Internet and reference material) (Questions 12 - 16)*

All the competitive examinations and some ad hoc ones bring candidates together in one room where they are supervised in order to prevent cheating. Only two competitive examinations are done using a computer (WTO and WB) and then without access to the Internet (owing to concerns about cheating). In other words, basic word-processing skills are

tested, but no research skills are activated. In the other cases, candidates have to write out their translation by hand, which means that typing and word processing skills are obviously not assessed and, possibly more importantly, conditions are far from similar to those on the job. In some competitive examinations, candidates can bring in dictionaries and glossaries.

Most ad hoc tests are taken off-site, and all off-site testing is computer-based. Candidates are free to consult any sources and are in many cases invited or specifically asked to check the organization's website, style manuals or certain documents for in-house terminology and style rules. This applies in seven out of the twelve IGOs in the survey. In the other five ad hoc examination processes, however, candidates still have to write out their examinations by hand. All the hand-written tests are timed. Only two organizations in the survey group arrange timed online tests (SELA and UNOV).

Translation services thus seem to have adopted a range of approaches towards the use of dictionaries, glossaries, thesauruses and other lexical tools during examinations: some consistently select texts that contain difficult vocabulary and they allow dictionary use, which would test research skills in some cases (e.g. IFAD) and give an advantage time-wise to candidates with superior lexical knowledge in invigilated competitive examinations, (e.g. EU); some consistently select texts that contain difficult vocabulary but the tests do not allow dictionary use, presumably to identify candidates with superior SL knowledge or the ability to guess from context (e.g. the competitive examinations at the UN and the ad hoc tests at UNOG); others choose texts that contain only straightforward vocabulary and do not allow access to resources (e.g. NATO); and others set tight deadlines for completing the paper to see how quickly and easily candidates can find good solutions (e.g. OAS).

#### *7.2.3.9 Test grading (Question 18)*

Few language services have written guidelines for grading test papers. Most work with some kind of intuitive system and assign a simple pass/fail grade. Assessment criteria and evaluation procedures, if discussed or specified at all, tend to be selected anew with each examination process. There are two notable exceptions which work with systematized grading schemes: one is a point-based system, the other uses a criterion-referenced band grading scheme, with minimum achievement levels. The corresponding organizations both provided their grading guidelines on the understanding that any reference to them would be made in such a way that they could not be identified. One of these is used for the analysis in Chapter 8.

Most organizations guard against subjectivity in grading by having test papers marked by more than one person and keeping test papers anonymous. In many language services, however, marking is left to one person, usually the head of service or a senior reviser. Only a few IGOs have written instructions and use marking sheets to try to ensure that graders apply the same criteria. No organization reported any steps to monitor the fairness of individual graders or how well they apply the established criteria.

### *7.2.3.10 General attitudes to errors in current testing practice (Question 19)*

How errors are viewed in the grading process is suggested by the answers to the questions on the seriousness of different error types. The average error ratings for IGOs in general are presented in Table 7.8).

These averages reflect the general weighting of skills and knowledge types in IGO test grading. No other error types were reported as being taken into consideration in the grading process, except at IFAD, where “[f]ollowing the source language syntax (especially for Arabic and Spanish)” will cause candidates to lose points (possibly explaining in part the significantly higher value attached to the ability to recast sentences among UN translators relative to EU translators shown in the impact questionnaire – see Table 5.2 in section 5.1.1). Both types of examinations penalize most candidates who cannot produce correctly written translations that convey the ST message in its entirety. The next filter includes more subtle errors: lack of flow, the mistranslation of individual words, loss of nuance, failure to flag an inconsistency, punctuation errors. The ranking of the severity of these error types differs quite considerably between the two types of examinations. Poor readability is viewed surprisingly leniently in the competitive examinations, for example, considering the importance awarded to flow and clarity in the impact questionnaire and the ‘clear-writing’ campaigns under way in the EU and the UN, both organizations that arrange competitive examinations. And then there are errors that seem almost forgivable inasmuch as they are associated with inside knowledge: adhering to in-house style, failure to use the correct technical term, and formatting. Of course, a multitude of minor errors could result in the same score being achieved as with just a few major ones, depending on the grading scheme applied. UNHQ pointed out that in its examination process, “no single error or category of errors would be ‘fatal’; rather a pattern of errors of a serious nature would lead to disqualification”. Other UN organizations and the EU reported that repeated errors, such as the consistent mistranslation of the same term were penalized as one error only.

Table 7.8. Average severity rating of errors across organizations\*

Severity ranking in ad hoc testing	Error type	Average rating in ad hoc examinations <sup>a</sup>	Average rating in competitive examinations <sup>a</sup>	Severity ranking in competitive examinations
1	Basic distortion of the text's overall meaning (wrong message conveyed)	4.92	4.57	1
2	Grammatical errors	4.25	3.86	3
4.5	Unwarranted additions	4.08	3.86	3
4.5	Unwarranted omissions	4.08	3.86	3
4.5	Internal inconsistencies	4.08	3.71	5
4.5	Spelling errors	4.08	3.43	6
7	Lack of flow, naturalness, readability	3.92	2.71	13
8	Punctuation errors	3.75	3.00	11.5
9	Lack of lexical precision (individual words mistranslated)	3.67	3.29	9
10	Inappropriate tone or register	3.58	3.00	11.5
11	Typing errors	3.50 <sup>b</sup>	2.60	17
12	Failure to detect, correct or point out inconsistencies or slips in logic	3.45 <sup>c</sup>	3.33	7
13	Loss of nuance, shift in emphasis, slight distortions	3.33	3.29	9
14	Failure to complete the assignment in the allotted time	3.00 <sup>d</sup>	3.29	9
15	Failure to use official/correct names for organizations, countries, acronyms	2.92	2.67	14.5
16	Failure to adhere to in-house style guidelines	2.88 <sup>e</sup>	2.60	7
17	Formatting errors	2.86 <sup>f</sup>	2.60	17
18	Failure to use the correct technical term	2.83	2.67	14.5

\*Errors rated in response to the question: How would the following errors/omissions in a candidate's test answers be viewed in your examination process? Please use the following 1-5 scale, where: 1= an irrelevant error (will not affect whether the candidate passes or fails); 5 = a fatal error (instant disqualification/elimination); and N/A = Not applicable due to the way the test is designed.

<sup>a</sup> Only includes answers when such errors are considered relevant at all (scored 2 or more)

<sup>b</sup> Not relevant at UNHQ or UNOG

<sup>c</sup> Not relevant at UNOV

<sup>d</sup> Not relevant at ECLAC, IFAD, IIC

<sup>e</sup> Not relevant at IIC, OAS, UNOG or WB

<sup>f</sup> Not relevant at ECLAC, UNHQ, UNOG or UNOV

The attitudes to error show that many of the key components of the skills-knowledge set identified in Part I are taken into account in current testing. The average ratings nevertheless conceal considerable variance in attitudes to errors across organizations. Although the answers of each survey participant are not reproduced here, in order to preserve the anonymity of the organizations that are the subject of the more detailed analysis outlined in Chapter 8, it can be reported that the organizations varied most in their ratings of the

failure to use the correct technical term and to detect inconsistencies, probably because the associated knowledge or skill is not required at every organization. They agreed that meaning errors and unwarranted omissions and additions are the most serious types of errors, and that errors in typing, formatting and using in-house style, are the least serious. There was also some coincidence on the severity of grammar, punctuation and spelling errors among those answering on the basis of competitive examinations.

Differences in attitudes to individual error types possibly reflect the type and purpose of the test in question. The mistranslation of a relatively unknown term, for example, is not viewed as seriously in an examination in which candidates were not allowed access to a dictionary as in an untimed online test, although, conversely, candidates who do get it right might be able to score positive points. Likewise, failure to format the text correctly is not viewed seriously by an organization that has a pool of text-processing assistants working for the translation service, but might be taken into consideration in organizations where translators are responsible for formatting their translations.

Average severity ratings across organizations varied from 2.56 to 4.29 on a scale of 1-5, as shown in Table 7.9. This difference in all probability reflects the differences in test conditions and expectations: the EU test, at the low end of the range, is a competitive examination procedure aiming to recruit translators for entry-level positions; the ECLAC (Sp) test at the other end is an on-line test with no time limit intended to identify experienced professionals who can produce publication-ready translations from day one.

Table 7.9. Average severity rating awarded to errors on a scale of 1-5, by type of examination

Ad hoc examinations		Competitive examinations	
ECLAC (Sp)	4.29	EPO	3.67
ECLAC (En)	4.19	WTO	3.28
SELA	4.00	UNHQ	3.22
OAS	3.78	WB	3.11
IFAD	3.65	IMO	3.06
EPO	3.61	NATO	2.72
UNOV	3.41	EU	2.56
IIC	3.29		
WB	3.28		
UNOG	3.22		
UNHQ	3.00		
IMO	2.89		

Allowing for different interpretations of the scale, the results suggest that the more technical organizations such as ECLAC and SELA, in their ad hoc testing processes, and

EPO and WTO in their competitive recruitment examinations, expect the most from candidates.

### 7.2.3.11 What organizations are looking for (Question 20)

In addition to having different attitudes to errors, organizations have different ideas about what positive features they want to find in candidates. Survey respondents were asked to comment and list, in order of importance, features of a translation (other than an absence of errors, if applicable) that could improve a candidate's chances of passing their examination. Their answers are set out in Table 7.10. Respondents were given a list of suggestions, hence the recurring wording in the answers.

Table 7.10. Positive features sought

Competitive examinations	EPO	No comment.
	EU	Strikingly apt translations, very good translations of difficult passages
	IMO	Technical and cultural knowledge, solutions to difficult problems
	NATO	Solutions to difficult problems, strikingly apt translations, fast completion
	UN	Accuracy, completeness, flair and style, evidence of a keen intellect and resourcefulness in dealing with challenges.
	WB	Solutions to difficult problems, evidence of technical knowledge
	WTO	Solutions to difficult problems, evidence of technical knowledge, fast completion
Ad hoc examinations	ECLAC (En)	Obvious familiarity with the subject matter and easy in handling it. Clever use of idiomatic expressions (adds to readability). Avoidance of calques.
	ECLAC (Sp)	Solutions of difficult problems, correction of inconsistencies.
	EPO	No comment
	IFAD	Solutions of difficult problems. Evidence of thorough research. Strikingly apt expressions. Omissions of redundancies.
	IIC	Pointing out problems in the source. Evidence of research. Elegant writing.
	IMO	Evidence of technical and cultural knowledge. Solution of difficult problems
	OAS	All of the above.
	SELA	Ability to solve difficult problems and evidence of thorough research.
	UNHQ	Ability to translate in a simple idiomatic style using readily understood terms.
	UNOG	Noticing and flagging (NOT correcting!) inconsistencies, errors in the original text, evidence of broad general knowledge, persuasiveness/ of the completed translation - admittedly, hard to define or quantify!
UNOV	Solutions of difficult problems, accurate capture of nuances and emphasis, strikingly apt expressions, evidence of technical or cultural knowledge, omission of redundancies, correction of inconsistencies, fast completion of the task, evidence of thorough research,	
WB	Knowledge of specialized subject matter, clever solutions to difficult problems	

The answers to this question reveal the slight differences in what organizations are seeking. For example, ECLAC (English), IIC and UNHQ are looking for readability; SELA for research skills; and WB for technical knowledge. WB, interestingly, rewards evidence of technical knowledge but does not, according to its error severity ratings, penalize mistranslations of technical errors heavily, suggesting a positive rather than a negative approach to grading. IIC attaches great importance to the ability to detect problems with the ST, while UNOG is looking for “persuasiveness and cogency” in the translation. Technical knowledge is a priority in the case of IMO and WTO, basic translation skills in the case of EU. The skills that UNHQ is seeking mirror the recruitment priorities identified for the organization in Part I (analytical skills and the ability to write particularly well in L1).

#### *7.2.3.12 Comments made (Question 21)*

Respondents chose to comment on different aspects of testing practice at their organization. The survey respondent for UNHQ (which is considering revamping its competitive examinations) commented on the problems with the screening process:

Examinations are scheduled in accordance with the capacities of the Examinations and Testing Service not necessarily when they are strategically needed by the ‘language department’. The screening of applications is labour-intensive and time consuming, and most likely lets through too many candidates who sit for the full-blown examination.

People involved in grading the UN examination papers expressed concerns that the criteria for test-text selection and grading varied from one year to the next, which possibly meant that candidates who failed one time around would pass the next, or that those who passed one year would not have passed the previous one. There were also concerns that the grading of the examinations was fitted in, as and when possible, around revisers’ other commitments, which possibly affected the consistency of the marking of test papers. On the whole they all agreed, however, that the UN competitive examinations were good at “finding the right people but not in sufficient numbers”. This coincides with the observations on the subject made by former head of English translation at the organization.<sup>16</sup>

---

<sup>16</sup> Interview with Stephen Sekel, former Chief of the English Translation Service and Director of the Documentation Division at United Nations Headquarters in New York. 5 October 2010.

Several organizations were worried that their test did not measure what it was supposed to “Tests do not often fit the profile” (IFAD); “Texts need to be updated and need to be more difficult” (OAS); “When we bring people in in-house, we have to rely on the parent organization’s testing procedures –which are not really geared to our particular kind of texts. That brings us (mostly) competent translators who need in-house training in our specialty” (ECLAC English). Concerns were also expressed about the lack of consistency in testing: there were calls for a “common approach”, “written guidelines” and “established criteria”.

WTO pointed out that “[c]andidates are interviewed after the written test”. Interviews are of course a fundamental recruitment tool. Some organizations make use of the interview to check knowledge of current affairs and of the organization, and most use it to check “soft-skills,” such as teamwork and other interpersonal skills. In the United Nations competitive examination for translators, the interview is considered the second part of the examination and includes a sight translation.

EU reported that its competition procedures were undergoing major changes, but that the translation tests would remain largely unchanged, except for the introduction of computer-based testing. WB and NATO reported satisfaction with their examination processes. WB was pleased with the results of its two-round testing procedure (candidates had to pass first a translation and then an editing test) “which was found to be very useful for obtaining a more well-rounded picture of the candidates”. NATO wrote: “The current feeling is that our examination system is stringent, difficult, but yields the best possible outcome. The new recruits that make it through the series of tests are reliable people who are able to pull their weight very quickly.” This is not achieved without considerable investment: “It is very time consuming though. Each test piece (we have 4 to 5 for each examination) is thoroughly researched and reviewed by the Head of Translation. Marking is also time consuming, and so are the interviews of the successful candidates afterwards.”

### **7.3 Discussion of the results of the testing survey and conclusions drawn about the coverage and weighting of skills and knowledge in current IGO testing practice**

The survey showed that several organizations are making efforts to adapt their examinations to the skills-knowledge profile they are seeking. In some cases, they are selecting specialized texts to measure subject knowledge and terminology skills. Some modify test-texts for the same reason. Many examinations now include, in addition to straightforward translation



papers, exercises such as editing, summary-writing and general knowledge tests, to reflect the demands of the job. All ad hoc tests and some competitive examinations are computer-based, and remote testing in particular closely simulates real working conditions. In some examinations, candidates are asked to work with style guidelines, consult websites and use in-house terminology tools. No organization is applying all of these practices, however, and testing procedures in fact vary considerably, especially as regards scoring procedures, which in many cases seem rather arbitrary. Several organizations could also be criticized for the lack of “authenticity” of their examinations since the source texts used and the test conditions bear little or no resemblance to the texts and conditions found in the domain. The extent to which current testing practice covers the particular skills and knowledge types that, according to the findings of Part I, make up the general ideal IGO candidate profile is summarized in Table 7.11.

Table 7.11 The extent to which components of the skills-knowledge set are covered in current testing practice, skills and knowledge grouped by type.

Skills/knowledge group*	Assessment in testing
<p><b>TL WRITING SKILLS:</b> the ability to select and combine words in the target language to capture the exact and detailed meanings (nuances) of the source text (17) (grouped with TL writing skills because SL and SC knowledge was not found lacking much among new recruits), the ability to produce idiomatic (natural-sounding) language in the target text (15), the ability to produce translations that flow smoothly even when the source text does not (16), the ability to recast sentences in the target language (to say the same thing in different ways) (18), the ability to produce an elegantly written target text regardless of how elegantly written the source text is (19), the ability to convey the source-text message clearly (20), the ability to convey the intended effect of the source text (21), the ability to achieve the appropriate tone and register in the target text (22), the ability to write correctly (12, 13, 14) and adequate TL vocabulary (11)</p>	<p>All organizations look for the highest level of accuracy (including nuance, intended effect and tone) and the ability to express the content of the ST in language that has the features listed here. One can presume that the ability to recast sentences is scored positively in the revision and editing tasks that are included in some examinations. Less than one third of the IGOs assess the ability to produce well-written TL versions of poorly written originals and then only sometimes, despite this being a common aspect of translation work at international organizations. Adequate word choice, grammar, punctuation and spelling are routinely assessed.</p>
<p><b>ANALYTICAL SKILLS:</b> the ability to detect inconsistencies, contradictions, nonsense, unintended ambiguities, misleading headings, etc. in the source text (9), the ability to work out the meaning of obscure passages in the source text (8) (grouped with analytical skills because SL and SC knowledge was not found lacking much among new recruits) and the ability to ensure the coherence of the target text (e.g. consistent terminology use, no contradictions, logical connections of ideas) (28)</p>	<p>None of the organizations that we have surveyed systematically tests the ability to detect inconsistencies in a ST or to work out obscure passages, and many never do so. Various aspects of coherence are often tested, and candidates are usually penalized for lack of internal consistency.</p>
<p><b>KNOWLEDGE:</b> subject knowledge (technical knowledge, e.g. of economics, international law, science, technology) (4), knowledge of the organization and how it works (5), general knowledge/knowledge of world</p>	<p>Subject knowledge is sometimes tested, but only through the ability to translate a text on the topic. Knowledge of the organization is sometimes tested in interview or, in the case of the EU, in a separate</p>

Skills/knowledge group*	Assessment in testing
affairs (repeatedly mentioned in the survey as a requirement), knowledge of SL (1) and knowledge of the SC (3).	test. General knowledge is similarly sometimes gauged in the translation task, separately or in interview, but not systematically. Knowledge of the SL and the SC is assessed indirectly through the measurement of accuracy in the translation.
<b>RESEARCH SKILLS:</b> the ability to track down sources of information to check facts (29), the ability to track down sources to obtain a better grasp of the thematic aspects of a text (understand the topic) (30), the ability to mine reference material for accepted phrasing and terminology (those used by the organization or in a specialized field) (31) and the ability to judge the reliability of information sources (32)	Another vital aspect of translation work at organizations (especially skill 31), these research skills are tested only by a few organizations in online tests for freelancers, and then not systematically. They are not considered to be as important or as rare among new recruits as are the other items presented in this table. A high proportion of respondents in the surveys commented that research skills could or have to be developed in-house, reflecting the fact that the sources and search tools employed are very organization-specific.
<b>OTHER SKILLS:</b> the ability to maintain quality under time pressure (34), adhere to in-house style (26) and interpersonal skills (teamwork, work with and in different cultures, etc.)	Most competitive examinations for in-house posts and some online tests consist of translation tasks with short turnaround times. Many organizations, however, provided generous amounts of time for completion of translation tasks. Only a few organizations expect candidates to demonstrate their ability to adhere to in-house style in their online tests for freelancers. Interpersonal skills are usually assessed in an interview.

\* Numbers are those used to identify skills and knowledge types in Part I

As to whether current examinations identify the right candidates, the testing survey results show that not all the important components of the skills-knowledge set are covered in current recruitment exams, thus confirming our third hypothesis. The most significant omission is the non-measurement of analytical skills. Other omissions, such as the non-measurement of research skills and the ability to adhere to in-house style, are more understandable since the skills in question are dependent on organization-specific knowledge or experience.

As noted above, how IGOs attempt to measure candidates' skills and knowledge varies considerably, not only from organization to organization, but even within the same organization. Unlike the translation work itself, the construction of translation tests is not a particularly regulated activity. Beyond the length and type (general or specialized) of the ST, there are no task specifications to guide test writers, and written specifications of grading schemes are rare. One notable exception was the case of one IGO that has standardized the grading scheme applicable to translation tests in all language combinations. The non-standardization of procedures seriously undermines the reliability of testing, as the IGOs themselves are most aware. "A greater attempt needs to be made to ensure a greater degree of

harmonization of testing practices across translation units as these vary to a certain degree from one unit to another,” noted one respondent, and another made a similar comment: “Each language section goes about it in its own way. We have no common guidelines or framework for establishing criteria.” The lack of standardization and reliability has resulted, in many cases, in IGOs only selecting those candidates who perform *exceptionally* well in the recruitment tests. Hence the less than 10 percent pass rates for some competitive examinations reported in the introduction to this thesis. Not only does the lack of reliability in translation testing force organizations to keep the bar very high, but it also raises questions about fairness. It may be easier for translators offering certain language combinations to obtain jobs, and translators who pass an exam one year might not have done the next and vice versa. Such perceptions undermine the credibility of the system.

Apart from reliability problems, the tests also raise validity concerns. Tests mainly consist of translating a well-written ST that is not wholly typical of the type of document translated by the service in question. In several cases candidates still have to write out their translation by hand and, even if they are asked to produce their TTs using a computer, they are sometimes denied access to the Internet or other resources. Test conditions are therefore often wholly unauthentic, which may have a major impact on the validity of inferences made about readiness for the job. The lack of similarity between test tasks and the tasks performed on the job raises doubts about whether all the components of the skills-knowledge set required are covered in examinations. Can they be considered to be being measured if they are assessed in a wholly different context? The experiment described in Chapter 9 below has attempted to begin to answer that question.

The survey also shows that only a few IGOs modify any aspect of test-texts other than length and then only sometimes. Text-selection procedures therefore mostly reflect a “real-life” approach to task design, which fails to optimize the “interactiveness” of the task, as Bachman (1990: 25) would put it. In many instances, test-texts are taken from external sources. If test-texts are different from the kind of texts that the organization usually translates, test graders may need to bear in mind that they should be identifying and assessing transferrable skills rather than evaluating the quality of the final product, unless they believe that the translators they are looking for must be able to translate all kinds of texts well. Test graders may be aware of this: of the 16 who answered the question “Do you evaluate the quality of translations only, or do you also evaluate the way the candidate translates?”, only three said they evaluated the quality of the translation only. All the others claimed to assess both the translation and how the translator translated. How they went about this was not made

clear, however, and might deserve further investigation. First, it should probably be clarified how they had understood the question, but the answers suggest that graders try to find evidence of skills and knowledge, rather than wholly acceptable translation solutions.

In terms of attitude to error, the general impression from the survey is that priority is awarded to candidates who have mastered at least fundamental translation skills (accuracy, grammar, spelling). All things being equal in that regard, the next level of skills and knowledge types is taken into consideration: the ability to capture exact nuances in smooth flowing prose, make particularly effective word choices and have a keen eye for detail. Attitudes to errors varied considerably among the IGOs participating in the survey, however, which suggests that candidates who do not succeed in the examination set by one organization might succeed in the examination set by another. As in the case of test-task variations, these differences between organizations are a sign of good practice if they do indeed respond to different recruitment needs; in other words, a horses-for-courses approach. They are a source of concern, however, if they reflect subjective approaches to test-task design and performance assessment that bear no relation to the candidate profile the organization should be seeking.

The survey shows that testing practice, in terms of the test task, attitudes to error and the grading schemes used, varies considerably from one IGO to another and even within the same organization. The weighting of the skills-knowledge set in current testing practice therefore needs to be analyzed at the level of the individual translation service. This is done in Chapter 8.



## **Chapter 8 Current testing practice: a closer look at two organizations**

### **8.1 Method for the analysis of testing practice at two organizations**

The purpose of the close analysis of testing at particular organizations was to compare the weightings of skills and knowledge types in their recruitment examinations with the weightings emerging from the results of Part I of the research. Two organizations were selected for the analysis: one because considerable information had been provided in the testing survey and in the follow-up communications with the person responding on behalf of the organization, and the other because I had access to those involved in test design and grading. Both asked to remain anonymous and are referred to as organization X and organization Y.

In the case of organization X, the method consisted of analyzing (1) the organization's responses to the testing survey, (2) two test papers it had used within the preceding 18 months and (3) the detailed written guidelines the organization issued to graders, with a view to determining how skills and knowledge seemed to be weighted in current tests. That weighting was compared with the hierarchy from the scatter graph obtained by cross-referring the results of the impact questionnaire and the recruits questionnaire for the organization.

In the case of organization Y, the same method was used but since there were no written guidelines for graders, the responses to a short survey of the graders at the organization were analyzed instead. In that survey, respondents were asked to rate how heavily they weighted evidence of different abilities and knowledge within the holistic grading scheme they used (see Appendix 10). The questionnaire was piloted on one grader before being distributed in January 2011.

### **8.2 The weighting of skills and knowledge types in testing at organization X**

In the large-scale competitive examinations at this organization, candidates take two translation papers, one from L2 and the other from L3; each consists of a translation of 450 words that has to be completed in 120 minutes. The texts cover general topics only, although the samples I obtained both dealt with issues facing the institution and its future, and knowledge of the organization might have been helpful to the candidates. Texts are taken

exclusively from external sources such as newspapers, magazines or journals. Texts are always well-written and often contain difficult vocabulary. The texts never contain logic errors, contradictions or inconsistencies, but often contain subtleties that could be missed. They sometimes contain sections that require considerable reformulation and culture-specific concepts that require explanation in the target language. The texts were reported to be fairly typical in terms of content, style and register. However, those that I examined involved an extended metaphor in one case and figurative language in the other, both typical of a news magazine editorial but certainly not of institutional texts. The tests are currently done on paper but the plan is for them to be done on computers in the near future. At the moment no resources, dictionaries, glossaries, etc., may be brought into the examination; previously they could.

Graders are given detailed written instructions and grading criteria, of which I obtained a copy on the understanding that it would not be made public. Graders are asked to look for candidates who have perfect mastery of their main language and a profound knowledge of the SLs. They must determine whether candidates have worked out exactly what the author was trying to say, have the necessary reasoning skills and are able to find satisfactory or even ingenious solutions.

The suggested procedure is as follows: after familiarizing themselves with the ST, graders first read through the translation to gain a general impression of how much the translation reads like an original, how clear the style and expression are, and whether a coherent TT has been produced. Next they rate it on a scale of 1 to 5 (where 2 is satisfactory), on the basis of the following criteria: comprehension of vocabulary and structures in the ST, conveyance of the sense of the original and style. The descriptors given for each point on the scale indicate the level of achievement in each of these three areas, but make no reference to coherence. The scheme has two other shortcomings: it does not allow for the possibility that candidates may score highly on style but very poorly on comprehension or vice versa; and the descriptors of each band do not always refer to the same thing. The level 2 descriptor, for example, refers to attempts to reflect stylistic features of the ST, while level 4 refers to the maintenance of consistent register.

The descriptors reveal the relative weighting of the three criteria (the actual wording has been paraphrased for the sake of confidentiality, unless otherwise indicated). According to the descriptor for level 3 on the scale, for example, candidates are expected to display a profound understanding of a broad range of vocabulary and structures and clearly plausible efforts to overcome specialized vocabulary problems. At the same time, the translation should

sound natural except when syntax is difficult. This shows that SL knowledge is weighted slightly more than meaning transfer skills and idiomaticity. And the descriptor for the top level shows that meaning transfer is more important than style because the meaning must be fully conveyed, including all subtleties, while one or two “natural” (original wording) flaws should be allowed in the style.

Next the graders assess the translation sentence-by-sentence against the original and make notes in the margin of the kinds of mistakes detected, as well as of any particularly ingenious solutions. This is useful for assessing inter-rater reliability since detailed comparisons of scoring can be made (although I do not know if this is done). The candidate is awarded a score on the basis of a positive and negative points scheme, and the pass mark is set before grading begins. Negative points are awarded according to the type and seriousness of the errors. The scales vary by error type (1-2 points in some cases 1-3, 1-4, and 1-6, in others). Positive points (2 or 4) are awarded for particularly ingenious solutions. In other words, a neat solution can earn 4 positive points and thus compensate a fairly major meaning distortion or a serious grammar mistake, which score up to 6 and 4 negative points respectively. The problem with this points-based scheme, as noted in the literature review, is that it is highly dependent on the level of difficulty of the test, and can only possibly work if the examination is exhaustively tried and tested to ensure that the points scheme and cut-off score help classify candidates usefully. This trialing would have to be done every time the source text is changed. Alternatively, source texts would have to meet the same set of highly specific criteria regarding the number and type of challenges they include, every time the examination is held.

The final point score is compared with the initial 1-5 rating on the basis of a predetermined equivalence scheme. According to the grading guidelines, there should not be a huge discrepancy between the results obtained using the two schemes. The validity of this equivalence would need to be demonstrated. The two schemes seem to reflect different approaches to assessment, a positive and a negative one: the rating scale admits the possibility of a candidate’s target text being publication-ready, while the points scheme sets a maximum possible score, and moreover sets it at the level that candidates start out with. In other words, the assumption is that negative points for error will always outweigh any positive ones for ingenious solutions.

Finally, graders are given the opportunity to recommend whether a candidate should pass or fail regardless of the rating level or point score achieved. They can also comment on any discrepancies between the initial rating and the point score. This is, presumably, a



safeguard measure designed to compensate for the shortcomings of the two marking schemes. For example, a candidate who comes up with ingenious ways of conveying the meaning of the original but has poor grammar and spelling will score the highest mark on the points-based scheme. It will have been impossible to assign the same candidate the highest rating in the more impressionistic first rating, however, because the scale descriptors refer to achievement in both meaning transfer and style. The graders will therefore have to overrule the grading scheme and apply their own criteria.

An examination of the points that can be awarded for each error type reveals that the components of the skills-knowledge set that are assessed, using the definitions from the skills-knowledge survey, are weighted, from heaviest to lightest, as follows:

- The ability to capture detailed levels of meaning (meaning errors: 1, 2, 4 or 6 points; unnecessary additions: 1 or 2 points);
- The ability to write correctly (grammar and syntax errors: 1, 2, or 4 points);
- The ability to write clearly and idiomatically using the appropriate register (style, clarity, register errors: 1, 2 or 3 points);
- The ability to ensure that the target text is complete (omissions: 1 or 2 points) and to spell and punctuate correctly (errors: 1 or 2 points).

This largely coincides with the findings of the testing survey, in which error types were rated on a scale of 1-5 by this organization, as follows:

- Severity rating 4: basic distortion of the text's meaning and unwarranted additions and omissions;
- Severity rating 3: loss of nuance, shift in emphasis, slight distortions; lack of lexical precision (individual words mistranslated); grammatical errors; spelling errors; punctuation errors; failure to complete the assignment in the allotted time; inappropriate tone or register; internal inconsistencies;
- Severity rating 2: lack of flow, naturalness, readability; failure to use the correct technical term, failure to use official/correct names for organizations, countries, acronyms.

The most noticeable differences are that omissions, punctuation and spelling are penalized more heavily in the survey responses than in the grading scheme, and that

naturalness is valued less. Also, lack of internal consistency is penalized as heavily as lexical precision and register, according to the survey results, but is not mentioned in either the first-impression rating (except for a reference to consistency in stylistic register in one descriptor) or the points schemes applied in practice. It should be noted that the testing survey respondent was possibly not a translation test grader, but a member of the human resources unit that administers examinations for the organization. This could account for these discrepancies. In all the other organizations, graders answered the questionnaire.

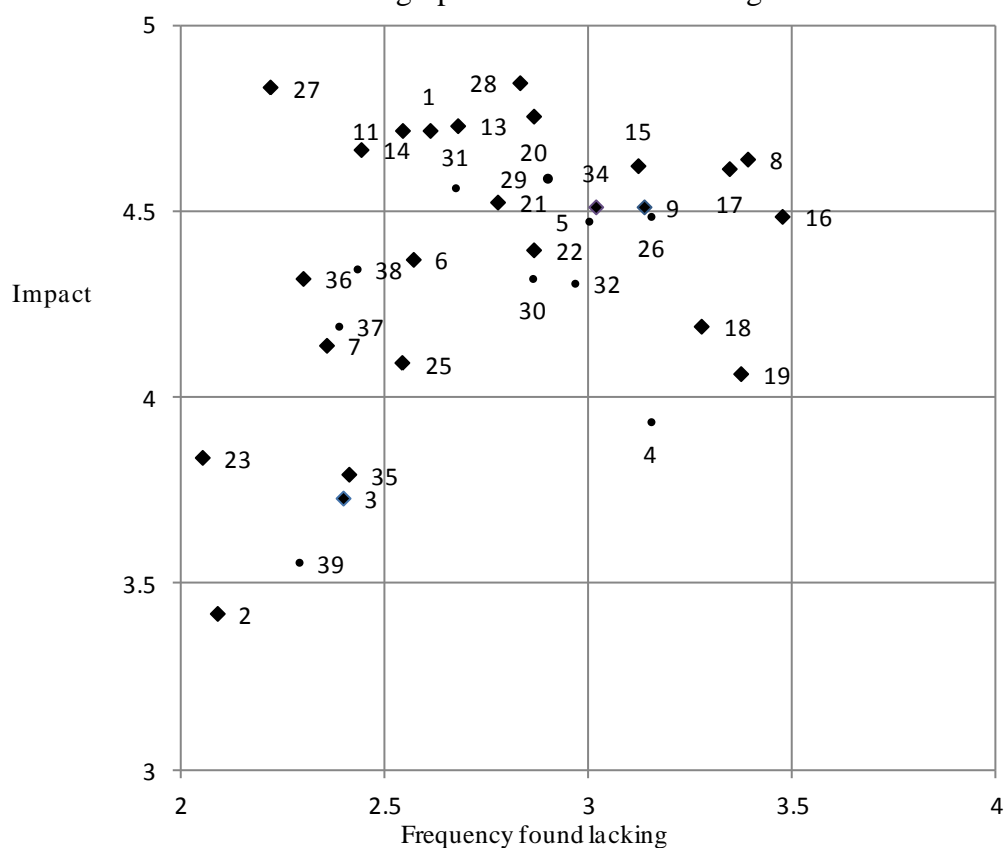
Based on the scheme that graders actually work with, cross-referring the information obtained from the instructions, the rating scale descriptors and the points system presented in that scheme, and in light of the STs analyzed, we can infer that, at this organization, skills and knowledge are weighted as follows in current testing practice, from heaviest to lightest:

- A. SL knowledge and the ability to capture detailed levels of meaning;
- B. TL grammar skills and the ability to produce translations that flow smoothly;
- C. The ability to write clearly, using natural-sounding TL and the appropriate register;
- D. Spelling, punctuation and the ability to ensure the completeness of the TT.

Since the grading scheme at this organization is applied by all the different language services and the same ST is used in all the examinations for translators with TLs other than the language of the ST, these weightings need to be compared with those of the general profile for translators from all the language services except the SL one at this organization. This was obtained by cross-referring the corresponding responses to the impact and recruits questionnaires for all respondents whose TL was not Spanish (since the sample test-texts examined were in Spanish) and plotting the average impact and frequency ratings for the skills and knowledge types that candidates were expected to have at the time of recruitment on a scatter chart, as shown in Figure 8.1. In the impact questionnaire, the respondents for this organization reported that the following skills and knowledge types are expected to be acquired after hiring at this organization (the numbers in brackets refer to the number used to identify the skill or knowledge type in the scatter chart; the percentages refer to the proportion of respondents who reported them as not being required prior to recruitment): work with translation memory software (37, 63.0%); work with electronic terminology tools (38, 55.6%); adhere to in-house style conventions (26, 40.7%); knowledge of the organization (5, 27.8%); handle more than basic Word functions (39, 14.8%); track down sources to check facts; mine reference material for phrasing (29, 31, both 13.0%); and subject knowledge (4,

9.3%). These are identified by dots instead of diamonds in the figure. It should be noted that, although the reviser/translator viewpoint variable was shown to have a bearing on the impact ratings, it was decided to include the opinions of both revisers and translators, on the premise that the translators' perspective might be of value.

Figure 8.1. Impact and frequency ratings for selected skills and knowledge types used in translating Spanish source texts at organization X



Key to Figure 8.1 (for full definitions, see Appendix 11)

- |  |   |
|--|---|
| 1 Knowledge of SL                              | 21 Convey the intended effect of the ST         |
| 2 Knowledge of SL varieties                    | 22 Achieve the right tone and register          |
| 3 Knowledge of SL culture(s)                   | 23 Knowledge of TL varieties                    |
| 4 Subject knowledge                            | 24 Knowledge of TL culture(s)                   |
| 5 Knowledge of the organization                | 25 Tailor language to the readers' needs        |
| 6 Understand complex topics                    | 26 Adhere to in-house style conventions         |
| 7 Master new subjects quickly                  | 27 Ensure the completeness of the TT            |
| 8 Work out the meaning of obscure passages     | 28 Ensure the coherence of the TT               |
| 9 Detect inconsistencies, contradictions, etc. | 29 Track down sources to check facts            |
| 10 Detect mathematical errors in the ST        | 30 Track down sources to understand the topic   |
| 11 An extensive TL vocabulary                  | 31 Mine reference material for phrasing         |
| 12 Knowledge of TL spelling rules              | 32 Judge the reliability of information sources |
| 13 Knowledge of TL grammar                     | 33 Type accurately and fast                     |
| 14 Knowledge of TL punctuation rules           | 34 Maintain quality even under time pressure    |
| 15 Produce idiomatic translations              | 35 Explain translation decisions and problems   |
| 16 Produce translations that flow smoothly     | 36 Follow complicated instructions              |
| 17 Capture nuances of ST                       | 37 Work with translation memory software        |
| 18 Recast sentences in the TL                  | 38 Work with electronic terminology tools       |
| 19 Write elegantly regardless of the ST        | 39 Handle more than basic Word functions        |
| 20 Convey the ST message clearly               | 40 Work with Excel and/or PowerPoint            |

The relative position of items on the scatter chart suggests that the different skills and knowledge types should be grouped as follows:

*Category A: High-impact, high-frequency components that need to be heavily weighted in recruitment (minimum rating of 4.5 on the impact scale and of 3 on the frequency scale).*

The components in this category are: detect inconsistencies, contradictions, etc. (9), work out the meaning of obscure passages (8), produce idiomatic, smooth-flowing translations (15, 16), capture nuances of the ST (17) and maintain quality under pressure of time (34).

*Category B: High-impact, low-frequency components that are either being adequately tested, or are commonly found among new recruits, or are not required often (minimum rating of 4.5 on the impact scale but less than 3 on the frequency scale).*

The components in this category are: convey the ST message clearly (20), ensure the coherence of the TT (28), knowledge of SL (1), ensure the completeness of the TT (27), knowledge of TL grammar (13), knowledge of TL punctuation rules (14), knowledge of TL spelling rules (12), an extensive TL vocabulary (11) and convey the intended effect of the ST (21).

Organization X should continue to check candidates have these skills and knowledge types, but they do not need to be weighted as heavily as those in category A, except possibly 20 and 28 (achieving coherence and clarity), which are considered of great importance. The positions of 12 and 27 on the chart suggest that spelling mistakes and omissions rarely account for revision work and are already being adequately tested.

*Category C: Low-impact, high-frequency components that are commonly lacking among new recruits but not rated as important as those in category A (less than 4.5 on the impact scale but more than 3 on the frequency scale).*

There are only two components in this category: write elegantly regardless of the ST (19) and recast sentences in the TL (18).

The absence of these TL writing skills generates a sizeable proportion of revision work, but does not apparently have as much impact on the effectiveness of translations. They need to be tested, but not to be weighted as heavily as those in category A. This may require adjustments in current testing practice.

*Category D: low-impact, low-frequency components (less than 4.5 on the impact scale and less than 3 on the frequency scale).*

There are several components in this category, which lends itself to being broken down into subgroups according to their position on the chart: (a) achieve the right tone and register (22), track down sources to understand the topic (30) and judge the reliability of sources (32); (b) understanding complex topics (6), mastering new subjects quickly (7), tailor language to the reader's needs (25) and follow complicate instructions (36); (c) knowledge of TL and SL culture(s) (3, 24,) TL and SL varieties (2, 23), explain translation decisions and problems (35) and type accurately and fast (33); and (d) detect mathematical errors (10) and work with Excel and/or PowerPoint (40).

These skills and knowledge types are being found. If they are being tested, the organization should check they are not being weighted unnecessarily heavily in the grading scheme used. They should not be weighted more than the skills and knowledge types in category B. If they are not being assessed they do not need to be incorporated into testing, with the possible exception of the research skills and ability to achieve tone and register in subgroup (a), which came close to being classified in the other categories.

A comparison of the categorization of the components of the skills-knowledge set outlined above and the weighting of competencies reflected in the grading scheme reveals certain differences, which suggest that in order to find more translators with the profile that it needs, organization X should adjust its examination to:

- Shorten the time allowed for task completion, in order to find translators who are good at maintaining quality under time pressure;
- Measure the ability to detect inconsistencies and work out the meaning of obscure passages;
- Measure the ability to produce translations that flow and have a good style even though the ST does not;
- Measure the ability to capture nuances more as a TL skill than as evidence of SL knowledge, since the absence of the former is generating revision work and the latter is more easily found;
- Award lower priority to register;
- Possibly assess the ability to track down information to understand topics and to judge the reliability of sources.

In addition to adjusting the weighting of the grading scheme, this would involve two innovations: measurement of the ability to detect inconsistencies, and assessment of the ability to produce translations that flow and have a nice style even though the ST does not. The latter is interesting inasmuch as several translators at the organization commented in the impact and frequency questionnaires on the need to handle poorly written STs. In the light of the findings in the literature review on the importance of domain relevance in proficiency testing, organization X should probably increase the authenticity of the test task in other ways, such as by using less atypical STs in its translation examinations.

### **8.3 The weighting of skills and knowledge types in testing at organization Y**

In the smaller-scale, less formal testing of translators at organization Y, candidates translate one L2 ST and one L3 ST, each 400-450 words long, in 60 minutes in each case. This is half the time allowed in organization X. The texts, however, are in many ways similar in type to those used at organization X. They cover general topics only and are taken exclusively from external sources such as newspapers, magazines or journals or specialist websites. The texts were reported in the testing survey to be often typical in terms of content, style and register, although the texts that I had the opportunity to examine were rather atypical inasmuch as they were news articles that contained a high proportion of direct speech. Translators at this organization rarely, if ever, have to translate the kind of language used by the person in the street, nor are they allowed to use a journalistic style in their work. Rather than knowledge of the organization, these texts call for knowledge of current affairs. Informal interviews with graders at the organization revealed that this was the norm. As in organization X, test-texts are always well-written and sometimes contain difficult vocabulary. The texts rarely contain logic errors, contradictions or inconsistencies, but often contain subtleties that could be missed. They sometimes contain sections that require considerable reformulation, and culture-specific concepts that require explanation in the target language. The tests are done on paper, and no resources, dictionaries, glossaries, etc., may be brought into the examination. Typically between two and five candidates take the examinations at a time, which are held once or twice a year.

Grading is shared out among the organization's revisers according to their language combinations. There are no written grading guidelines. Graders are simply instructed to decide whether they would be willing to revise the candidate's work, which almost satisfies Kiraly's requirement for quality to be measured in terms of revision time (2000: 162-163).

All papers are graded by two revisers, and then a brief meeting is held with the head of service. At this meeting, all the revisers discuss each candidate's performance on the two tests and decide who should pass or fail. The problem with this method is that the outcome can depend not only on subjective attitudes as to what marks a strong or weak performance, but also on the assertiveness of individual graders, i.e. their willingness or inclination to contradict the opinion of another reviser or their supervisor. How graders' attitudes vary in organization Y is shown in the summary of the results of the graders' survey in Table 8.1.

Table 8.1. Weighting of evidence in test grading\*

	Reviser						Average weighting
	A	B	C	D	E	F	
Evidence of ability to write clearly	4	5	5	5	5	5	4.8
Evidence of ability to write correctly (e.g. grammar, spelling, capitalization)	5	5	4	5	4	5	4.7
Evidence of knowledge of the source language	4	5	5	5	3	5	4.5
Evidence of the ability to produce a coherent target text (no internal inconsistencies, logical connection of ideas)	4	4	4	5	4	4	4.2
Evidence of ability to capture exact and detailed meanings (nuances)	3	4	4	5	4	3	3.8
Evidence of ability to produce target texts that flow smoothly	4	3	4	5	4	2	3.7
Evidence of ability to convey the intended effect of the source text	2	3	5	5	3	4	3.7
Evidence of the ability to achieve the appropriate tone and register	4	3	3	4	3	3	3.3
Evidence of general knowledge	3	0	3	4	2	2	1.8

\*Ratings awarded on a 1-5 scale, where 1=not taken into consideration and 5=the most heavily weighted. If there was no such evidence to be weighted, respondents were asked to put "0".

The points on which the revisers most coincided were the weighting of the ability to write clearly and correctly, to make the text internally consistent, and to achieve the right tone and register. Regarding the other six types of evidence presented for consideration, opinions differed by more than 2 points on the 1-5 point scale.

Respondents were also asked to comment on any instant disqualifiers. Four mentioned grammar mistakes, agreeing that more than one in a text would lead to disqualification. This attitude was also displayed in a meeting I had the opportunity to attend in February 2011: "I'm sorry but if there is more than one major grammar mistake, I tend not to look further", commented one reviser during the discussion, and that was the end of the consideration of that candidate. Completely misunderstanding a sentence ("Wow! He got the wrong end of the stick there") is also a direct route to failure. In the survey, "writing patent nonsense",



“embroidering the author’s message” and “serious meaning errors not attributable to unfamiliarity with vocabulary” were also reported as instant disqualifiers.

Additionally, the survey respondents made the following comments:

Reviser A: Although I did not consciously feel that “general knowledge” was a major criterion, I realized afterward that it might have played a subconscious role in my approach. For example, one text dealt with pollution, including marine pollution, an example of which was “red tide”. I was much impressed by the candidates who dealt with this subject comfortably, especially considering that the source language was Chinese.

Reviser B: I would divide this [SL knowledge] into two criteria: evidence of the ability to parse a sentence in the SL (crucial -5) and evidence of knowledge of SL lexis (important, but less so -4).

Reviser C: What I really look for is an ability to write.

Reviser D: I also consider evidence of close reading of the original - spotting subtleties or problems and dealing with them intelligently rather than translating unthinkingly.

Reviser E: I am happy to find a workmanlike translation: no serious misunderstandings, fairly idiomatic, elegance and inspiration can come later. I don’t mind if a translator doesn’t know obscure words (what are dictionaries for?), but I do mind if they misinterpret common phrases or constructions.

There therefore seems to be a general consensus on the need for grammatically correct TL and the relatively greater importance of knowledge of SL structures and phrases than SL terminology. A high level of accuracy is expected, and some revisers seem to be looking for the same “intelligence,” be it manifested in the application of general knowledge or the handling of inconsistencies. There is greater discrepancy regarding the writing skills sought, particularly as regards style and flow.

The inconsistency in the criteria applied possibly reflects not only subjective preferences but also the different types of texts used in the same examination, and from one examination to the next. The features of the L2 ST can be quite different from the features of the L3 ST, for example. This would affect the criteria and weighting applied by the various revisers, who do not all revise the same language combinations. An analysis of one set of examination papers revealed, for example, that one ST contained several lexical items that were unknown to the five candidates in question, but whose probable meaning they should have been able to guess from context. Neither of the other STs posed this kind of challenge.

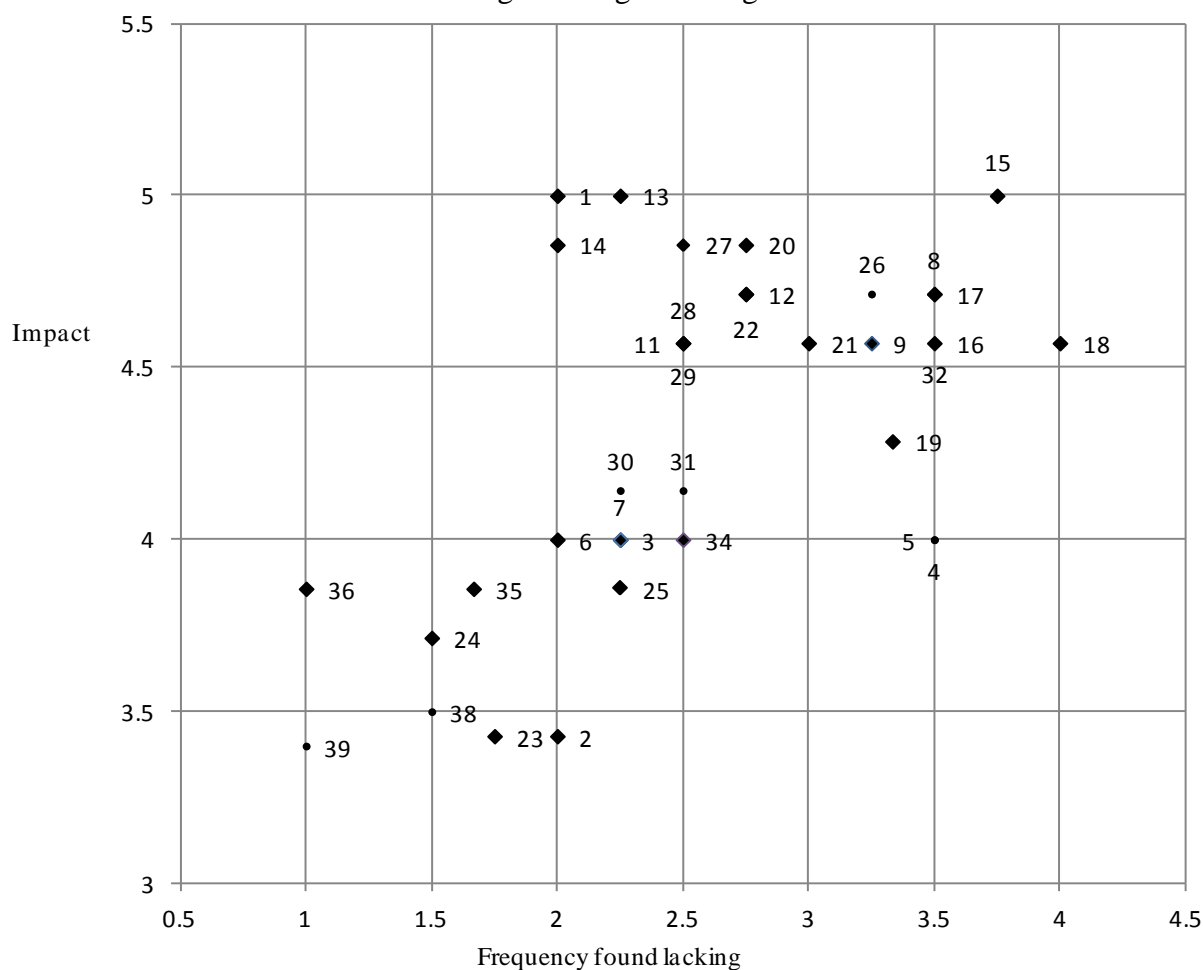
One contained more SL-specific challenges that called for considerable reformulation in the TL, and the third (apparently, because I do not know the SL in question) did not contain particularly difficult words but posed a major comprehension challenge because the message was couched in the complicated phrasing and structures typical of that language.

The average weightings obtained from this survey largely coincide with the findings of the testing survey, in which error types were rated on a scale of 1 to 5 by the head of service at this organization, as follows:

- Severity rating 5: basic distortion of the text’s meaning and grammatical errors;
- Severity rating 4: loss of nuance, shift in emphasis, slight distortions of meaning; lack of flow, naturalness, readability; internal inconsistencies; spelling errors; punctuation errors; and unwarranted additions and omissions;
- Severity rating 3: lack of lexical precision (individual words mistranslated), failure to complete the assignment in the allotted time and inappropriate tone or register;
- Severity rating 2: failure to use the correct technical term and failure to use official/correct names for organizations, countries, acronyms.

The recruitment priorities identified in Part I for this organization, or more specifically for its English-language service (as TL was shown to be an important variable) are revealed in the scatter chart presented in Figure 8.2. Again the importance awarded to the items by both revisers and translators was taken into account. In this case, however, unlike in organization X, exams are set and graded by individual language sections, and the ST depends on the TL of the exercise. In other words, the Spanish ST for the French translation examination is not the same as the Spanish ST for the English translation examination. Only the impact and frequency ratings awarded by members of the English unit were therefore taken into consideration.

Figure 8.2. Impact and frequency ratings for selected skills and knowledge types used in translating into English at organization Y



Note: See the key to Figure 8.1 for the skills and knowledge to which the numbers refer.

By eliminating the components reported as being expected to be acquired on the job,<sup>17</sup> which are identified with a dot rather than a diamond on the scatter chart, and using the same procedure and cut-off points as used for organization X above, the following groupings are obtained:

<sup>17</sup> The skills and knowledge reported as expected to be acquired post-recruitment were: 26 - adhere to in-house style conventions (25%), 38 - work with electronic terminology tools (40%), 4 - subject knowledge (20%), 31 - mine reference material for phrasing (30%), 5 - knowledge of the organization (20%), 29 - track down sources to check facts (20%), 36- follow complicated instructions (10%), and 32 - judge the reliability of information sources (20%), where the number in brackets corresponds to percentage of respondents who reported them thus. The abilities to work with translation memory software and Excel and PowerPoint and to detect mathematical errors were reported as not required and therefore do not appear on the chart.

*Category A: High-impact, high-frequency components that need to be heavily weighted in recruitment.*

Produce idiomatic translations (15), recast sentences in the TL (18), detect inconsistencies, contradictions, etc. (9), capture nuances of ST (17), understand complex topics (6), work out obscure meanings (8), produce translations that flow smoothly (16) and convey the intended effect of the ST (21).

The top recruitment priorities are TL-writing skills (15 and 18). The analytical skills (8 and 9) in this group are not tested in this organization.

*Category B: High-impact, low-frequency components that are either being adequately tested for, or are commonly found among new recruits, or are not required often.*

Knowledge of SL (1), achieve the right tone and register (22), ensure the coherence of the TT (28), convey the ST message clearly (20), ensure the completeness of the TT, knowledge of TL grammar (13), knowledge of TL punctuation rules (14), knowledge of TL spelling rules (12) and an extensive TL vocabulary (11).

The analysis of the grading scheme showed that poor grammar was an instant disqualifier, and the analysis of the test papers showed that the exams assessed knowledge of SL vocabulary. It is therefore not surprising that these are rated 5 in importance, but not commonly found lacking among new recruits (candidates without the corresponding knowledge are effectively filtered out by current testing practice).

*Category C: Low-impact, high-frequency components that are commonly lacking among new recruits but not rated as important as those in category A.*

Write elegantly regardless of the ST (19).

The inability to handle poorly written STs is not considered so important but does create revision work.

*Category D: low-impact, low-frequency components.*

Knowledge of TL and SL culture(s) (3, 24), TL and SL varieties (2, 23), maintain quality when translating under pressure (34), master new subjects quickly (7), tailor language to the readers' needs (25), explain translation decisions and problems (35) and handle more than basic Word functions (39).

The presence in this category of maintaining quality when translating under time pressure and of working with more than basic Word functions may be explained by the fact

that translators at this organization rarely work with more than basic Word functions and are not responsible for formatting their documents, and that the current recruitment test is a high-pressure task (translation of 450-500 words in 1 hour).

A comparison of the weightings derived from the results of the impact and frequency questionnaires and those reflected in the grading criteria of test graders at this organization reveals fewer differences than in organization X. In fact, the differences suggest that in order to find more translators with the skills-knowledge set sought, organization Y needs to make only the following adjustments to its examinations:

- Assess the ability to detect inconsistencies and work out obscure passages
- Weight the ability to produce idiomatic, smooth-flowing translations and recast sentences as highly as grammar and SL knowledge.

In other words, the revisers at this organization on average seem to have the right approach inasmuch as the priorities of most of them when they are grading coincide with the recruitment priorities identified in the skills and knowledge survey. The problem at this organization is that grading can be heavily skewed by the outlying opinion of a particularly assertive reviser, and, as in organization X, the test seems to measure the ability to translate texts that bear little resemblance to the texts translated on the job. The validity and reliability of testing at organization Y is also undermined by the lack of standardization of ST difficulty across languages.

#### **8.4 Conclusions of the close analysis of testing at two organizations**

The analysis of the test papers and the grading criteria used at organizations X and Y shows that the weighting of the components of the skills-knowledge set is closer to recruitment needs in organization Y than in organization X, but still not optimum, thus largely confirming the fourth hypothesis of the research: *H<sub>4</sub>: The weighting of skills and knowledge types in current recruitment examinations does not correspond to that of the profile required in the IGO in question.*

In neither organization do recruitment tests cover all the skills and knowledge required, the most notable omission in both cases being the analytical skills used to detect inconsistencies and to work out obscure passages, as noted in the survey on testing in IGOs in general. This is further evidence of the validity of the third hypothesis: *H<sub>3</sub>: Not all the*

*important components of the skills-knowledge set required for IGO translation work are tested in current recruitment examinations.*

Beyond the coverage and weighting of the components of the profile sought, there seem to be reliability and validity issues that raise questions about how effectively current recruitment tests identify ideal candidates at both X and Y. The points-grading scheme used by X and the intuitive scoring and lack of standardization of STs across languages at Y possibly undermine the reliability of their tests, and the lack of authenticity of the test task in terms of content (ST type) and conditions (hand-written without access to resources) at both organizations raises validity concerns. In neither organization does it therefore seem that the current testing practice is optimum for identifying and ranking candidates according to whether they have the desired skills-knowledge set.



## **Chapter 9 The design of a test trial to compare performance on a typical recruitment test and a profile-adapted one**

The analysis of current testing practice at IGOs presented in Chapters 7 and 8 has shown that not only do current recruitment tests not necessarily cover all the components of the skills-knowledge set they need to measure, but the weighting of the grading schemes used may not coincide with the weighting of the profile of the candidates they should identify. Moreover, the lack of authenticity of the test tasks possibly diminishes the validity of inferences made about future performance on the job, and the lack of standardization of test features and scoring criteria undermines the reliability of the examination process.

How important are those findings? The experiment described in the next three chapters was designed to begin to answer that question by comparing performance on a traditional test and on a profile-adapted test, on the premise that if both tests ranked candidates the same way, then the findings of Part I would be merely of interest to translator trainers, while if the tests ranked candidates in a significantly different way, then the findings of Part I would also have implications for those responsible for recruitment test design at IGOs.

The experiment consisted of comparing the scores and rankings of the same group of test-takers on two tests: (1) a mock traditional test, and (2) a profile-adapted test designed specifically to identify candidates with the profile sought. To neutralize the influence of certain variables that, according to the literature review, may have a bearing on performance, I standardized the test length, language combination and testing conditions and decided to eliminate the need for subject knowledge. This made it possible to focus the analysis on the influence of task type and grading scheme. Both tests therefore involved translating, within the same time limit, a text of the same length on a non-technical topic that would not require any specialized or organization-specific or other knowledge beyond what an average Master's student might be expected to know. Master's-level students were taken as the benchmark since, according to the testing survey results (see chapter 7), they would have the minimum academic qualifications required for applying for translation posts at IGOs.

Both the traditional and the profile-adapted tests were constructed with a real IGO language service in mind, since the organization and the target language had been found in Part I to significantly influence the hierarchy of skills and knowledge required. The organization in question was selected because, according to the testing survey results, its testing practice was fairly typical in many ways. The organization agreed to participate on condition of anonymity and is therefore referred to as organization Z.



The length of the STs in the two tests was set at approximately 500 words. Ideally, participants would have been asked to do more or longer translations in order to obtain a larger ratable sample, but 500 words was selected for two reasons: this was close to the average length of STs in traditional tests at this and most IGOs, and it was as much as could probably be expected from the test-takers and graders participating in the experiment, especially as they were unpaid volunteers. The original completion time of 1 hour and 15 minutes was selected in part for the same reason and in part because the ability to maintain quality under pressure of time was a top priority for organization Z. The language combination was set as Spanish-English translation for the sake of convenience, as it is my area of specialization. Also for convenience, it was decided that both tests would be done on a computer, even though in some IGOs, including organization Z, translation tests that form part of competitive examination processes are still done by hand. Apart from the text length and the time allowed, the two tests were very different. The design of each is described in detail in the next two sections.

## 9.1 The design of the mock traditional test (Test One)

Since organization Z does not work with test specifications or standardized grading guidelines, the design of the mock traditional test was based on the results of the testing survey, the direct observation of testing practice at Z and consultations held with some of its examiners. The mock traditional test was referred to as Test One throughout the trial so as not to influence the attitude of participants (test-takers or graders) towards the test, and it is referred to this way here as well.

### 9.1.1 The test task

The main test task at organization Z is the translation of a text or texts from L2 to L1, and the source text for Test One was selected and prepared using the same procedures used at organization Z. Newspapers, magazines, journals in printed and online format were scoured to find a suitable Spanish text that had not already been translated into English and was of the right (typical) length (approximately 500 words). The text *Aprendizaje frustrado*, an editorial taken from a Uruguayan newspaper,<sup>18</sup> was chosen because it did not require much, if any,

---

<sup>18</sup> Posadas, Juan Martín. 2011. "Aprendizaje frustrado". In *El País*, edition of 13 March 2011. Available online at: <http://www.elpais.com.uy/110313/predit-553004/juanmartinposadas/aprendizaje-frustrado/>

specialized knowledge (of a technical nature or of a particular IGO), was 525 words long and quite well written. The register of the text was neither particularly high nor low and it contained several relatively obscure words, which would challenge non-native speakers but be possible to guess from context, as well as figurative language and complicated structures. It also had a clear argument to follow. In that regard, it was similar to the Spanish STs analyzed in the closer inspection of testing at the larger of the two IGOs presented in Chapter 8 (organization X), as well as some past papers at Z that I was shown. A comparison of some of the features of those STs and the one used in Test One is presented in Table 9.1. Most of the features are those recommended by Nord for ST analysis (Nord 2005: 41-141). I added the total word count and number of low-frequency words that I thought would not be easy to guess from English.

Moreover, the features coincided with those reported as typical for testing at Z in the testing survey, namely: vocabulary is difficult but does not include culture-specific or organization-specific concepts; the ST includes stylistic features unique to the SL but is not written in a particular rhetorical style; there are no internal inconsistencies or coherence problems, although the meaning may be somewhat obscure in parts and there are many subtleties that might be missed; and some sections require considerable reformulation in the TL. Two heads of service responsible for test-text selection were shown the text and both agreed that it was fairly typical of the test-texts used at Z. They felt that some of the expressions were too obscure, however, and would be difficult to guess from context. These were therefore removed or reworked, and a few other minor editing changes were made for the sake of clarity or length. Otherwise, no major changes were made to the original. The text used in the trial, with the changes marked, can be found in Appendix 12.

Table 9.1. Comparison of source texts used in translation tests at organizations X and Z and in Test One<sup>19</sup>

Factor	Sample test-text of organization X	Sample test-text of organization Z (i)	Sample test-text of organization Z (ii)	Test One test-text
Sender	Unidentified columnist or essayist	Unidentified essayist	Unidentified, but is famous Mexican poet and essayist	Unidentified columnist or essayist
Sender's intention, motive/function of the communication	Communicate opinion and make the reader reflect. Entertain with metaphor.	Communicate opinion and make the reader accept argument	Transmit description	Communicate opinion and make the reader reflect. Entertain with descriptions used to make points.
Audience	Readers of news articles on current affairs	Readers of essays on current affairs or culture	Readers of essays on current affairs	Readers of news articles on current affairs
Medium	Written text taken from unidentified journal or newspaper.	Written text taken from unidentified work, possibly an abstract	Written text taken from unidentified work, actually part of a collection	Written text taken from unidentified newspaper.
Place	The region in which org X operates "identified by the use of <i>nosotros</i> ("we")	Probably the West	Mexico presumably	Uruguay presumably
Time	Not stated, would be presumed to be during the 12 months before the examination	A few decades before the examination; date given indirectly (1990s)	Not stated, not apparent, actually published in 1950	Not stated, would be presumed to be during the 12 months before the examination
Subject matter	The status and workings of the institution	The importance of art in the face of technical progress.	The Mexican psyche	The economic boom in Uruguay
Content	Identification and criticism of factors hampering the effectiveness of the institution; no gaps of cohesion or coherence	Brief reference to threats of annihilation posed by technical progress followed by description of civilizing effect of art; no gaps of cohesion or coherence, but seems to lack development and conclusion	Description of "El Mexicano" to portray the attitudes of Mexicans towards others – possibly of Mexico to the outside world; no gaps of cohesion or coherence, but seems to lack conclusion	Description of reactions to economic boom and impact on consumers, warning that undeserved economic boom teaches nothing; no gaps of cohesion or coherence

<sup>19</sup>Features are taken mainly from the factors of ST analysis recommended by Nord (2005: 41-141).

<b>Factor</b>	<b>Sample test-text of organization X</b>	<b>Sample test-text of organization Z (i)</b>	<b>Sample test-text of organization Z (ii)</b>	<b>Test One test-text</b>
Presuppositions	Basic knowledge of the institution, current international affairs and cooking	Knowledge of current affairs	Knowledge of Mexican history and Greek mythology (reference to Narcissus looking at his reflection)	Knowledge of Uruguayan current affairs (President mentioned by name)
Composition	Six well structured paragraphs; no title, headings or other markers; thematic progression shaped by cooking metaphor, from deciding the menu to tasting the product	Six paragraphs, main point stated in one-sentence fifth paragraph; others are of equal length	Four well structured paragraphs that fit well together, but the fourth paragraph is not meant to be the last; clearly taken from a longer piece	Seven paragraphs, each introducing new angle; thematic progression shaped by graphic down-to-earth imagery, parallel structures and repetition
Non-verbal elements	None	None	None	None
Lexis	Extended metaphor used throughout occasionally creates unusual word combinations; several figures of speech; generally high register, rather scholarly language; appellative use of first person plural	Largely abstract vocabulary; scholarly language; appellative use of first person plural	Very poetic; many marked collocations and unusual images; use of repetition, parallel structures; appellative use of first person plural	Juxtaposition of concrete and abstract terms; figurative language used, including popular sayings and metaphors; appellative use of first person plural
Sentence structure	Sentences short for Spanish texts; no unusual sentence structures; use of punctuation and parallelism to maximize clarity	Long sentences throughout	Careful phrasing; deliberate juxtaposition of short and long sentences	Use of parentheses; generally long sentences; use of short balanced sentence to highlight main point
Number of low-frequency words <sup>20</sup>	7	11	12	12
Number of words and sentences	504 words in 21 sentences (average 24 words per sentence)	544 words in 23 sentences (average 24 words per sentence)	490 words in 29 sentences (average 17 words per sentence)	504 in 24 sentences (average 21 words per sentence)

<sup>20</sup> This refers to words M.A. students might be expected to want to look up in a dictionary, based on their frequencies and their similarity to English equivalents. Frequencies were calculated using a corpus (Davies, Mark. (2002-) *Corpus del Español: 100 million words, 1200s-1900s*. Available online at <http://www.corpusdelespanol.org>), in which words occurring less than 2,500 times in a corpus of 100 million words were considered low frequency.

<b>Factor</b>	<b>Sample test-text of organization X</b>	<b>Sample test-text of organization Z (i)</b>	<b>Sample test-text of organization Z (ii)</b>	<b>Test One test-text</b>
Special uses of punctuation	None	Use of inverted commas to draw attention to certain words and expressions	Careful phrasing, use of inversion of subject and verb, and of clauses, use of inverted commas to set off phrases,	Stylistic punctuation to add expressivity and emphasis words such as dashes and ellipsis
Effect	Reader admires author for successful use of metaphor to present his/her opinion and the amusing imagery it creates; reader is probably satisfied with small portion of “food for thought” received	The reader is left unsure of whether the article is complete as the message has not been effectively developed	Reader admires author’s style and powerful verbal portrait of the mistrustful yet dignified Mexican; reader wants to read more	Reader receives the message and is aware of an issue probably not considered beforehand

### 9.1.2 Test conditions

Test-takers were asked to produce their translation at one sitting using a computer and Microsoft Word. In keeping with testing practice at Z, they were asked not to consult any resources, human or otherwise.

### 9.1.3 Instructions for test-takers and mode of delivery

The test paper and the instructions for test-takers were prepared for delivery in electronic format. The instructions were modeled almost verbatim on those used by Z, apart from the reference to using a computer for word-processing and the submission of completed translations by e-mail. The instructions were very brief and contained no context information. They are presented below.

#### **Test One**

**The time allowed: 1 hour and 20 minutes<sup>21</sup>**

#### **Instructions**

Without sacrificing accuracy, candidates should strive to achieve a good style in translating the Spanish text below into English. No alternative translations should be submitted (i.e. please do not give two different translations for the same piece of Spanish text).

Candidates are asked to type up their translations in Word and to save them as “TEST ONE XXX”, substituting XXX with their own name.

Candidates need not concern themselves with formatting beyond paragraphing and the use of italics for emphasis as considered appropriate.

During the test, candidates must not use the computer for anything other than word processing and may not consult resources of any kind, online, human or otherwise.

Please note that only complete translations will be marked. Candidates are therefore strongly advised to manage their time carefully.<sup>22</sup>

---

<sup>21</sup> The time was lengthened from 75 to 80 minutes after the pilot.

<sup>22</sup> This warning was included after the pilot in the desire to obtain ratable scripts.

### 9.1.4 Grading scheme

Two traditional grading methods, similar to those Z has applied in the past, were prepared. The first was an intuitive scoring method, in which graders would assign the translation a score out of 20, with 12 being the pass mark. This method was intended to replicate the kind of holistic assessments made by graders working with schemes like that used most often by organization Z (and Y). The number 20 was chosen in order to provide a broad enough scale to discriminate between test-takers. The second method was a points-deduction one based very closely on that used by organization X, but without the grader being told how many points “in the bank” candidates would start out with or what the pass mark was, so as not to influence how many negative points the grader would be willing to deduct. Although largely dismissed by language testing and translation scholars as incapable of producing meaningful scores, the method was thought potentially useful for ranking test-takers and thus able to provide input for the research. This two-method scheme is referred to as scheme A.

### 9.1.5 Instructions for graders

The instructions for graders used in the trial are shown below. Given the shortage of graders available for the trials,<sup>23</sup> it was decided to have the same graders apply both methods in order to reduce the number needed. Graders at organization X do in fact work with a dual scheme similar to this one (see Chapter 8). The instructions for the intuitive method reflect those usually given orally to graders. The instructions for the points-based one are modeled closely on those used by organization X.

GRADING SCHEME A- SCORECARD AND FEEDBACK	
Grader: _____	Test (please tick):      One <input type="checkbox"/> Two <input type="checkbox"/>
Candidate No.: _____	
Score out of 20: _____	
Points balance (+/-): _____	
Finished: Yes <input type="checkbox"/>	No <input type="checkbox"/>
Strengths/weaknesses: _____	

<sup>23</sup> It was decided to have the scripts graded by professional revisers from organization Z, as recommended by Hatim and Mason (1997: 203).

### GRADING SCHEME A - INSTRUCTIONS

Thank you so much for volunteering to help trial different grading schemes on different translation tasks. The results will provide valuable input for recruitment testing, so your participation is much appreciated. You are of course assured total anonymity.

#### Instructions

- Please remember to grade the attached set of scripts in the order they are presented to you.
- Please try to grade several scripts at one sitting as you are more likely to be consistent (and will probably get through them more quickly).
- Please make a note of how much time you spend grading the scripts, either altogether or individually.

#### Procedure

- Read through the translation once and assign the candidate a score of between 1 and 20, where 12 and above is a pass, and the pass/fail criterion is whether, as a reviser, you would be willing to revise the work of the candidate if they joined the translation service.
- Write your score in the space provided on the score card. If you give two or more candidates the same score, please indicate, if applicable, which candidate was better than the other(s) using letters after the score (15a, 15 b, 15c, etc.)
- Go through the translation again marking the text and writing in the margin the points to be deducted or added using the scheme provided on the attached card.
- Total the points and write the balance on the scorecard in the space provided. **DO NOT CHANGE YOUR FIRST (INTUITIVE) SCORE.**
- Please indicate any unfinished scripts (ones in which the candidates seems to have run out of time at the end rather than missed bits in other parts of the text).
- Please comment on the candidates' performance. If you need more space, use the back of the scorecard. Your identity will of course not be revealed to the candidate (or anyone else for that matter).
- Please return the scripts, the test papers, the scorecards and your feedback to the chief grader when you have finished.

### GRADING SCHEME A - POINTS

<b>MEANING ERRORS:</b>		<b>SPELLING AND PUNCTUATION:</b>	
Minor	—1	Minor	—0.5
Serious	—2	Serious	—1
Very serious	—3		
<b>GRAMMAR/SYNTAX ERRORS:</b>		<b>OMISSIONS:</b>	
Minor	—0.5	Minor	—0.5
Serious	—1	Serious	—1
Very serious	—2		
<b>STYLE/CLARITY /REGISTER ERRORS:</b>		<b>INAPPROPRIATE ADDITIONS:</b>	
Minor	—0.5	Minor	—0.5
Serious	—1	Serious	—1
Very serious	—1.5		
<b>BONUS POINTS FOR INGENIOUS SOLUTIONS + 1 or +2</b>			



## 9.2 The design of the profile-adapted test

The preparation of the profile-adapted test was a far more complicated undertaking. As its name suggests, the test was heavily based on the profile sought. The design process started with the identification of the skills and knowledge the test was intended to measure using the methodology described in Part I.

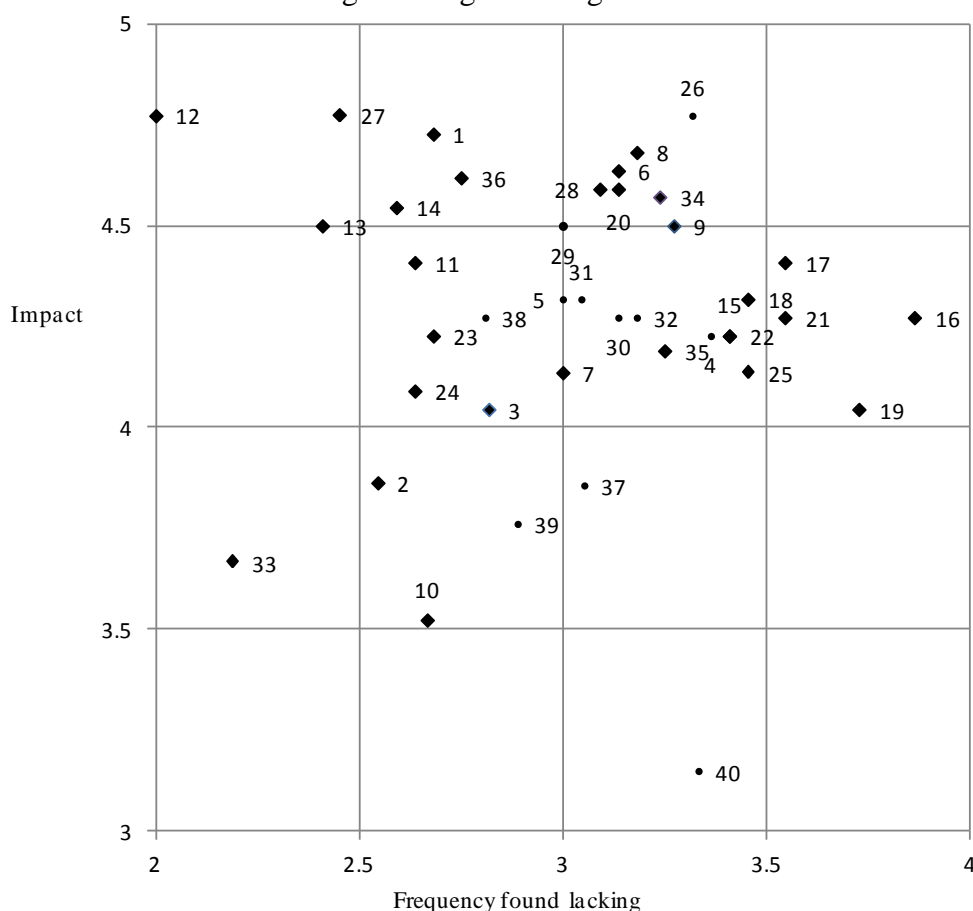
### 9.2.1 *Identification of the construct(s) to be measured*

According to the literature on language testing reviewed for this research, test design usually begins (or should begin) with a definition of the domain to which we wish to generalize, mainly to induce the constructs to be measured (Bachman and Palmer 1996: 117). We could use the survey instrument we had designed to obtain empirical data from experienced professionals in the domain to not only identify those constructs but to prioritize them. The testing priorities for organization Z were determined on the basis of the scatter chart presented in Figure 9.1. The data correspond to translation into English only, and the impact ratings are simple averages of the ratings awarded by translators and revisers alike. The skills and knowledge types that new recruits are not expected to have prior to joining the organization or are best acquired in-house (the research skills, subject knowledge, knowledge of the organization and the ability to adhere to in-house style) are identified by small dots instead of diamonds.<sup>24</sup>

---

<sup>24</sup> The following skills and knowledge types were reported in an interview with four revisers working for organization Z as not being required prior to recruitment: adhere to in-house style conventions, subject knowledge, knowledge of the organization, work with translation memory software, handle more than basic Word functions, work with electronic terminology skills, Excel or PowerPoint, track down sources to check facts, mine reference material for phrasing and judge the reliability of information sources.

Figure 9.1. Impact and frequency ratings for selected skills and knowledge types used in translating into English at organization Z<sup>25</sup>



Note: See the key to Figure 8.1 for the skill or knowledge type to which each number refers or refer to the categorization below.

The groupings in the scatter chart suggest that the components of the skills-knowledge set required prior to recruitment at organization Z should be prioritized as follows:

*Category A skills and knowledge types (high-impact, high-frequency components):* work out the meaning of obscure passages (8), detect inconsistencies, contradictions etc. (9), understand complex topics (6), convey the ST message clearly (20), maintain quality under pressure of time (34) and ensure the coherence of the TT (28).

These components of the skills-knowledge set should not only be assessed in the test, but also heavily weighted.

<sup>25</sup> It should also be remembered that other skills and knowledge, not included in the list, would need to be factored into decision-making (knowledge of current affairs/general knowledge and interpersonal skills). They were excluded from the analysis since they would be extremely difficult to assess in a short text-based translation test.

*Category B skills and knowledge types (high-impact, low-frequency components):* SL knowledge (1), follow complicated instructions (36), ensure the completeness (27), the correctness (grammar (13), punctuation (14), spelling (12)) of the TT and an extensive TL vocabulary (11)

These components are as important as the category A ones, but do not necessarily generate much revision work. They should be tested to uphold the communicative impact of the IGO's translations, but not weighted as heavily as the category A components.

*Category C skills and knowledge types (high-frequency, low-impact components):* master new subjects quickly (7), the ability to capture nuances (17) and strong TL writing skills (recast sentences (18), tailor to the readers' needs (25), convey intended effect (21), idiomaticity (15), flow (16), tone and register (22) even when working from poorly written STs (19))

The relatively frequent absence of these components accounts for much of the revision load at the organization. They need to be tested to improve the productivity of the translation service, but not weighted as heavily as the category A components.

*Category D skills and knowledge types (low-impact, low-frequency components):* knowledge of SL and TL varieties (2, 23) and cultures (3, 24) and typing (33)

The greater importance of knowledge of regional varieties of English over knowledge of varieties of the SL shown in the scatter chart is intriguing, but not particularly relevant as none of the components in this category are that scarce and necessarily worth testing.

The implications of the scatter chart analysis for the test are as follows:

- In general, measurement efforts should focus on the category A skills and knowledge types.
- In terms of coverage, all the category A, B and C components should be measured in the test, if possible. Certainly, all category A components must be assessed.
- The D components do not need to be assessed in the test.

The category A components should be weighted more heavily than the others. Any differences in the weighting of category B and C components should depend on the

organization's priorities: is it looking for translators who are more likely to ensure it achieves its communicative aims and protects its reputation for quality, or for translators who can help it boost its efficiency? If both, then the category B and C components should be equally weighted behind the A components. This was the approach chosen at this stage for the trial. It was also decided at this point to exclude the abilities to master new subjects and to understand complex topics, as they would be very difficult to measure in a short translation test and could possibly be measured by another means. Accordingly, the profile-adapted test for organization Z should (using the definitions used in the skills and knowledge survey), in terms of coverage, aim to measure the following:

The ability to follow complicated instructions and detect, when translating into English Spanish texts typical of those translated by the organization, inconsistencies and slips in logic in the ST, such as missing markers, tautologies, nonsense, redundancies, empty references, impossible sequences, confusion of cause and effect and misconceived headings;

and

The ability to draw on linguistic and other skills and knowledge to produce (by recasting sentences as necessary) correctly written, idiomatic, clear and coherent English translations (of Spanish texts typical of those translated by the organization and which contain some obscure passages) that flow smoothly, capture the exact and detailed meanings of the ST, achieve the appropriate tone and register and convey the intended effect of the ST, even when the ST is poorly written and does not flow smoothly and even when required to do so under time pressure.

In terms of weighting, the profile-adapted test for organization Z should aim to weight the ability to detect inconsistencies, work out obscure texts and produce a coherent and clearly written TT quickly more heavily than the other components, which should be weighted equally.

### *9.2.2 Preliminary considerations for the design of the test task*

According to the findings of the literature review, the authenticity, interactiveness and level of difficulty of test tasks are key considerations in language testing. Their application in the design of a profile-adapted translation test is discussed below.

### 9.2.2.1 Authenticity

“Authenticity” for our purposes refers to being domain-relevant. In other words, an authentic task is a task that is highly similar to one that is performed in the domain to which we wish to generalize, and in the case of translation tasks, an authentic source text is one that shares key features of source texts translated in the domain. The source text we use should therefore be typical, but not necessarily real. A domain analysis was carried out, therefore, not to identify what skills and knowledge we wished to measure, but to maximize the authenticity of the test task. This meant identifying text types, turnaround times on rush jobs, the key features of the STs and the challenges they typically pose, the tools translators use, and the conditions under which they work. The analysis was performed by directly observing the work carried out by translators in organization Z. The results are summarized below.

Translators at Z mainly translate information documents covering a wide range of economic and social issues. The documents include political and legal texts. Some are highly technical, but most are report-like in style and, in terms of subject knowledge, are no more difficult to understand than newspaper articles on technical issues. Apart from the linguistic challenges typical of source texts at any institution, they pose two other challenges: they require careful consultation of reference documents because they form part of a chain, and they can be difficult to understand because they are sometimes poorly written. Any content errors that cannot be fixed without blatantly mistranslating the source text are flagged for revisers to follow-up on with the author (or the ST editor if there is one). Although the texts are sometimes highly technical, terminology is not seen as a major challenge, as the translators have access to a wide array of electronic and bibliographical sources of information to help them understand texts and find the right technical words for the translation. Translators at Z may not know who the authors are, but they are very aware of what the translation will be used for, and they adapt accordingly. They produce their translations using Word (some use speech-recognition software applications). They are not responsible for more than basic formatting. For example, they do not have to format tables or graphs and can give instructions in their translation to the text-processing unit about the text that should be inserted in a figure. Most work is done without too much time pressure, although during certain periods, extremely short turnaround times can become the norm (e.g. 400 words have to be translated in 45 minutes).

### 9.2.2.2 *Activation power*

Bachman refers to the need to maximize the “interactiveness” of language test tasks, which he defines as “the extent and type of involvement of the test-taker’s individual characteristics in accomplishing a test task” (Bachman 1990: 25). Tasks must activate as many components of the construct as possible, and as often as possible. In language testing, this is achieved by determining “what combination of test method facets is most likely to promote an appropriate interaction of a particular group of test-takers with the testing context” (Bachman 1990: 317). In translation testing, our needs are slightly different: we are not worried that test-takers might be unable or unwilling to perform or produce a ratable sample because they cannot relate to the form or content of the input material. The corresponding concern in translation testing is ensuring what I will call the tasks’ *activation power*, which is their power to force test-takers to apply the skills and knowledge we wish to assess. This power is not just a matter of quality, but also of quantity to generate sufficient examples of probable manifestations of the components of the construct for us to make meaningful inferences about the construct in relation to each test-taker. The ST in the profile-adapted test therefore would have to be effectively sown with several challenges that would activate the skills and knowledge types we sought to assess. This is reminiscent of Nord’s practice of selecting texts that pose particular problems (2005: 162-177).

### 9.2.2.3 *The level of difficulty*

The determination of the desired level of difficulty was shaped by two considerations. The first was the domain tasks that successful candidates would be expected to perform, as identified by the domain analysis. Test tasks should not be more difficult than domain tasks but should stretch test-takers and enable the strongest to perform to the best of their ability. The discriminatory power of the test and hence the validity of inferences made on the basis of results would be more problematic if the test was too easy. The other consideration was what could reasonably be expected of candidates. This had been addressed in the consideration of post-recruitment skills acquisition during the identification of the skills-knowledge hierarchy and the definition of the construct.

In short, what was needed was a task that would activate all the category A, B and C components of the construct in an authentic way and would make it possible to discriminate among candidates. Evidence of the A components, in particular, had to be as abundant as possible.

### 9.2.3 *Activating the A components*

The A components were: the ability to maintain quality even under time pressure, four largely analytical skills (work out the meaning of obscure passages, understand complex subjects, detect inconsistencies in the ST and ensure the coherence of the TT), and the ability to convey the ST message clearly. Although TL writing skills have to be employed to achieve coherence and it is measured in its manifestation in the TT, coherence “is not an inherent quality of the text itself, but rather comes from the accuracy of the writer’s assessment of what the reader will be able to infer from the text” (Cushing Weigle 2002: 21). It is therefore classified as an analytical skill here.

Obtaining evidence of the ability to meet standards when translating quickly would be addressed by establishing a relatively short turnaround time for the task (500 words in 1 hour and 15 minutes), as is done in traditional testing. At organization Z, productivity targets for translations requiring considerable reference checking are approximately 300 words per hour. Asking candidates to translate 500 in 1 hour and 15 minutes would thus be appropriate if the goal was to see how they performed under time pressure. There would of course be a danger that test-takers would not finish and therefore not produce a ratable or comparable sample.

Apart from the ability to produce a coherent TT, analytical skills are rarely measured in IGO translation tests, and that applies in organization Z as well. According to the testing survey results, a few IGOs ask translators to edit texts written in their TL, but that possibly requires a different skills-knowledge set and has them perform a task they would not perform if hired, i.e. an unauthentic task that raises validity concerns. Translating obscure passages is one of the challenges of IGO translation work. It often requires not only linguistic skills but contextual knowledge, which can sometimes be found in the text itself but is frequently subject- or organization-specific. Accessing that kind of knowledge would call for displays of resourcefulness, such as consulting the author and colleagues, or ploughing through background documents, and that cannot be practically arranged for in examinations. However, some aspects of working out obscure passages can be more readily assessed in a translation test: for example, the ability to detect nonsense or contradictions that contribute to obscurity, and the ability to express ideas clearly in the TT regardless of how unclearly they have been expressed in the ST. It

was therefore decided to limit the assessment of the broader ability to work out obscure passages to the measurement of those two particular skills.

According to the literature on IGO translation and as confirmed by comments made in the impact and frequency questionnaires, as well as the direct observation of the tasks regularly performed by translators at several IGOs, including Z, not only do IGO translators often have to translate poorly written STs and work out obscure meanings; they are also expected to point out any inconsistencies in the ST for follow-up. It was therefore decided to ask test-takers in the profile-adapted test not only to translate a ST into their mother tongue, but also to flag any inconsistencies in the ST that would need addressing. This two-fold task would not only be authentic, it would activate another key aspect of the construct: the ability to produce well-written TTs from poorly written STs. Specifically, it would help assess another category A component, the flipside of the analytical coin: the ability to convey the ST message clearly.

#### 9.2.4 *Activating all the components*

The B and C components can be grouped into (1) skills and knowledge used for comprehension (of the SL, the instructions and the readers' needs), (2) skills and knowledge used for transferring the message accurately (the ability to capture nuances and convey the intended effect of the TT), and (3) target-language writing skills and knowledge (TL vocabulary and the ability to recast sentences and produce flowing, idiomatic, correctly written translations). The ability to respond to readers' needs and to follow complicated instructions would be activated by giving test-takers context information and instructions (see section 9.2.8), but was actually not explicitly measured, a point noted and lamented by one of the graders (see section 10.4). The others would be activated by the source text used in the translation task.

To try to make sure that the ST would have the necessary activation power, a list was made of ST features that were critical to the kinds of inferences we wished to make (Bachman and Palmer 1996: 46). In addition to being the activators of the components of the construct, these features would be the difficulty drivers of the task, one of the ways recommended by Cushing Weigle (2002: 50) to enhance test validity. The results of the domain analysis were also taken into account to maximize the authenticity of the ST. The linguistic features are listed in Table 9.2 using Bachman's classification of the nature of language input (Bachman 1990: 130-136). They were intended not only to guide the



preparation of the ST for the profile-adapted test, but also to serve as a reference for any future repetitions of the experiment.

Table 9.2. Linguistic features of the source text sought for the profile-adapted test

---

PROPOSITIONAL CONTENT (characteristics of the information):

- a. *Vocabulary*. The vocabulary should be on the whole straightforward. It should not pose major comprehension challenges. The ability to find terminology fast is not included in the construct and could be activated if the ST contains difficult L2 vocabulary. However, the language used has to be typical of that used in domain STs, for the sake of authenticity. In parts the wording and phrasing should defy literal translation (e.g. include false friends, L2-unique expressions and cultural references) to activate the ability to produce an idiomatic TT. There should be inelegant repetitions or excessive wordiness that require sentences to be recast in the TL. In parts, the vocabulary should express subtleties to activate the ability to capture precise levels of meaning and emphases in the TT.
- b. *Degree of contextualization*. The ST should be fairly context-reduced inasmuch as it should contain information that is largely expected to be new to the test-takers, in order to avoid bias.
- c. *Type of information*. There should be a mixture of both abstract and concrete, positive and negative, factual and counterfactual information, given that all types are found in the domain STs.
- d. *Topic*. The topic should not call for specialized subject knowledge, including knowledge of the organization or current affairs, beyond what any university graduate in the English-speaking world may be expected to know.
- e. *Genre*. The ST should not belong to a genre that may be unfamiliar to test-takers. The most common genre at Z is a descriptive factual account, which is usually published with a view to making a point (and thereby indirectly calling for action) or to informing decision-making.

---

ORGANIZATIONAL CHARACTERISTICS (the use of conventions for joining utterances or sentences):

- a. *Grammar, spelling, punctuation*. Grammar, punctuation and spelling should largely be correct and only faulty to interrupt clarity, flow or internal logic to test the ability to produce clear, flowing, logical prose in the TT.
  - b. *Cohesion*.<sup>26</sup> The cohesion of the text should be poor, to activate the analytical skills. Some parts need to be poorly written in terms of articulation of ideas and correctness, e.g. miscollocations, unnecessary wordiness, faulty parallelism, incoherent chains of cause and effect. The ST should contain enough inconsistencies to ensure the discriminative power of the task, e.g. empty references, illogical sequences, tautologies, obvious factual errors (slips), etc. Target: at least one per 100 words of text.
  - c. *Rhetorical organization*. The text should be organized on the basis of the conventions for factual description and then, possibly, an appeal, related to achieving an effect on the reader.
- 

<sup>26</sup> “Cohesion” refers to ways of explicitly marking semantic relationships, such as ellipsis, conjunction, reference and lexical cohesion (Halliday and Hasan 1976).

---

## PRAGMATIC CHARACTERISTICS

1. *Illocutionary force.* The language function is to inform and possibly to elicit understanding and support for a position (intended effect). Ideally, the ST would have a clearly intended effect (e.g. a message beyond the mere presentation of facts), although this might be difficult to find or develop in a short text.
  2. *Sociolinguistic features.* The ST should be written in standard Spanish, not a noticeable dialect or variety (since such knowledge is not required). The language should be natural in terms of being rather dry, colorless prose that is typical of domain STs, but possibly rather unnatural in terms of sloppiness and lack of clarity. The ST should have the tone and register typical of IGO documents.
- 

### 9.2.5 *Preparation of the ST*

It would have been impossible to find a genuine text that was short enough to have all the above linguistic features. It was unexpectedly easy, however, to find authentic material to use as the starting point for designing the ST that would meet the necessary criteria in terms of activation and discriminatory power as well. The texts in question were found on the website of the United Nations Economic Commission for Latin America and the Caribbean, which produces a wide range of publications in Spanish. Natural disasters were chosen as a topic that test-takers were likely to be familiar with but not experts in. The base texts are presented in Appendix 13.

Changes included removing technical vocabulary, inserting inconsistencies, inserting known Spanish-English translation challenges and ensuring an adequate range and density of construct activators to create a ST that was more likely to produce ratable samples from which more valid inferences about abilities might be drawn. The idea was to ensure the ST contained the kind and number of challenges needed in order to obtain as reliable a measure as possible of each component of the construct and to discriminate between test-takers. The changes were also influenced by the results of the domain and target-population analyses. The ensuing ST is therefore an imitation of domain STs in which the typical challenges that domain STs pose have been necessarily concentrated into just a few paragraphs. It is interesting to note that several of the inconsistencies and features of poor writing existed in the original base texts, reinforcing the validity of including a poorly written ST in the task in the interests of authenticity. An annotated version of the ST showing the changes made to the base texts is presented in Appendix 14.

The inconsistencies and logic problems that test-takers were expected to spot in Test Two were as follows:

1. An incorrect superordinate is used to introduce a list.
2. A reference to a table is faulty (the text says it is “above” when it is “below”)
3. The wrong date is given in one instance, as can be deduced from the text.
4. Certain numbers in the table are expressed incorrectly for the English version (decimal commas instead of points)
5. A mathematical error: the proportions are said to add up to 1000% instead of 100%.
6. A logic error/faulty reference: the text states that the disasters will affect the “aforementioned least-vulnerable” countries the most, when it has referred to the “most” vulnerable countries.
7. Another logic error: the text refers to “the United States and the industrialized countries”, suggesting that the United States is not an industrialized country.

This approach to ST preparation differs considerably from current testing practice at Z and other IGOs, in which heads of service select a “real” text, usually taken from an external source, that they think will make a good test. Such practice is more in keeping with what Bachmann calls the “real-life” approach and the “new behaviorism” referred to by Chapelle and the underlying pragmatic theory of validity, as mentioned in the literature review, in which “performance is viewed as a sign of underlying traits and is influenced by the context in which it occurs and is therefore a sample of performance in similar contexts” (Fulcher and Davidson 2007: 17). The problem with the “real-life” approach is that generalizations about competence are limited to the ability to perform a specific type of translation task in a specific context because the test has not been designed to distinguish or provide much information on separate abilities. If the test STs are not authentic in type (as is the case on our mock traditional test and often is the case in IGO testing), and no effort has been made to ensure the task activates the skills and knowledge the organization needs to find, then there is a danger, from the recruitment perspective, of hiring the translation equivalent of a golf player when you need a footballer. Moreover, in translation testing, “real-life” STs, even if they are authentic in type, are highly unlikely to contain the necessary range and density of activators of the constructs we should be measuring. i.e. have sufficient activation power. Tests by definition will produce samples of performance, but we need those samples to be as evidence-rich as possible.

If the problem with “real-life” translation tasks is their relative lack of activation power, the problem with tasks with excellent activation power is their potential lack of authenticity, which could undermine the very construct validity upon the basis of which

inferences are justified. The ST prepared for Test Two was shown to ten Spanish-English translators and revisers from organization Z. They all agreed, when shown the text, that all the challenges it posed were ones they had come across in the documents they had translated over the course of their careers. They were surprised to find so many problems in such a short text, but did not immediately conclude that it must have been created artificially for the test trial. “Where did you find it?” was the question asked, not “How long did it take you to put it together?” Of course, they may have just presumed that I would use a genuine text, but the fact it was not patently obvious that the text was artificially created for examination purposes suggested that a reasonable balance between authenticity and activation power had been struck. I had created a ST that was a condensed imitation of a domain ST and should activate the various components of the skills-knowledge set sufficiently to be able to obtain enough evidence to make some inferences about abilities, while still meeting minimum authenticity and practicality requirements. With regard to its level of difficulty, all those consulted also agreed that the ST would be relatively easy for an experienced professional to translate but probably challenging for a novice. That hypothesis would be tested in the pilot and the trial itself.

#### 9.2.6 *Test conditions*

For the sake of authenticity, test-takers would produce their translations using a computer and would be free to consult any resources they wished, except human ones. Consulting human sources was prohibited to ensure the translation would be the test-taker’s own work.

#### 9.2.7 *Instructions for test-takers and mode of delivery*

Subjects would receive the STs in electronic format together with instructions and a fictitious context designed to elicit the type of performance sought. The instructions were detailed and contained a clear description of the criteria for success, as recommended by Alderson et al. (1995: 21) in language testing and by advocates of the *Skopos* approach in Translation Studies, to ensure that test-takers would have the opportunity to perform to the best of their ability. The instructions for Test Two are presented below.

**TEST TWO****Time allowed: 1 hour and 20 minutes****Instructions**

- Please read through the instructions and context information carefully and then translate the text that follows.
- Your translation will be rated according to the style and correctness of your writing and the clarity and accuracy with which you manage to convey all the levels of meaning of the original. Please do not give two different translations for the same piece of Spanish text.
- Your ability to detect inconsistencies in the source text, as explained below, will also be measured.
- You are free to use the Internet to consult websites, dictionaries and glossaries, etc. for researching terminology or background information as you think necessary. You are not allowed to use the Internet to run the source text through a machine translation application or to e-mail, chat or communicate with other people in any way during the test.
- Please type your translation. You need not concern yourself with formatting beyond paragraphing and the use of italics for emphasis, as you think appropriate.
- Please remember that incomplete translations will not be marked.
- Please remember to complete the questionnaire at <http://www.surveymonkey.com/s/TESTTYPE2> when you have handed in your translation and please be a bit patient: you might need to wait a few seconds at the end for the system to save your answers or to move from one page to the next, especially if lots of other people are answering at the same time.

**Context and description of the task**

The text below is a summary of a report published in December 2010 that is to be used as input for a meeting on disaster prevention and preparedness in Latin America and the Caribbean. The Spanish text was put together in a hurry at the end of the year and was never edited. It may therefore contain some parts that are poorly written. The English translation needs to be done now regardless. Editors in the Spanish Section will work on the original later. If you detect any inconsistencies, i.e. slips in logic (but not stylistic problems) that should be brought up with the Spanish editors and/or authors, please bring these to the attention of the reviser by inserting a note in square brackets next to the corresponding part of your translation and highlighting it in yellow. The reviser will then follow up.

For example, if the original says: *El programa tiene tres objetivos: mejorar el clima de negocios y atraer inversiones*. Your translation might read: The programme has two [N.B. the original says three] objectives: to improve the business climate and attract investment. There is no need to explain why you put two instead of three. In this case you could resolve the problem. If you cannot, just translate the text literally and put a comment like [this contradicts what was said in line 21] or [this makes no sense]. Please note that inconsistencies do not refer to stylistic flaws such as poor punctuation or syntax, but to content errors that affect the logical coherence of the text.

If you prefer, instead of highlighted insertions in square brackets you can flag the inconsistencies using the “comments” function in track changes. Please remember that all translations are expected to meet the highest standards of excellence. Thus, apart from some duly highlighted inconsistencies which the reviser will deal with, your English version of the document should be fit for publication, regardless of the state of the original.

N.B. You are not expected to transcribe the numbers in the table, just give instructions for the Text-Processing Unit (TPU) who will format the document. Put your instructions at the position the inserts should go in the text and highlight them in yellow. For example:

**TPU:**

column headings: xxxxx, yyyy, zzzzz

row headings: aaaa, bbbbbb, ccccc,

Please transfer numbers as they are.

Alternatively, if you have an electronic version of the original, you can cut and paste the table and type over the text

The amount of contextual information provided in Test Two marks another way in which it differs from Test One. For comparison purposes, the main features, again based on Nord's analysis scheme (2005: 41-141) with some additions, are presented in Table 9.3.

Table 9.3. Comparison of the source texts used in Test One and Test Two<sup>27</sup>

<b>Factor</b>	<b>Test One test-text</b>	<b>Test Two test-text</b>
Sender	Unidentified columnist or essayist	Several, unidentified, report writers, presumably from a specialized agency/unit (see instructions for test-takers in section 9.2.7)
Sender's intention, motive/function of the communication	Communicate opinion and make the reader reflect; entertain with descriptions used to make points	To provide input for regional meeting (see instructions for test-takers in section 9.2.7)
Audience	Readers of news articles on current affairs	Participants in meeting of experts on natural disasters (see instructions for test-takers in section 9.2.7)
Medium	Written text taken from unidentified journal or newspaper	Written report
Place	Uruguay presumably	Latin American and the Caribbean
Time	Not stated, would be presumed to be during the 12 months before the examination	End of 2010 (see instructions for test-takers in section 9.2.7)
Subject matter	The economic boom in Uruguay	Impact of natural disasters on the region
Content	Description of reactions to economic boom and impact on consumers, warning that undeserved economic boom teaches nothing; no gaps of cohesion or coherence	Description of impact of natural disasters in 2010, followed by statement that the region needs to hold talks to ensure the nations responsible for climate change cover the costs; text contains redundancies and inconsistencies
Presuppositions	Knowledge of Uruguayan current affairs (head of State mentioned by name)	Basic knowledge of current affairs, the region and basic economic terms
Composition	Seven paragraphs, each introducing new angle; thematic progression shaped by graphic down-to-earth imagery, parallel structures and repetition	Six paragraphs and one table; faulty references interfere with progression
Non-verbal elements	None	Table of data typical of text type

<sup>27</sup> Features included are taken mainly from the factors of ST analysis recommended by Nord (2005: 41-141).

<b>Factor</b>	<b>Test One test-text</b>	<b>Test Two test-text</b>
Lexis	Juxtaposition of concrete and abstract terms; figurative language used, including popular sayings and metaphors; appellative use of first person plural	Plain language; some abstract language; miscolllocations; register is typical of official reports; two technical words recognizable from English
Number of low-frequency words <sup>28</sup>	12	4
Number of words and sentences	505 in 24 sentences (average 21 words per sentence)	508 in 10 sentences (average 51 words per sentence)
Sentence structure	Use of parentheses; generally long sentences; use of short balanced sentence to highlight main point	Very long sentences; use of parentheses and semicolons to try to clarify
Punctuation	Stylistic punctuation, such as dashes and ellipsis, to add expressivity and emphasize words	Use of parentheses
Effect	Reader receives the message and is aware of an issue probably not considered beforehand	Reader receives a lot of data and the suggestion that the region should enter into talks to hold the United States and other countries accountable for climate change

### 9.2.8 *Design of the grading scheme in light of the skills-knowledge hierarchy*

In keeping with the findings of the literature review, a multi-trait grading scheme was chosen as the most appropriate way to score performance. This is referred to as grading scheme B. Grading the components of the construct individually would enable them to be weighted according to the profile sought. This meant designing rating scales for evidence of the following: flow, clarity, coherence, correctness (grammar, punctuation and spelling put together to reduce the number of scales used, as the instrument was becoming rather cumbersome and because these particular qualities were not that highly rated in the scatter chart analysis), tone and register, accuracy (understood to cover the ability to capture nuances and convey the intended effect of the TT), idiomaticity, completeness and target language vocabulary. Accuracy is heavily weighted in this scheme because, in addition to rating the ability to capture nuances, graders rate SL knowledge and TL vocabulary, which also contribute to accuracy. Clarity is also weighted heavily because correctness, flow, idiomaticity and coherence all contribute to clarity. The abilities to recast sentences in the

<sup>28</sup> This refers to words M.A. students might be expected to want to look up in a dictionary, based on their frequencies and their similarity to English equivalents. Frequencies were calculated using a corpus (Davies, Mark. (2002-) *Corpus del Español: 100 million words, 1200s-1900s*. Available online at <http://www.corpusdelespanol.org>).

TL, tailor language to the readers' needs and translate poorly written STs are tested through the task design but are not rated separately as they are reflected in all the different aspects of TL writing that are individually assessed. The abilities to follow complicated instructions and to maintain quality under time pressure are similarly measured indirectly. The detection of inconsistencies is measured objectively (inconsistency detected or not), reminiscent of Nord's scoring of the successful resolution of translation problems (2005: 188-189). Bearing in mind the fineness of the distinctions that graders would be able to make, as recommended by Alderson et al. (1995: 111), five-level scales were initially designed for flow, clarity, correctness, accuracy idiomaticity, and target language vocabulary, and three-level scales for tone and register, coherence, and completeness. In both cases the midpoints were intended to indicate levels of performance bordering on the unacceptable.

The scales, component definitions and instructions for grading scheme B are presented below, together with the lists that were given to graders of challenges posed and features to look out for in each test. The changes made to the rating scales in the light of the results of the pilot and the training session held for graders are also indicated. The wording of the descriptors and the component definitions was intended to encourage graders to base ratings on all the evidence of a component found in the TT and deliberately echoes that used in the construct definition to enhance the validity of the process, as recommended by Cushing Weigle (2002: 51). As Fulcher and Davidson also point out, it is essential to "ensure that the rating scales focus attention on the constructs of interest, and that each of these constructs is carefully and empirically linked to the target domain of interest" (2007: 257). Descriptors were also provided for each level of each rating scale to try to keep graders on the same page. Even though the terms "some", "several" or "many" might be open to highly subjective interpretations, I thought it was better to use them when possible than to specify a maximum number of errors for each level as graders would then be able to factor in the severity as well as the frequency of any failing.

Graders were asked to indicate whether a candidate should pass or fail, in order to help determine the pass mark once the scores were in. Since the scheme would be applied to the mock traditional test as well, the instructions also refer to Test One. The small changes in the rating scales introduced after the scheme had been tried out on a few scripts in Test Two at the beginning of the trial (see section 9.4.7.2) are indicated and explained in footnotes.



### INSTRUCTIONS FOR GRADING SCHEME B

Dear volunteer,

Thank you for volunteering to help grade the test papers. Under this scheme you will be asked to rate various aspects of the target text as manifestations of different components of the profile sought. Each aspect or component is defined and has its own rating scale. The ratings scales vary because different concepts lend themselves to different degrees of specificity. Some features of the translation may affect the rating of more than one component. Do not worry about this, just rate each component individually. The final score will be calculated using a weighted formula that is based on empirically identified recruitment priorities.

#### Suggested procedure

- Bear in mind that candidates are applying for entry-level positions and are not expected to be familiar with in-house spelling or usage rules.
- Before you begin, familiarize yourself with the source text and the definitions of the components and their rating scales.
- Read through the translation without stopping, in order to get an idea of the flow and clarity of the translation. Rate those two components using the marking sheet.
- Read through the translation again more slowly and rate the other components. You can mark the text in any way you wish and you can adjust your flow and clarity ratings if necessary.
- In test two, test-takers were asked to flag and, if possible, correct any inconsistencies in the original. A check list is provided for those they are expected to detect.
  - If they noticed that the text says *cuatro decadas* and the table *1970-2000*, but did not realize it was the table that was wrong, not the text, put ½ next to the yes.
  - If they fail to detect one of the listed inconsistencies, penalize them in the checklist and maybe under coherence, not under accuracy.
  - If they point out an inconsistency but do not correct it even if they could, do not penalize them.
  - If they point out inconsistencies that are not in the list, ignore them or adjust the coherence, knowledge of Spanish and/or accuracy ratings as appropriate.
- When you have finished, make sure to indicate whether you think the candidate should pass or fail and to comment on the candidate's apparent strengths and/or weaknesses. These comments will be forwarded anonymously to the test-taker.

#### CHALLENGES POSED AND THINGS TO LOOK OUT FOR IN TEST ONE

- Title: Does the translation of the title make sense? Has the idea of a missed learning opportunity been conveyed? Use it to help gauge internal logic (coherence).
- Figurative language: The metaphors and figures of speech and the sentence about hitting the mall might be useful for rating idiomaticity and TL vocabulary.

- Words and expressions test-takers are not expected to know: cardinales, cátedra, abarrotar, rémoras, tesón, en tiempos de vacas gordas, rienda suelta. They should nevertheless be able to guess them or at least put something sensible.
- Tone and register: journalistic, newspaper editorial, popular sayings and contractions are appropriate. Idiomaticity is very important in this text.
- The long sentence “quiero señalar....” (line 24-29) is good for assessing ability to write clearly.

### **CHALLENGES POSED AND THINGS TO LOOK OUT FOR IN TEST TWO**

Logical thinking: This test aims to measure candidates’ analytical skills by seeing how well they handle a poorly written source text. Even though the text is not well written, the candidates are expected to be able to work out what is meant and produce a well-written target text that makes sense. Nonsense should be penalized under accuracy, knowledge of source language and/or coherence, as applicable, using the scale descriptors.

Although about natural disasters, the text is not very technical and individual words are not expected to pose comprehension problems. Test-takers had access to the Internet, if not much time, to research terms.

General knowledge is not being measured in this test, so do not penalize candidates for their ignorance of the fact, for example, that Santa Lucia is Saint Lucia in English.

In addition to the ability to detect inconsistencies, look for the ability to produce a well-written translation despite the poor quality of the original.

See annotated test paper attached hereto.

### **Rating scales**

#### **Smoothness of flow**

- 5 Ideas and sentences flow smoothly throughout
- 4 Minor interruptions of flow do not affect overall readability
- 3 Some interruptions of flow affect readability in parts
- 2 Readability poor in several parts
- 1 Readability poor throughout

#### **Clarity**

- 5 Strikingly clear throughout
- 4 On the whole clear throughout
- 3 Meaning unclear in some parts
- 2 Meaning unclear in several parts
- 1 Incomprehensible in many parts

**Tone and register**

- 3 Appropriate throughout
- 2 Appropriate in almost all of the text
- 1 Inappropriate in more than one part

**Evidence of knowledge of source language**

- 5 No evidence of failure to fully understand the source language
- 4 Some evidence of failure to understand obscure words but not structures
- 3 Evidence of failure to understand obscure words and/or the odd structure<sup>29</sup>
- 2 Several examples of apparent misunderstandings of vocabulary and/or structures
- 1 Many examples of apparent misunderstandings of vocabulary and/or structures

**Evidence of ability to transfer detailed levels of meaning (accuracy)**

- 5 Every nuance and emphasis of the original is captured in the translation
- 4 Most nuances are captured. Slight loss of meaning does not affect the message received
- 3 Several small distortions at detailed level of meaning begin to affect overall level of accuracy/A serious distortion affects part of the message received
- 2 Reader receives wrong information on a few occasions
- 1 In several instances the reader receives the wrong message

**Evidence of broad active target-language vocabulary<sup>30</sup>**

- 5 Excellent word choice (including technical vocabulary) throughout enhances all aspects of translation
- 4 Word choice on the whole enhances all aspects of translation
- 3 Word choice adequate for purpose
- 2 Inadequate word choices occasionally affect transfer of meaning and style of translation
- 1 Several wrong word choices and combinations affect transfer of meaning and style

**Idiomatic language use**

- 5 Language is wholly idiomatic throughout, strong collocations
- 4 The occasional example of unidiomatic language use
- 3 A few examples of unidiomatic language use
- 2 Several examples of unidiomatic language use
- 1 Does not read like an English text

<sup>29</sup> The “and/or” possibility was introduced as being more useful since it made the scale cover all possibilities. The piloted version had read as follows:

- 3 Evidence of failure to understand obscure words and, on the odd occasion, structure
- 2 Several examples of apparent misunderstandings of vocabulary and structures
- 1 Many examples of apparent misunderstandings of vocabulary and structures

<sup>30</sup> The references to terminology were removed from the piloted version since graders said it made them think of technical vocabulary which was not really applicable. The piloted version read as follows:

- 5 Excellent word choice throughout enhances all aspects of translation
- 4 All terminology correct. Word choice on the whole enhances all aspects of translation
- 3 Most terminology correct. Word choice adequate for purpose.
- 2 Some terminology problems. Inadequate word choices occasionally affect transfer of meaning and style of translation.
- 1 Several wrong word choices and combinations affect transfer of meaning and style.

**Coherence (internal logic)<sup>31</sup>**

- 5 Wholly consistent and logical throughout
- 4 Minor internal inconsistency or slip in logic does not affect message received or interfere with comprehension
- 3 Minor internal inconsistency or slip in logic affects message received or interferes with comprehension
- 2 Inconsistent in a more than one instance
- 1 Inconsistent in several instances

**Correctness<sup>32</sup>**

- 5 All grammar, punctuation and spelling correct
- 4 A few punctuation and spelling errors, all grammar correct
- 3 The odd grammar problem or several punctuation and spelling errors
- 2 Several grammar problems or many punctuation and spelling errors
- 1 Several grammar problems and many punctuation and spelling errors

**Completeness**

- 3 Target text is complete: no sentences or sentence parts have been omitted
- 2 Part or all of a sentence has been left out
- 1 More than one phrase or sentence has been left out

**Inconsistencies and logic problems checklist (only for Test Two, of course)**

Inconsistency spotted? Indicate Yes or No

Line 14: *servicios básicos* are not a risk (the lack of them is)

Line 15: *Cuadro arriba* is a faulty reference (it is below)

Line 15: *Cuatro décadas -1970-2000* (should be 2010)

Table: 1000% (should be 100 %)

Table: change decimal commas

Line 34: *Los menos vulnerables* (should be the most)

Line 37: *EEUU y las naciones hoy desarrolladas* (should be the United States and *other* countries)

**The candidate should (please circle): pass fail****Strengths and weaknesses:**

<sup>31</sup> Graders found the piloted version did not have sufficient levels to reflect differences in performance. It had read as follows:

- 3 Wholly consistent and logical throughout
- 2 Minor internal inconsistency or slip in logic does not affect message received or interfere with comprehension
- 1 Internal consistencies or slip in logic affect message received or interfere with comprehension

<sup>32</sup> Grammar is weighted slightly more heavily than punctuation and spelling; this was deliberate. The concepts were rated together on the same scale because I feared having too many scales. After the trial, however, I realized grammar should have been rated separately.

## COMPONENT DEFINITIONS

**Smoothness of flow:** affects the readability of the text

*What to look for: cohesion, easily perceptible connectors between sentences, sequence of tenses, syntax, word order, use of clauses, cadence, emphasis, correct parallel structures, absence of awkward syntax.*

**Clarity:** reader gets the message immediately and without any problem

*What to look for: ideas are stated as directly as possible using as few words as needed, clear use of reference markers (this/that, clauses, etc.), clear identification of what-goes-with-what, correctly placed modifiers, judicious use of punctuation, repetition and syntax to ensure there is no ambiguity, no redundancies.*

**Idiomatic language use:** language of the target text sounds natural, wholly English

*What to look for: observes stylistic and rhetorical preferences of English, correct collocations, colligations, set phrases, plural nouns for generalizations, etc.*

**Word choice as evidence of target language vocabulary:** range and appropriateness of words used.

*What to look for: correct usage of terminology, appropriate use of vivid language, use of synonyms to avoid inelegant repetitions, words that are appropriate for institutional documents, as applicable.*

**Evidence of ability to capture all the nuances of the original:** how well and to what extent the detailed levels of meaning of the original (and no more) are conveyed. Be careful you are not grading source-language comprehension or target-language skills here (which are rated separately).

*What to look for: slight distortions or losses of meaning, shifts of emphasis, inappropriate additions. Check modifiers are modifying what they should be and to the right extent. Omissions and additions that do not affect the basic message, but do affect the nuances are penalized here. Penalize blatant omissions under "completeness".*

**Evidence of knowledge of the source language:** as displayed in apparent comprehension of source text

*What to look for: inaccurate translations attributable to a misunderstanding of the original, distinguish between apparent mistranslations of individual words and misunderstandings of structures.*

**Coherence (internal logic):** the internal consistency and logic of the target text

*What to look for: consistent use of terminology (even if wrong), logical connection of ideas and sentences, no new contradictions introduced into text (such as inappropriate or misleading additions) no nonsense or tautologies.*

**Correctness:** grammar, spelling and punctuation of the target text

*What to look for: grammar, punctuation and spelling (including capitalization) errors. Test-takers are expected to apply the same spelling and punctuation rules throughout (i.e. US or UK; Oxford commas or not, etc.).*

**Appropriate tone and register:** level of formality

*What to look for: choice of neutral or politically correct terms, as appropriate, relatively high register, avoidance of journalistic style in Test Two, reflecting awareness of*

*appropriate language for use in international organizations, connotations of words and norms of English rhetoric.*

**Completeness:** the extent to which the content of the original is in the translation  
*What to look for: This refers to blatant omissions of content that seem to be the result of oversight rather than deliberate choice. Deliberate omissions should be evaluated under the coherence criterion.*

### 9.2.9 Calculating the final scores

Given the priorities established in the scatter-chart analysis, the final scores of each candidate were calculated by weighting the scores of the category A components of the skills-knowledge set, both individually and as a group, more heavily than the scores of the non-A components. The weighting used and its impact on scores is presented and discussed in Chapter 10.

### 9.2.10 Fairness, validity and reliability considerations

Validity, reliability and fairness are key considerations in any testing exercise. The measures taken to maximize these qualities in the profile-adapted test are summarized below.

#### 9.2.10.1 Fairness

Fairness is a quality that most, we hope, would claim should underpin all testing, especially high-stakes recruitment testing. In the domain we are interested in, fairness affects not only the test-takers, but also the test users, the organizations themselves. They need to be sure that the candidates who pass their tests do so for the right reasons.

Several measures were taken to ensure fairness in the profile-adapted test:

1. Bias was minimized by removing the need for particular subject knowledge, and no research skills beyond checking a term in an online dictionary or encyclopedia were expected to be activated.
2. Test-takers would take the tests under the same conditions as far as equipment options, access to resources and timing were concerned.
3. Instructions regarding expectations and criteria for success were clear so that all participants would be attempting to achieve the same thing.

4. Every effort would be made to ensure that graders adhered to the same criteria by training them, trialing the scales and measuring inter- and intra-rater reliability if possible.
5. As to cheating, some test-takers would take the test under invigilation; others would have to be trusted not to cheat. It was hoped that anonymity would also help keep participants honest.

#### 9.2.10.2 *Validity considerations*

The following measures were taken to gauge and maximize the validity of any inferences made on the basis of the test results (based on the checklist set out in Alderson et al. 1995: 192-193):

1. *Content validity*. A domain analysis was performed and ten experts from the domain were interviewed to determine how representative the test task was of domain tasks.
2. *Face validity and response validity*. According to Alderson et al. “face validity will affect the response validity of the test” (1995: 173). The feedback questionnaire designed for participants in the trial included questions intended to determine whether the design of the test affected attitudes towards it (and hence possibly performance), whether the task activated the skills and knowledge types it was supposed to and whether test-takers tried their hardest and took the exercise as seriously as hoped. The questions asked are presented in Appendix 15.
3. *Concurrent (criterion) validity*. Rankings would be compared with those provided by teachers in the case of the university students and by the simple self-assessments reported in the feedback questionnaire. The performances of students and professionals would be compared.
4. *Construct validity*. The test task and the grading scheme were designed very much with the definition of the components of the construct in mind. The task design was moreover almost entirely based on the domain we wished to generalize to, the constructs we wished to measure and the test-takers. This is reflected in the task’s authenticity and its activation power. The measures taken to maximize these two qualities within the practical limitations of the experiment (time and resource constraints) are outlined below:
  - a. *Measures to maximize authenticity*. The task design is based on the analysis of the domain. Some aspects of authenticity, namely the level of SL and topical knowledge and terminology and fact-checking required to complete domain tasks,

are sacrificed to minimize bias and for practical reasons. The possible influence of the interplay with these other skills on performance as a whole are therefore not taken into account, and any inferences made are limited to translation work not involving research skills or advanced lexical knowledge in L2. The concentration of construct activators in the ST also undermines the task's authenticity, but enhances the scope and relevance of the content (important for content validity) and the amount of evidence available for assessing each construct component (important for reliability).

- b. *Measures to maximize activation potential.* The task is designed to activate the constructs by systematically concentrating the kinds of features in the ST that call the components of those constructs into play. These features are also the difficulty drivers of the tasks. This is done to increase the construct validity of any inferences made about abilities. As Cushing Weigle puts it, “[c]onstruct validity is enhanced when the factors that contribute to difficulty are those that are included in the definition of the construct” (Cushing Weigle 2002: 50).

### 9.2.10.3 Reliability considerations

For the purposes of comparing performance on two different types of test, validity considerations are more important than reliability considerations. Nevertheless, one planned measure to increase the reliability of the test scores was to have all scripts marked anonymously by two different graders. It was also hoped that it would be possible to examine inter-rater and intra-rater reliability. The task specifications provided in this chapter should also make it possible to repeat the experiment and perhaps perform some kind of parallel-form reliability testing. Such an investigation, however, would require several trials and subsequent observations of performance over time; they lie far beyond the scope of this project.

## 9.3 The pilot

The two tests were piloted in February 2011. The two main objectives of the pilot were (1) to check whether the level of difficulty of the tasks was appropriate, and (2) whether the grading schemes were manageable. It was essential that both acceptable and outstanding candidates should be able to perform to the best of their ability and that graders would be



able to work with the schemes. Participants were asked to complete a feedback questionnaire after each test. Questions on the test conditions, timing and clarity of the instructions were included. The participants were also all interviewed in person or by e-mail.

### 9.3.1 *Participants*

A profile was drawn up of the typical applicant for in-house posts at organization Z in terms of age, sex, educational background, experience and expected attitude to the task. The information was obtained from vacancy announcements and informal consultations with those involved in recruitment at Z. This research showed that applicants would mostly be aged between 22 and 35, have at least a Bachelor's degree and usually one or two years' experience, but little experience in IGO translation. Over 60% would be female. An analysis of the profiles of *successful* candidates in previous examinations for permanent positions at Z revealed a slightly different demographic: 52% of successful candidates were female and their distribution by age group was as follows: 25.9% were aged 25-30 years, 18.5% 31-35 years, 18.5% 36-40 years, 14.8% 41-45 years, 14.8% 46-50 years, and 7.5% 51-55 years. Age (and/or experience) and being male seems to have a bearing on success in this organization's current examinations.

In light of these findings, three people were selected to represent the range of the typical target population in the pilot of Test One and Test Two: a 23-year-old female student taking a part-time M.A. in Translation at a university in the United Kingdom; a 27-year-old male translator who had been working in-house for an IGO for just under a year; and a 45-year-old female professional translator with 12 years' experience in translation, including seven years in IGOs.

### 9.3.2 *Test conditions and mode of delivery*

All participants received the tests by e-mail and sent in their translations by the same means. They completed the mock traditional test on one day and the profile-adapted test on another. All followed the instructions regarding timing and the use of resources very closely. They were all unpaid volunteers.

### 9.3.3 Grading

All the scripts were graded by the same experienced IGO reviser: once using the two mock traditional methods (scheme A) and once using the multi-trait scheme designed for the profile-adapted test (scheme B). Neither scheme was found to be particularly difficult to apply.

### 9.3.4 Results of the pilot

Both tasks and grading schemes discriminated among the three participants as expected, with the older professional scoring the highest on both tests, albeit only just, and the student the lowest, as shown in Table 9.4.

Table 9.4. Scores obtained in the pilot of Test One and Test Two

Test-taker	Female M.A. student		Recent male recruit		Experienced female professional	
	One	Two	One	Two	One	Two
Test						
Grading Scheme A intuitive (20)	12	10	15	12	17	18
Grading Scheme A points	-22.5	-14.5	-10	-9	-5.5	-8.5
Grading scheme B simple total	28/44	21/52	39/44	43/52	38/44	48/52
Flow (5)	4	1	5	5	4	5
Clarity (5)	4	1	5	3	4	5
Tone and register (3)	1	2	3	3	3	3
SL knowledge (5)	3	3	4	5	4	5
Accuracy (5)	2	3	4	2	4	4
TL vocab (5)	3	2	5	5	4	4
Idiomacity (5)	4	3	5	5	5	4
Coherence (3)	2	1	2	3	3	3
Correctness (5)	2	1	3	4	4	5
Completeness (3)	3	1	3	3	3	3
Detection of inconsistencies (8)	n/a	3	n/a	5	n/a	7

All participants said the instructions for both tests were clear. Both the student and the young professional said that both tests were challenging in different ways, but not impossible to complete, although the student failed to finish the last sentence of the profile-adapted test because, by her own admission, she failed to manage her time correctly. The student found the inability to consult resources frustrating in Test One. She found Test Two difficult to complete on time:

It was much easier than test one as we were allowed to consult dictionaries and the internet. On reflection, I would say that test one was more of a 4 on the above scale

[1-5 for difficulty] whilst I would rate this one a 3 as the inconsistencies made it a difficult task but owing to the text type and the resources available, it was easier to understand and less difficult to find translations for this text. The text type is what I expected from a test piece for [org Z]. I feel like I was more able to show how I can translate with this test. However, I did struggle to understand some of the long sentences and produce an eloquent translation in the time given, although I think this was mainly down to my poor time management.

The recent recruit had the opposite experience:

The fact that I had access to resources suggested to me that the test might be more difficult. I was not mistaken! While I succeeded in conveying the message of the source text, I was ultimately frustrated that I was unable to polish the text as I had done the previous one. To an extent, not having access to resources was liberating as I could focus on the translation instead of worrying about the time I was taking to look things up and whether the translation I found was the most appropriate one. While what I produced was passable, I am convinced that I could do a better job with additional time. Again, the test was fair to the extent that any test is fair. I felt that my time-management and analytical skills were being scrutinized more than my creative flair. The time constraints put me in mind of an urgent translation job, insofar as I was being tested on how well I could overcome the poor draftsmanship of the original (in this case) to complete the translation within the specified time frame.

The more experienced translator commented:

It [Test Two] was a very different test. I found it harder to produce a decent translation because I was trying to strike the balance between accuracy and readability. Test One was a far more enjoyable experience even if I struggled with some of the vocabulary. Test Two was far more typical of the work I regularly do at [Z] although I usually work from French.

According to the grader, both grading schemes were manageable, although she felt it would take practice to feel confident using the scales of scheme B. Obviously the

detection of inconsistencies had to be ignored when scheme B was being applied to Test One. The coherence scale was also slightly more difficult to apply to the mock traditional test, but not impossible.

### *9.3.5 Outcomes of the pilot*

In light of the comments on timing, it was decided to shorten the profile-adapted text from 569 to 500 words and to cut the mock traditional one to the same length (from 506 words). It was also decided to extend the time allowed for the completion of both tasks slightly to 1 hour and 20 minutes. This would also enhance the comparability of the two tasks. I was reluctant to extend the time frame by much more for fear of putting people off participating in the trial and because the ability to maintain quality while under pressure of time was a key component of the skill-set sought at organization Z. Given that it is practically impossible to rate incomplete translations in any sensible way, it was also decided to stress the need to finish the task in the instructions sent to test-takers in both tests. This would not be wholly unauthentic since translators from time to time simply have to sacrifice accuracy and style for the sake of expediency, even at IGOs, including Z. Of course, the best translators in some respects are those who can sacrifice the least.

## **9.4 Organization of the test trial**

The trial was held in April-July 2011. The test administration phase ran from April to May, and the test grading phase from June to July. Participants consisted of university students and professional IGO translators and revisers. All were unpaid volunteers.

### *9.4.1 The test-takers*

I decided to invite students on programs that had a reputation for producing graduates of the desired caliber to participate in the test trial. Not only would there be a strong chance that some of them would match the profile very closely, but all of them would, at least in theory, be interested in practicing taking a recruitment test that would be graded by experienced revisers at an organization for which they might one day wish to work. Invitations were therefore sent to the coordinators of Master's translation programs at the universities of Bath, Leeds and Westminster, in the United Kingdom, the University of

Geneva in Switzerland, and the Monterey Institute for International Studies (MIIS) in the United States. All the institutions agreed to participate and to provide expected performance indicators of the test-takers. Between four and ten students from each institution volunteered, and a total sample of 40 student participants was thus obtained. Eight translators from different branches of organization Z were also invited to participate for the purpose of criterion referencing. All 48 test-takers were promised anonymity and signed release forms.

#### 9.4.2 *The graders*

It was thought important to have the translations graded by insiders, by people with a clear idea of what a good translation is by the standards of the organization, as recommended by Hatim and Mason (1997: 205), namely revisers and heads of service at branches of organization Z. This actually proved to be one of the aspects of the trial that had the universities interested in participating. They wanted to see if their assessment of their students tallied with that of professionals in the field.<sup>33</sup> The revisers, for their part, were interested in the outcome of the research, having taken, graded and in many cases set translation tests over the course of their careers. Information on the graders is presented in Table 9.5.

Table 9.5. Distribution of the 10 revisers who graded the translations in the test trial, by sex, age and experience

Sex		Age (years)				Experience as a translator/reviser (years)			
M	F	40-49	50-59	60-69	Over 69	10-15	16-20	21-30	>30
7	3	2	5	2	1	3	3	3	1

#### 9.4.3 *The administration of the two tests*

In some cases, the participants from the same institution took each test at the same time and in the same room; in others, they took them at home and at slightly different times. In all cases, the following testing conditions were applied:

<sup>33</sup> At the end of the trial, the coordinators and the students received a table of the scores and rankings obtained on both tests in individually addressed e-mails. In the table, the test-takers and the institutions were identified by numbers. The coordinators of the Master's programs and the students were told the number of their institution; the students were told their individual test-taker numbers as well.

- All test-takers received the written instructions.
- Each test was taken in a separate session (i.e. they were not taken immediately after each other). The idea was to ensure there was a break and a chance to refresh the brain between tests.
- Test One was taken before Test Two.<sup>34</sup>
- Both tests were taken on computers: they were received by e-mail, completed in Word and submitted by e-mail.
- The time limits were strictly adhered to. The times on the e-mails to which completed translations were attached could be used to ensure this.
- No sources of any kind were consulted in Test One.
- Internet and bibliographical (but no human) sources could be consulted during Test Two.

Of course the non-consultation of resources in Test One in some cases was a matter of trust, as was the non-consultation of other people in both tests, but having offered anonymity, there was little reason to suspect participants would cheat.

#### *9.4.4 Feedback from participants*

All participants were asked to complete a questionnaire after each test (see Appendix 15). The objective of the questionnaires was to identify the age, sex, experience and academic background of the participants and to obtain their own appraisal of their performances, as well as their observations about the tests.

#### *9.4.5 Preparation of the test TTs for grading*

All TTs, once received, were converted into the same standard Word format (Times New Roman 12, justified, double spaced). Names or student numbers were removed, and a unique candidate number (a random number between 10 and 99) was inserted in the header. The scripts were then saved under the candidate number, e.g. Test One 24.

---

<sup>34</sup> I was not checking for learning effect, in part because I could not control what experiences the test-takers would have between tests. This could be an area for further research.

#### 9.4.6 *The organization of the grading*

Having 48 participants meant 384 assessments to arrange (each translation to be graded by two graders using two schemes:  $48 \times 2 \times 2 \times 2 = 384$ ). The problem was, of course, that IGO revisers and heads of service are busy people who hardly relish the idea of giving up their time to grade students' translations. To make the grading feasible, it was decided that I would be the first grading team for both tests and both schemes (since I did relish the idea of poring over all the scripts), but would leave a gap of one month between applying one scheme and the other to minimize the influence of one experience on the other. The other graders would make up the second team, which would be split into those using the combined intuitive/points-deduction scheme (scheme A) and those using the profile-adapted multi-trait scheme (scheme B). To avoid a systematic influence, team 1 and team 2 received their scripts in two different orders, and all graders had at least one translation done by one of the professional translators participating in the trial in their set of scripts.

#### 9.4.7 *Preparation of the teams of graders*

##### 9.4.7.1 *Those working with scheme A*

As in most traditional practice, and as usually happens at organization Z, the 10 graders working with the intuitive and points-deduction methods were just given the written instructions, but no other guidance. They were not, for example, given samples of translations that would achieve certain scores or descriptions of the types of errors that would be considered minor or major, and no practice session was arranged.

##### 9.4.7.2 *Those working with scheme B*

Alderson et al. describe the difficulties involved in writing band descriptors but point out that “[u]ltimately, since we are not creating an equal interval scale, what will matter is whether assessors can use the scales and agree on their understanding of the descriptions that define the levels” (1995: 82). The grader in the pilot had also recommended that the graders practice. A training session for scheme B graders was therefore held to practice applying the scales and try to get all graders to apply them the same way. As Cushing Weigle points out, it is not just a matter of having appropriate scoring criteria, but of making sure graders adhere to them (2002: 51).

Two of the three graders in team 2 who were to apply scheme B met with the team 1 grader (me). The third could not attend the session. Three basic texts, which at first glance seemed to represent a weak, a mediocre and a strong performance, were selected for practice. After working with those three scripts, another six scripts were graded by the three graders individually and the scores were compared. These scripts were chosen at random, but known to contain at least two translations done by professionals, in order to include the full range of performances expected. The training session lasted seven hours, by the end of which all participants said that they felt comfortable using the scales and felt that they were on the same page. A few adjustments were found to be necessary to make the scales as usable and meaningful as possible. These are indicated on the grading sheets presented in section 9.2.8.

The graders occasionally consulted with one another over the next two weeks about how to rate a particular aspect of a translation, but only rarely. The researcher met briefly with the fourth grader to hand over and explain the written instructions, but this grader did not have the opportunity to practice grading.

#### 9.4.7.3 Feedback from graders

All graders were asked to complete the form below after handing in their score sheets and feedback for the test-takers.

<p>Feedback from graders using scheme A</p> <p>How many scripts did you grade? _____</p> <p>Which test did you grade? _____</p> <p>How long did it take you to grade all the scripts/each script? _____</p> <p>Comments on the scheme. Did you find it difficult to assign the score out of 20? If yes, why? Did you find it difficult to use the points scheme? If yes, why? Perceived advantages, disadvantages, flaws, virtues? Suggested modifications? Comparisons with other schemes you have used in the past?</p>
---

<p>Feedback from graders using scheme B</p> <p>Circle which test you graded: One    Two</p> <p>How many scripts did you grade? _____</p> <p>How long did it take you to grade all the scripts/each script? _____</p> <p>Please comment on the scheme. Was the scheme simple or complicated to use? Were there particular components that were difficult to rate? Perceived advantages, disadvantages, flaws, virtues? Suggested modifications? Comparisons with other schemes you have used in the past?</p>
--





## Chapter 10 Results of the test trial

Of the 48 participants in the test trial, 46 completed both Test One (the mock traditional test) and Test Two (the profile-adapted test) and their translations are therefore the ones included in the analysis.

### 10.1 The test-takers

Of the 46 test-takers included in the analysis, 37 were female and 9 were male. All held Bachelor's degrees and 10 had Master's degrees (5 were taking a second M.A.; the other 5 were professional translators). Of the 38 who answered the optional open question on the subjects studied, 29 had a degree in languages or languages and literature. The other 9 had degrees in global studies, chemistry, art and design, sociology, American sign language interpreting, physics, social work, aviation, and politics and international studies. The distribution of the participants by age, sex and experience in each university group and among the professional translators working for IGOs is shown in Tables 10.1, 10.2 and 10.3. Just over half (24/46) of the participants (all students) were under 26; the fourteen other students were aged between 26 and 35. One fifth of participants were male. With one or two exceptions, none of the students had experience working for IGOs, and few had more than 1 or 2 years' experience outside IGOs either, if at all. The IGO group is the most heterogeneous one in terms of age and experience.

Table 10.1. Participants, by institution, sex and age

Institution	Sex		Age (years)							
	Male	Female	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60
Uni-1		10	8	1	1					
Uni-2	2	6	4	1	3					
Uni-3		4	3	1						
Uni-4	1	6	4	1	2					
Uni-5	3	6	5	4						
IGOs	3	5		1		1	2	1	2	1
Total	9	37	24	9	6	1	2	1	2	1

Table 10.2. Participants, by institution and years of translation experience in IGOs

Institution	none	<6 m	6-12 m	1-2 ys	3-5 ys	6-9 ys	10-15 ys	16-20 ys	21-30 ys
Uni-1	8	1	1						
Uni-2	6			1	1				
Uni-3	3	1							
Uni-4	7								
Uni-5	7	1	1						
IGOs				3		2		1	2
Total	31	3	2	4	1	2		1	2

Table 10.3. Participants, by institution and years of translation experience outside IGOs

Institution	none	<6 m	6-12 m	1-2 ys	3-5 ys	6-10 ys	11-15 ys
Uni-1	4	4	1	1			
Uni-2	3	1	1	3			
Uni-3	2	2					
Uni-4	4	1	1		1		
Uni-5	4	2	1	1	1		
IGOs	2	2		3			1
Total	19	12	4	8	2		1

## 10.2 The test scores

The scores obtained in Test One by applying the dual intuitive and points methods (scheme A) are shown in Appendix 16. Five revisers, referred to as B, C, D, F and H, made up the second team of graders; I was the first team. The scores obtained in Test Two by applying the multi-trait grading scheme are shown in Appendix 17. Four revisers, referred to as A, B, G and K, made up the second team of graders; I was the first team.

## 10.3 Comments made by test-takers

### 10.3.1 On the instructions

All participants reported that the instructions in Test One were clear, although one person commented that they were slightly “unfriendly”, which was interesting since they were based on those actually used by Z. One person also said the instructions in Test Two were “much better”, but without explaining why. All but two participants said the instructions in

Test Two were clear. Those two said they wondered why they were supposed to leave decimal commas in the English. They had obviously misunderstood the instructions which had given “Please transfer numbers as they are” as an example of the type of in-text instructions that should be given to the text-processing unit (see section 9.2.7). They should have realized, like all the other test-takers, that in this case, the instructions should have referred to the need to change the decimal commas. Possibly rather harshly, they consequently ended up being penalized under the “detection of inconsistencies” criterion. But as one grader pointed out, being able (or willing) to follow instructions is rather vital.

### 10.3.2 On the timing

The answers to the open-ended questions “How long did it take you to produce a first version? To complete the translation?” are presented graphically in Figures 10.1 and 10.2. It is apparent that producing a first draft was easier in Test One, and that more test-takers needed the full allocated time to complete the task in Test Two. Three times as many test-takers in Test One as in Test Two did not use all the time available. Only one person felt very rushed in Test One, and six said they would have liked an additional ten minutes or so, while ten people complained of feeling rushed in Test Two and almost half the test-takers said they would have liked another 10-30 minutes to polish their translations. Considering that all the respondents completed the task, this raises the question of how different their scores would have been if they had had more time. But that is another area for further research.

Figure 10.1. Distribution of time in Test One, by total time spent

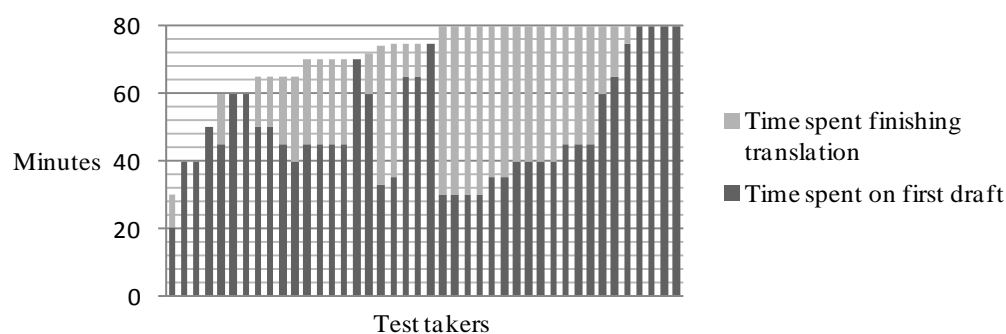
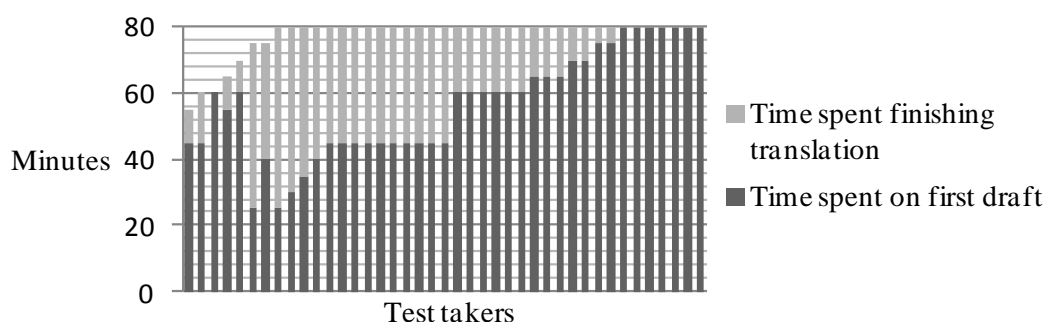


Figure 10.2. Distribution of time in Test Two, by total time spent



### 10.3.3 On their experience as test-takers

In response to the invitation to comment on how fair they thought the tests were and whether they were what they had expected, participants said that both tests seemed fair. Most of the students complained about not being allowed to consult resources in Test One, however, which they felt put them at a disadvantage and undermined the authenticity of the test. The comments below are typical:

The text wasn't too bad and although there were some vocab issues, it was still possible to make reasonable guesses. However, I do think translating without a dictionary/online resources is a bit of an artificial exercise as a translator would never have to do it for a paid job.

I did not feel rushed, but rather felt the lack of external resources and references.

The challenges posed by Test Two were perceived to be different from those encountered in Test One, as these comments show:

I don't think the translation itself was very difficult, but it was harder trying to work out the sentence structure and the meaning behind some of the sentences. If there had been more time to do more research on the topic, that would have helped to untangle some of the cumbersome sentences. I think it probably is fair as that would be the sort of text you may get in [Z], but a little more time would have been helpful.

I think it was fair as the terminology was not overly technical, but the time factor made it difficult. Translation 1 was easier because there was more time to check over the

translation and the text was written more clearly. The sentences in this text were long and packed full of ideas and information that needed to be taken apart and reformulated to make sense in English.

Test 2 was definitely better in that we were allowed to look up vocab and therefore didn't have to guess, so I felt that it was slightly more accurate than Test 1, but more time would have allowed to me to make sure that the English flowed better and the information was correct.

These comments suggest that Test Two achieved some of its aims inasmuch as it seems to have activated the constructs it was supposed to: analytical skills and the abilities to recast sentences, produce a translation that flows well even if the ST does not, and maintain quality under time pressure.

#### **10.4 Comments made by graders on the schemes used**

The graders working with scheme A had mixed opinions of the intuitive method. Three of them felt the scheme was quite straightforward and that the range (0-20) was appropriate. Another, who graded Test Two scripts, commented that “assigning a score to test papers displaying such a range of virtues and vices feels very artificial”. The points scheme came in for various criticisms: it was time-consuming, the categories were not useful and it was hard to distinguish between a serious and a non-serious error. Another grader pointed out another, unforeseen drawback: “The closer a translation comes to a satisfactory rendering, the more prominence the (remaining) blemishes and imperfections take on. I am very conscious of having been much more critical of the good papers than the bad.” This refers to the tendency to pick up minor errors in good work that would be ignored when fixing major errors in poor work. The same phenomenon is seen in revision, where the number of changes made may be influenced by how much red ink (real or virtual) revisers are willing or able to use on a page.

The graders working with scheme B felt some of the scales should be modified slightly and had suggestions for improving them even after the adjustments made in the practice session. One wanted to merge the scales for flow, clarity and idiomaticity, while another wanted to separate grammar from punctuation and spelling, and another wanted to

be able to rate the fact that a translator had not added information (“overtranslation”) and had adhered to instructions. At the same time, as shown in section 10.3.1, failing to do the latter would be penalized, albeit indirectly. All the scheme B users said that the scheme was slightly cumbersome at first but became easier and quicker to apply the more familiar they became with the source text, the definitions and the scale descriptors. They also all commented that it was at times difficult to apply to Test One since evidence for some of the components (e.g. coherence) was difficult to find.

## Chapter 11 Analysis of the results of the test trial<sup>35</sup>

It is probably worth recalling at this point that the objective of the test trial was to determine whether the outcome of profile-adapted assessment processes would be noticeably different from the outcome of traditional ones in the IGO context. The analysis of the results was performed solely with this in mind.

### 11.1 The reliability of the grading schemes

Before comparing performance on the two tests, we needed to have an idea of how reliable the scores were. Ideally, to explore intra-rater reliability, the graders would have re-marked scripts they had already graded, and to explore inter-rater reliability better, there would have been more overlap among graders instead of just two teams and only one grader in team 1 (me). Unfortunately, that was not possible to arrange. The reliability analysis was therefore limited to the examination the inter-rater reliability of the scores obtained through the application of the three grading methods by the two teams to the two tests. The scores obtained by applying scheme B to Test One are shown in Appendix 18; those obtained by applying scheme A to Test Two are shown in Appendix 19.

#### 11.1.1 The reliability of the scores obtained using scheme A (intuitive and points scoring)

The inter-rater reliability of the two methods of scheme A was analyzed by calculating Spearman's rank order correlation coefficients and performing t-tests on the two grading teams' scores. Those calculated for Tests One and Two are presented in Table 11.1. The same combination of graders applied both the points and the intuitive methods to each script.

Table 11.1. Spearman's rank correlation coefficients and p-values for paired t-tests for the intuitive and points scores awarded by grading teams 1 and 2 in Tests One and Two

	Intuitive scoring		Points scoring	
	Test One	Test Two	Test One	Test Two
Spearman's rank correlation ( $\rho$ )	0.640	0.621	0.632	0.550
P-value for paired t-test	0.273	0.320	0.563	<0.05

<sup>35</sup> Special thanks to Dr. Muhammad Atiyat for his advice on the statistical analysis performed in this chapter.



In terms of ranking, all the rank correlations shown in Table 11.1 are statistically significant (the critical value for a sample of 46 is 0.291 for a 95% confidence level). They are not particularly strong, however. In terms of scores, the difference between the scores with the weakest correlation (the two grading teams' points scores in Test Two) is also statistically significant (p-value of the paired t-test<0.05). The probable source of this difference is the fact that graders I and J consistently deducted far fewer points than I did. This is shown in Figure 11.1. The difference between the two grading teams' points scores in Test One was much smaller, as shown in Figure 11.2.

Figure 11.1. Points deducted in Test Two by grading teams 1 and 2\*

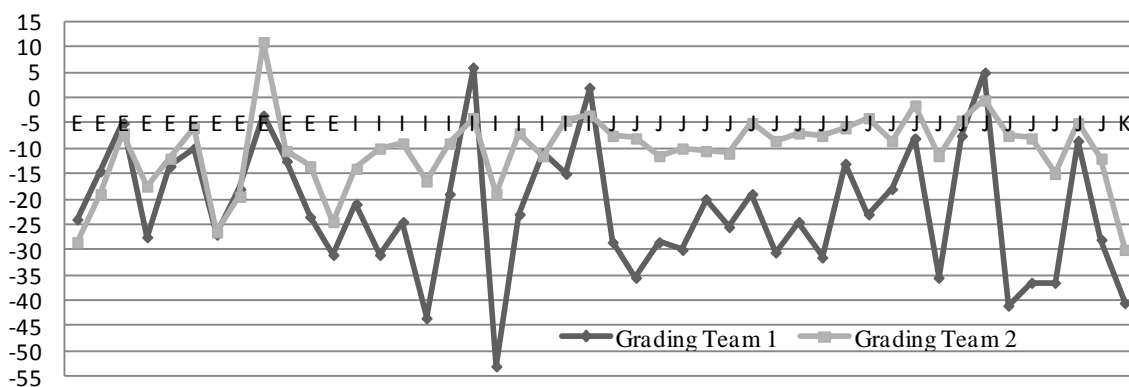
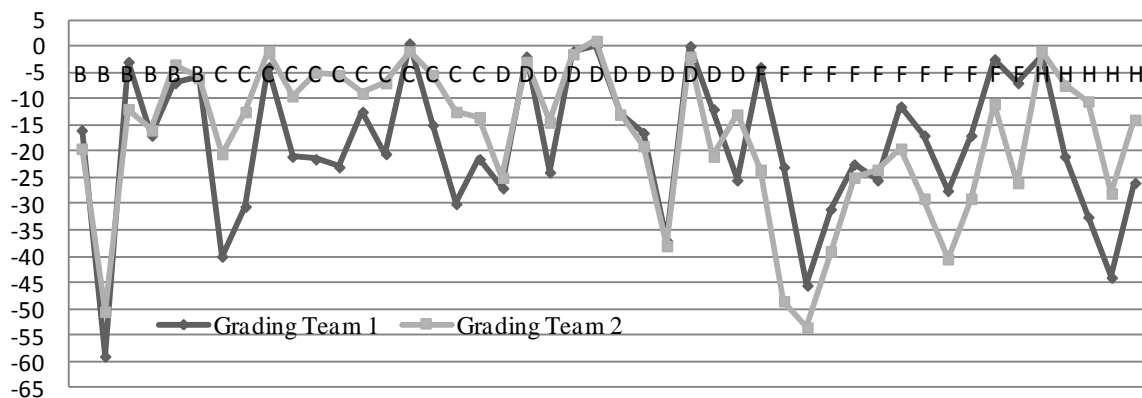


Figure 11.2. Points deducted in Test One by grading teams 1 and 2\*



\*Grading team 2 members identified by letters.

In conversations with Grader I, who worked closely with J, I learned that instead of viewing the intuitive and points methods as separate schemes, they had understood that the points would be deducted from 20, the maximum intuitive score. This considerably affected the amount of points they were willing to deduct. For example, where I deducted 54 points (the lowest score shown on Figure 11.1), Grader I deducted 20, what she thought was as low as she could go. Moreover, not only had Graders I and J awarded and deducted

points from an imaginary starting point of 20, they had applied different criteria when assigning the intuitive score. They had judged the test-takers' performance in terms of what could be expected of recent M.A. graduates or novice translators, not in terms of their readiness for IGO translation work. This could explain why the intuitive scores diverged more from mine than the points balances, as shown by the correlations in Table 11.2, which presents the correlations between the scores awarded by each member of grading team 2 with those awarded by grading team 1 (me). However, Grader E's ranking of the test-takers also diverged more from mine when applying the intuitive scores than when applying the points method. This could reflect a highly subjective or inconsistent approach on my part or the difficulties of applying the traditional scoring methods to the profile-adapted test mentioned by one reviser.

Table 11.2. Spearman's rank correlations ( $\rho$ ) for scores awarded by team 1 and the individual members of team 2 using the intuitive scoring and points scoring methods

Grader	Intuitive scoring		Points scoring	
	Test One	Test Two	Test One	Test Two
B (n=6)	0.986		0.714	
C (n= 12)	0.685		0.726	
D (n=11)	0.955		0.855	
F (n=12)	0.780		0.717	
H (n=5)	0.925		0.900	
E (n=12)		0.668		0.825
I (n=11)		0.714		0.750
J (n= 22)		0.555		0.612
K (n=1)		n.a		n.a

Notes: n= number of scripts graded

Regardless of the source of the variance, all the rank correlations shown in Table 11.2 are also all significant for a 95% confidence level according to the critical values for Spearman's rank correlations. More than half of the correlations for the intuitive scoring of Test One are also above 0.85, which, according to Waddington, is the minimum reliability coefficient recommended for written production in language testing (Waddington 2001b: 31). By that yardstick, the intuitive scores are slightly more reliable than the points balances.

Some conclusions can be drawn about the reliability of the scores obtained using scheme A:

- The rank order correlations (Spearman’s rho) between the two grading teams are all positive.
- The reliability of the intuitive scores and the points scores is higher in Test One than in Test Two.
- On the whole, in Test One, the correlation between the two team’s intuitive scores is greater than between the points scores.
- In Test Two, the correlations between individual members of team 2 and team 1 are higher for the points scores than the intuitive ones, but the large difference in the point balances means that taken as a whole, the correlation between the points-based rankings of team 1 and team 2 is lower than the correlation between the two teams’ intuitive rankings.
- Issuing clear instructions and meeting with graders before they begin to make sure they apply the methods the same way is essential for maximizing the reliability of the scores.

#### *11.1.2 The reliability of the scores obtained using scheme B (multi-trait grading)*

The differences between the component ratings of the two grading teams only differed on two occasions by more than one point on the 1-3 and 1-5 point scales: the rating of TL vocabulary in the case of candidate 39, and the rating of the idiomaticity of the TT in the case of candidate 1, both in Test Two. The rank correlation coefficients for the two grading teams shown in Table 11.3 are all significant and fairly high. For all components, except accuracy, the mean score of grading team 1 was slightly higher than that of team 2, suggesting that I tended to be more generous than grading team 2. The differences between the means are statistically significant in more cases in Test One than in Test Two, suggesting that the scoring of the profile-adapted test was slightly more reliable. The relatively low inter-grading-team reliability of the TL vocabulary scores in Test One suggests that one or both teams had difficulty applying the scale consistently. This could reflect different appreciations of the journalistic style called for in the translation task in Test One or problems with the scale. The slightly lower correlations of the correctness ratings could reflect descriptor wording that was open to subjective interpretations. Graders were free to determine what represented “few” or “several” errors when referring to three different things: grammar, punctuation and spelling.

Table 11.3. Spearman's rank order correlation coefficients and p-values for paired sample t-tests calculated for component scores awarded by grading teams 1 and 2

	TEST ONE			TEST TWO		
	Spearman's rank correlation (rho)	Grading team with the higher mean score	P-value for paired t-test	Spearman's rank correlation (rho)	Grading team with the higher mean score	P-value for paired t-test
Flow	0.787	1	0.002	0.808	1	0.014
Clarity	0.858	1	0.000	0.863	1	0.090
Tone	0.643	1	0.001	0.629	1	0.005
SL knowledge	0.774	1	0.077	0.858	1	0.256
Accuracy	0.882	2	0.660	0.777	1	0.278
TL vocab	0.718	1	0.025	0.827	1	0.417
Idiomacity	0.848	1	0.009	0.848	1	0.088
Coherence	0.847	1	0.040	0.889	1	0.003
Correctness	0.737	1	0.003	0.727	1	0.049
Completeness	0.912	1	0.183	0.843	1	0.183

A look at the correlations with individual team 2 graders again reveals the importance of training graders. Table 11.4 shows the Spearman's rank correlation coefficients for the component scores awarded by me and the individual graders of team 2 in both tests. The size of the scale used to rate the component is indicated in brackets. Assuming a certain level of consistency on my part, Grader G's scores correlated the least with mine, especially in Test Two, whose scripts he graded first. Tellingly, Grader G was the one reviser who did not attend the training session.

Table 11.4. Spearman's rank order correlation coefficients (rho values) for the different components measured by team 1 and the different graders of team 2 using the multi-trait grading scheme (n= number of scripts graded)

Team 1 vs. grader...	Test One			Test Two			
	A (n=16)	B (n=24)	G (n=6)	A (n=12)	B (n=15)	G (n=18)	K (n=1)
Flow (1-5)	0.823	0.742	0.600	0.879	0.8375	0.447	1
Clarity (1-5)	0.797	0.831	0.914	0.837	0.881	0.703	1
SL knowledge (1-5)	0.730	0.787	0.957	0.962	0.829	0.787	1
Accuracy (1-5)	0.929	0.866	0.857	0.909	0.877	0.452	1
Tone (1-3)	0.716	0.694	0.786	0.540	0.621	0.587	1
TL vocab (1-5)	0.781	0.695	0.529	0.857	0.824	0.574	1
Idiomacity (1-5)	0.892	0.755	0.700	0.865	0.908	0.773	1
Coherence (1-5)	0.931	0.813	0.900	0.913	0.772	0.859	1
Correctness (1-5)	0.838	0.701	0.871	0.738	0.713	0.577	1
Completeness (1-3)	1	1	1	0.885	0.759	0.921	1

The low correlations of the scores for tone and register, especially in Test Two, possibly reflect problems applying the scale and the fact that the size of the scale used

affects the degree of correlation likely to be achieved: a difference of one point has a greater effect on a scale of 1-3 than on a scale of 1-5. This is not apparent in the coefficient for completeness because the vast majority of test-takers scored 3 (i.e. they finished the translation). Of the components rated on 1-5 scales, the scores for TL vocabulary in Test One and for correctness (grammar, punctuation and spelling) in Test Two seem to be the ones applied least consistently by all the team 2 graders in relation to team 1, possibly reflecting a problem with the scale as mentioned earlier. Grader K awarded the same ratings to each component as I did, hence the perfect correlations.

Importantly for this study, the inter-grading-team correlations in Test Two for the subjectively-rated A components, clarity and coherence, are acceptably high (0.863 and 0.889, respectively). These components, together with the objectively rated ability to detect inconsistencies, are supposed to be weighted the most heavily in the grading process, and it is therefore important that the corresponding scores are as reliable as possible. The weighting of the components to obtain composite scores is discussed in section 11.2.

### *11.1.3 Conclusions about the reliability of the scoring methods and treatment of the scores*

On both tests, the rank correlations for the component scores awarded by the two grading teams using scheme B were higher than the rank correlations for the intuitive or points scores of scheme A. This confirms the greater reliability of multi-trait scoring, as reported by Cushing Weigle (2002: 72-73).

Of the two methods used in scheme A, in terms of ranking by the two teams as a whole, the points method was the less reliable. In Test One, the inter-rater reliability of the intuitive scores was higher in all cases than for the points scores. Since the purpose was to compare performance on a traditional test and a profile-adapted test and that would necessarily focus on ranking, I decided to base the remainder of the analysis on the more reliable intuitive scores for Test One and not to use the points scores. I only needed one traditional scoring method and I opted to continue on the basis of the one I had found to be most reliably applied in Test One. To enhance the reliability of the intuitive scores further, I called in a third grader to score the scripts in which there was a difference of three or more points between the two grading teams and then to continue the analysis on the basis

of the mean of the three intuitive scores.<sup>36</sup> Alderson et al. recommend bringing in a third grader when scores are “two points or more apart on a five-point scale” (1995: 133). Considering the intuitive scores were on a scale of 1-20, I chose to err on the side of caution. The final mean intuitive scores for Tests One and Two are shown in Appendix 20. The intuitive scoring method discriminates well between candidates on both tests, with the scores on Test One ranging from 32.5% to 97.5% and on Test Two from 20% to 92.5%.

The differences between the component ratings of scheme B only very rarely (twice) differed by more than one point on the 1-3 and 1-5 point scales, which, by the rule of thumb proposed in Alderson et al., was acceptable. The consistent bias towards leniency on my part would be neutralized by averaging the two teams' scores. I therefore decided to proceed with the analysis on the basis of composite scores obtained by combining the mean scores for each component. The calculation of those composite scores is described in section 11.3. The mean component scores for Tests One and Two are presented in Appendix 21.

## 11.2 Calculating composite scores using scheme B

For the purpose of comparison, composite scores based on the component ratings obtained using scheme B were needed. From the outset, the goal had been to weight the A components ( the ability to detect inconsistencies in the ST, write clearly and produce a coherent TT) more heavily than the non-A components. The exploration of possible ways to achieve this consisted of first examining the self-weighting of the test and then considering different weighting options. The analysis focused on Test Two, just as the analysis of the traditional scheme A methods focused on Test One, given that the most important comparison to be made would be between performance on Test One under a traditional grading method and on Test Two using the multi-trait method.

---

<sup>36</sup> There were also practical considerations: ideally I would have called in a third grader to do the same for the points scores and analyzed their reliability again, but the method was very time consuming and no more volunteers were forthcoming.

### 11.2.1 The self-weighting of components in Test Two

The self-weighting of a test refers to the extent to which individual component scores determine the composite score. Components whose ratings vary considerably and that are not correlated highly with the other components will contribute more to the composite score than those that do not discriminate much between candidates (Bachman 1990: 122). Table 11.5 shows the rank correlation coefficients for the mean component scores awarded in Test Two. Table 11.6 shows the variance and the coefficient of variation (variability in relation to the mean) of each component.

The components that generally correlate the least with the others are the abilities to ensure the completeness of the TT and detect inconsistencies (see Table 11.5). The coefficients of variation for these two components (see Table 11.6), however, are relatively low, suggesting they do not discriminate among candidates more than the others. The markedly low variability of the completeness component is explained by the heavy concentration of perfect scores. On the basis of the data in Table 11.5 and Table 11.6, Test Two does not appear to be heavily weighted towards one component or a particular set of components. To increase the discriminatory power of the A components, they would need to be weighted in the calculation of the composite scores.

Table 11.5. Test Two: Spearman's rank correlation coefficients of mean component scores (rho values)

	Flow	Clarity	Tone	SL knowledge	Accuracy	TL vocab	Idiomacity	Coherence	Correctness	Completeness
Flow										
Clarity	0.89									
Tone	0.81	0.76								
SL knowledge	0.77	0.74	0.65							
Accuracy	0.72	0.80	0.61	0.78						
TL vocab	0.83	0.82	0.77	0.64	0.59					
Idiomacity	0.81	0.80	0.65	0.75	0.69	0.75				
Coherence	0.758	0.78	0.63	0.65	0.65	0.77	0.63			
Correctness	0.689	0.77	0.72	0.60	0.70	0.62	0.64	0.58		
Completeness	0.355	0.403	0.325	0.43	0.41	0.39	0.32	0.42	0.40	
Detection	0.505	0.519	0.364	0.53	0.46	0.50	0.62	0.52	0.291	0.31

Table 11.6. Test Two: variance (sigma squared) and coefficient of variation of the component scores, component scales given in brackets

Component	Variance (sigma squared)	Coefficient of variation (CV)
Flow (1-5)	1.267	35.59
Clarity (1-5)	1.138	35.17
SL knowledge (1-5)	0.354	24.21
Accuracy (1-5)	1.363	42.13
Tone (1-3)	1.005	42.30
TL vocabulary (1-5)	1.148	39.92
Idiomacity (1-5)	1.405	43.98
Coherence (1-5)	1.407	48.71
Correctness (1-5)	0.847	32.56
Completeness (1-3)	0.112	11.78
Detection of inconsistencies (1-7)	1.962	31.59

### 11.2.2 Weighting options

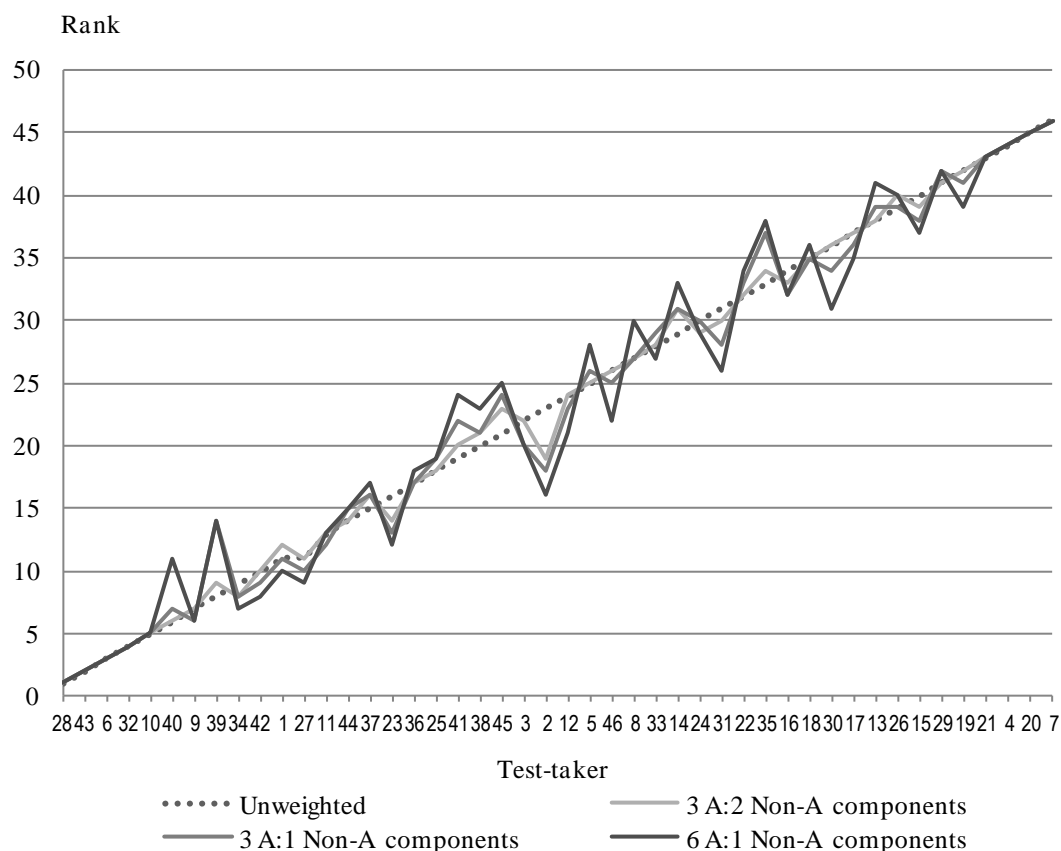
How to weight priority components is ultimately a subjective decision. One option is to use a simple formula, such as those listed below.

1. Weighting the A components 3:2 to the other components, which would mean that the A components would together account for 36% of the final score;
2. Weighting the A components 3:1 to the other components, which would mean that the A components would together account for 53% of the final score;
3. Weighting the A components 6:1 to the other components, which would mean that the A components would together account for 69% of the final score.

The impact of these weightings on rankings are shown in Figure 11.3 and their impact on scores and ranking are shown in Appendix 22. Figure 11.3 shows that only among those test-takers who ranked highest and lowest does the weighting of the A components have no impact on ranking. These are test-takers 28, 43, 6, 32, 10, 21, 4, 20 and 7. This is because they either scored consistently well or consistently badly on all the components. The weightings affected the rankings of all the test-takers in between. In other words hiring on the basis of unweighted scores would not affect who gets selected if the organization only hires the top four candidates. If the organization needs to fill more than four vacancies, however, the weighting scheme could make all the difference to who gets hired.



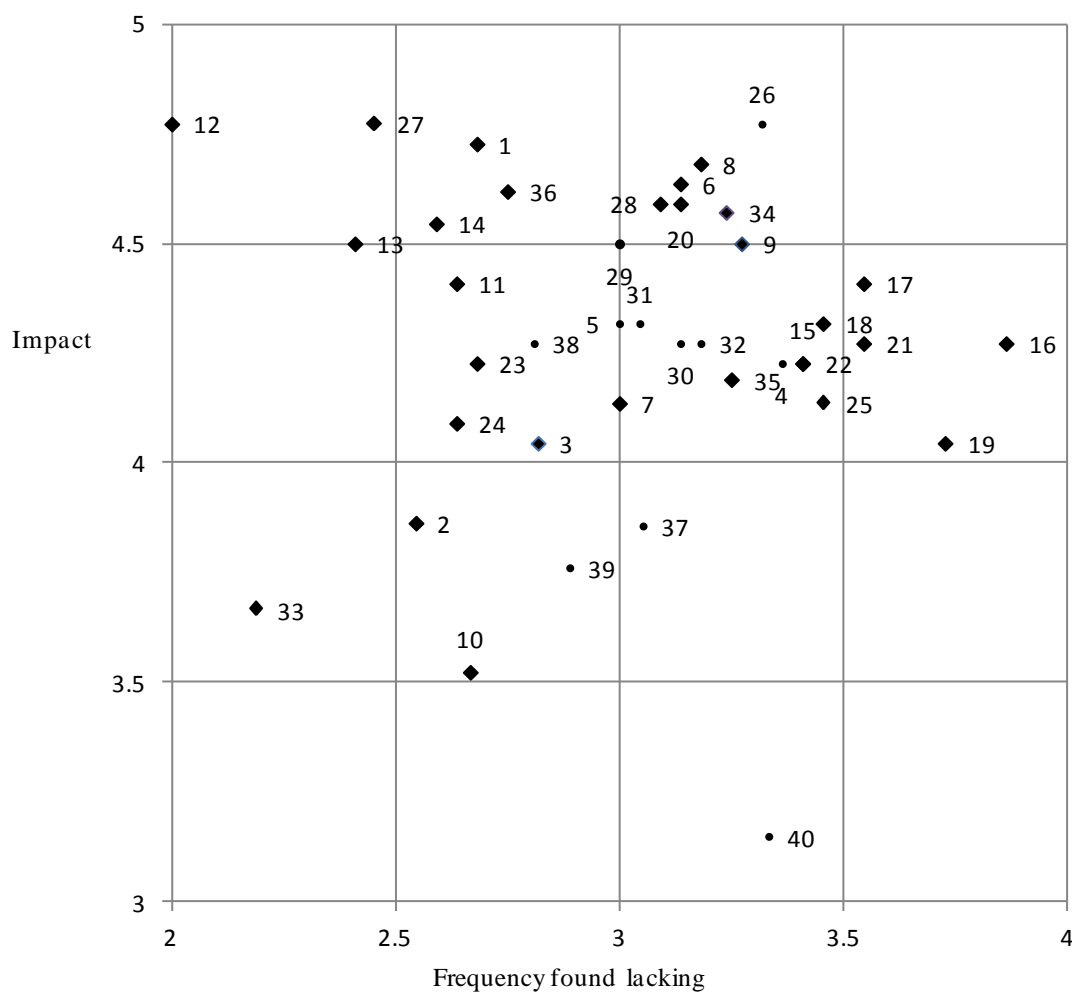
Figure 11.3. Rankings on Test Two, by different weighting of A and non-A components



The problem of calculating composite scores this way is that the weighting is ultimately rather arbitrary. According to Alderson et al. “the best advice is not to weight components by some formulaic adjustment of scores, but rather to weight components at the test design stage by including more or fewer items in the individual components” (1995: 150). Test Two was designed to activate the category A skills in particular, but the analysis of the self-weighting of the test shows that they are not naturally self-weighted components.

Our profile-adapted approach offers a less arbitrary means of determining weightings. The composite scores can be calculated by weighting the component scores on the basis of their relative position within the empirically identified skills-knowledge profile, in other words, on the basis of their frequency and impact ratings. This is demonstrated below. Figure 9.1, which depicts the scatter graph for organization Z, is reproduced here for convenience.

Figure 9.1. Impact and frequency ratings for organization Z



Note: For key to skills and knowledge types, see Figure 8.1.

The weightings of each component according to the skills-knowledge profile for organization Z are shown in Table 11.7. These were calculated by applying the following logic:

- To weight the component scores (s) in terms of the frequency with which the corresponding components are lacking, they should be multiplied by their frequency rating (f).
- To weight the component scores (s) in terms of the impact of the components, they should be multiplied by their impact rating (i).
- To weight them by both f and i, they should be multiplied by both f and i.

The weighted composite score, as a percentage, (*wcs*) is then calculated as follows:

$$wcs = \frac{100}{m} \sum_{j=1}^c \bar{x}_j f_j i_j a_j$$

where:

- c* = number of components
- m* = maximum possible score
- $\bar{x}_j$  = mean score for component *j*
- f<sub>j</sub>* = frequency for component *j*
- i<sub>j</sub>* = impact for component *j*
- a<sub>j</sub>* = adjustment for component *j*

and

$$a_j = \frac{dm_j}{am_j}$$

where:

- dm<sub>j</sub>* = desired maximum possible score
- am<sub>j</sub>* = actual maximum possible score

Table 11.7. Empirical weighting of profile components on the basis of impact and frequency ratings

Skill/knowledge	f	i	fxi	
16	Produce translations that flow smoothly	3.9	4.3	16.51
26	Adhere to in-house style conventions	3.3	4.8	15.84
17	Capture nuances of ST	3.5	4.4	15.63
21	Convey the intended effect of the ST	3.5	4.3	15.15
19	Write elegantly regardless of the ST	3.7	4.0	15.08
18	Recast sentences in the TL	3.5	4.3	14.92
8	Work out the meaning of obscure passages	3.2	4.7	14.90
34	Maintain quality even under time pressure	3.2	4.6	14.80
9	Detect inconsistencies, contradictions, etc.	3.3	4.5	14.73
6	Understand complex topics	3.1	4.6	14.54
15	Produce idiomatic translations	3.4	4.2	14.41
22	Achieve the right tone and register	3.4	4.2	14.41
20	Convey the ST message clearly	3.1	4.6	14.40
25	Tailor language to the readers' needs	3.5	4.1	14.29
4	Subject knowledge	3.4	4.2	14.22
28	Ensure the coherence of the TT	3.1	4.6	14.19
35	Explain translation decisions and problems	3.3	4.2	13.62
32	Judge the reliability of information sources	3.2	4.3	13.60
29	Track down sources to check facts	3.0	4.5	13.50
30	Track down sources to understand the topic	3.1	4.3	13.40
31	Mine reference material for phrasing	3.0	4.3	13.15
5	Knowledge of the organization	3.0	4.3	12.95
36	Follow complicated instructions	2.8	4.6	12.70
1	Knowledge of SL	2.7	4.7	12.68
7	Master new subjects quickly	3.0	4.1	12.41
38	Work with electronic terminology tools	2.8	4.3	12.00
14	Knowledge of TL punctuation rules	2.6	4.5	11.78
37	Work with translation memory software	3.1	3.9	11.77
27	Ensure the completeness of the TT	2.5	4.8	11.71
11	An extensive TL vocabulary	2.6	4.4	11.62
3	Knowledge of SL culture(s)	2.8	4.0	11.40
23	Knowledge of TL varieties	2.7	4.2	11.34
39	Handle more than basic Word functions	2.9	3.8	10.87
13	Knowledge of TL grammar	2.4	4.5	10.84
24	Knowledge of TL culture(s)	2.6	4.1	10.79
40	Work with Excel and/or PowerPoint	3.3	3.2	10.50
2	Knowledge of SL varieties	2.5	3.9	9.83
12	Knowledge of TL spelling rules	2.0	4.8	9.55
10	Detect mathematical errors in the ST	2.7	3.5	9.40
33	Type accurately and fast	2.2	3.7	8.02

The hierarchy obtained through the impact-frequency categorization (A, B, C, D components) used in the design stage (see section 9.2.1) was slightly different from the hierarchy obtained by multiplying  $f$  and  $i$  shown in Table 11.7. This is because the former was based on the use of different cut-off marks to distinguish between high and low: 3 in the case of frequency (i.e., at least sometimes) and 4.5 in the case of impact (which split the group of items approximately in half). The weightings, on the other hand, are based on the premise that the impact and frequency ratings are equivalent, i.e. that a 3 in frequency

is equivalent to a 3 in impact. The categorization for test-design purposes was not performed on that basis because we would have ended up with four categories of high-impact components (very-high-frequency, high-frequency, low-frequency and very-low-frequency). We wanted to spread the influence of impact and frequency more evenly. In the  $f_{xi}$  weightings, however, the greater spread of the frequency ratings along the length of the scale results in them having greater discriminatory power than the impact ratings. By assuming the scales are equivalent in calculating the weightings, the frequency ratings assume a greater influence over the relative position of items in the hierarchy. Multiplying  $f$  by  $i$  pushes relatively low-impact, but high-frequency components, such as flow (16) and capturing nuances (17), for example, to relatively higher positions than those they occupied in the categorization scheme (category C), and relegates high-impact, relatively low-frequency ones, such as ensuring coherence (28), to relatively lower positions (it was in category A). To apply the same logic to the weightings as applied in the categorization, the weightings formula would have to be adjusted to account for the different cut-off points used. The shifts are not that dramatic, however, and for the purposes of this demonstration we will presume that the organization is bent on improving its productivity (which in fact it is) and is happy to award slighter greater importance to the scarcer components, particularly flow.

The weighting factors that apply to the components actually assessed in Test Two are listed in Table 11.8.

Table 11.8. Profile-adapted weighting factors applied in the multi-trait grading of Test Two

Component (using the definitions used by graders)	Weighting factor (f.i)
Smoothness of flow	16.5
Evidence of ability to transfer detailed levels of meaning (accuracy)	15.6
Inconsistencies and logic problems	14.7
Coherence (internal logic)*	14.6
Clarity	14.4
Idiomatic language use	14.4
Tone and register	14.4
Evidence of knowledge of source language	12.7
Completeness	11.7
Evidence of broad active target-language vocabulary	11.6
Correctness (grammar, punctuation and spelling)	10.7**

\* Average of the individual weightings for working out obscure meanings and ensuring internal consistency as explained in section 9.1

\*\*Average of the individual weighting factors for grammar, punctuation and spelling was used.

By applying the formula above we obtain the weighted composite scores presented in Appendix 23. Of course, the problem with this profile-adapted weighting method is that not only does it rest on the assumption that frequency is as important as impact, as noted above, but it assumes that each component is a discrete unit. One of the problems in language and translation testing is how to distinguish sufficiently between what are often inter-related abilities (Bachman 1990: 32). Ultimately the decision about how to weight the components is up to the test user.

Interestingly, the profile-adapted weighting produces an almost identical set of scores as the 3:1 weighting, which are also shown in Appendix 23 for comparison. The mean scores in Test Two under the 3:1 weighting scheme ranged from 27.0% to 95.1%. The mean scores under the profile-adapted weighting ranged from 28.3% to 95.0%. The correlation between the rankings obtained from the two sets of scores is almost perfect ( $r_s=0.99$ ). The difference between the means is statistically significant (p-value for paired t-test =0.001), but only 1.2 points on a scale of 1-100, and hence of no practical difference. For convenience, it was therefore decided to continue the analysis on the basis of composite scores obtained using the straightforward 3:1 weighting scheme. The composite scores for Test One obtained by applying the 3:1 weighting scheme to the mean scores for each component are presented in Appendix 24. It should be noted that scheme B discriminates well among the candidates on both tests, with composite scores ranging from 31.8% to 98.4% on Test One and 27% to 95.1% on Test Two.

### **11.3 Comparison of performance on the two tests**

To compare performance on the mock traditional test (Test One) and the profile-adapted test (Test Two), we need to compare the rankings obtained by applying a traditional grading method (in this case, intuitive scoring) to Test One and by applying the multi-trait grading scheme to Test Two. The rank correlation coefficient for that combination is 0.602, which is still statistically significant, according to the Spearman's table, but certainly not high. That suggests that the type of test used has a considerable impact on the outcome of the testing process. The implications this has for test-takers and test users is discussed in section 11.3.2.

### 11.3.1 *The relative influence of the test task and the grading scheme used*

The analysis of the correlations between scores and rankings obtained by the application of the intuitive and the multi-trait scoring methods to the same test (see Table 11.9) and the application of the same methods to both tests (see Table 11.10) suggests that the test task had a greater influence on ranking than the grading scheme used.<sup>37</sup> This is because the correlations in Table 11.9 are higher than the correlations in Table 11.10. In other words, the rankings changed more between Test One and Test Two under the same grading scheme than when different grading schemes were applied to the same test. Of the grading schemes, the multi-trait one seems to have had a greater impact on ranking in Test Two. This is possibly due to the influence of the detection of inconsistencies component which is not measured in Test One but is heavily weighted in the scheme when it is applied to Test Two. The correlations in Table 11.10 are low, which suggests that indeed the two tests measure different things. Even if the difference between the means is not statistically significant when the multi-trait scoring method is applied, the correlation between the rankings is notably low.

Table 11.9. Spearman's rank correlations between scores obtained using intuitive and multi-trait scoring in Test One and Test Two

Correlations between rankings	Test One	Test Two
Intuitive vs. multi-trait scoring	0.82	0.67

Table 11.10. Spearman's rank correlations (rho) and p-values for paired sample t-tests for the intuitive and multi-trait scores on Test One and Test Two

	Intuitive scores Test One vs. Test Two	MT scores Test One vs. Test Two
Spearman's coefficient	0.609	0.502
P-value for paired t-test	0.001	0.087

### 11.3.2 *The implications for test-takers and test users*

How the differences in scores and ranking affect candidate selection depends on the purpose of the test and whether a norm- or criterion-referenced model is applied. IGOs select candidates in two ways: they either place all the test-takers who achieve above a

<sup>37</sup> The composite scores obtained by weighting the mean component scores awarded under the multi-trait scheme to Test One are shown in Appendix 24.

certain score on a roster for future employment (either to perform freelance work, fill temporary vacancies or assume permanent posts) or they select only as many candidates as there are vacancies that need filling. We therefore need to examine how the mock traditional and the profile-adapted tests discriminate in terms of both scores (pass marks) and rankings. The comparisons between performance on the two tests in this chapter are based on the intuitive scores for Test One and the composite multi-trait scores for Test Two.

In the intuitive scheme of Test One, the pass mark was pre-determined; in the multi-trait scheme of Test Two, it can be determined in different ways. One way is to decide the cut-off point on the basis of the pass/fail indications given by the graders just after they rate each component. These coincided at around 70% on both tests. These may be holistic ratings, but they are not wholly intuitive in the way they are in scheme A because they are decided upon after completing the multi-trait scoring, not after a simple read-through of the text. Alternatively a panel of experts could be set up to examine the borderline cases identified by the grading teams and establish a pass mark that way. In both cases, the pass mark could be combined with filters to isolate candidates who meet certain criteria or minimum standards in particular areas. Selecting all test-takers who scored over 70% and scored at least 4 on flow, for example, would in theory isolate the candidates most likely to help improve the productivity of the service, since the ability to produce translations that flow was identified as an oft-lacking skill among new recruits. This is one of the advantages of multi-trait scoring.

The final scores and pass marks on both tests are shown in Figure 11.4 and the final ranks on both tests are shown in Figure 11.5. In terms of pass rates, the two tests differ considerably. Just over half (55%) of the test-takers passed Test One, and only 30% passed Test Two (based on the 70% pass mark), suggesting that the profile-adapted test was more difficult than the mock traditional one, as reported by most of the student test-takers.

Moreover, it is apparent from the scores shown in Figure 11.4 that the group of test-takers who passed Test One is not the same as the group who passed Test Two. Nine participants passed both tests (test-takers 44, 23, 34, 40, 39, 11, 42, 1, 6, 32, 27, 10, 28, 43), while eleven passed Test One but not Test Two (test-takers 16, 26, 8, 38, 25, 22, 33, 18, 12, 29, 37), and only one passed Test Two but not Test One (test-taker 9). If the organization was filling a roster on the basis of selecting all those who passed Test One, it would take on many more translators who, according to Test Two results, had not displayed the right combination of skills and knowledge, and if it filled the roster on the

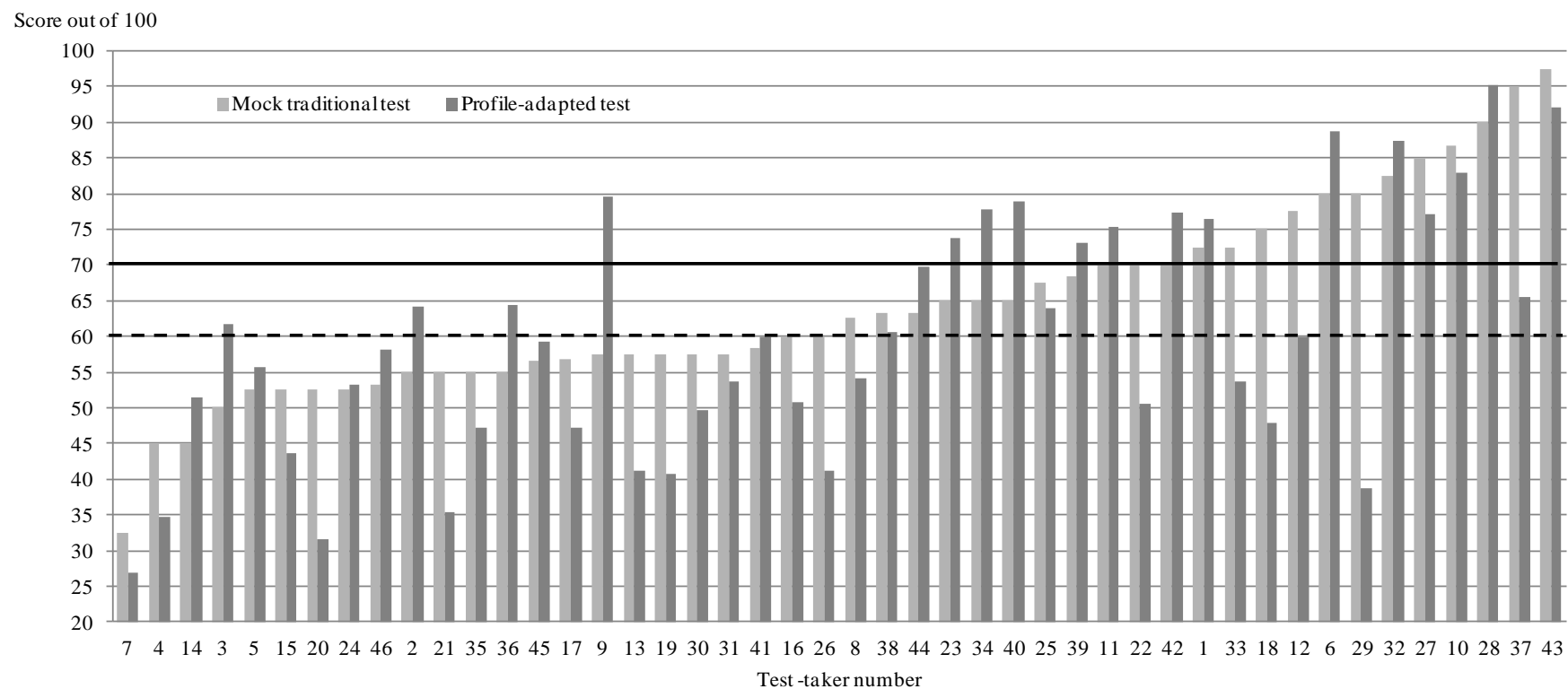


basis of those who scored well on Test One (say over 75%), it would possibly be missing out on candidates who did match the profile.

The difference in the outcome of the two tests is even clearer when we examine the rankings. Figure 11.5 shows that the relative performance of test-takers within the study group varied considerably from one test to the other. In 24 cases, the difference in ranking is over 5 places, in 19 cases it is over 10, and in 9 it is over 15 places. If the organization had ten vacancies to fill, on the basis of the Test One results, it would select, in order, candidates 43, 37, 28, 10, 27, 32, 29, 6, 12 and 18; on the basis of the Test Two scores, it would select 28, 43, 6, 32, 10, 9, 40, 34, 42 and 1. Only five of the test-takers fall into both groups. So Test One clearly ranks candidates differently than Test Two. This has major implications for recruitment testing. At organization Z, intuitively scored, domain-irrelevant tests do not rank candidates the same way as profile-adapted tests, and recruitment decisions based on the results of one or the other would result in different translators being hired. This is a key finding of our research.

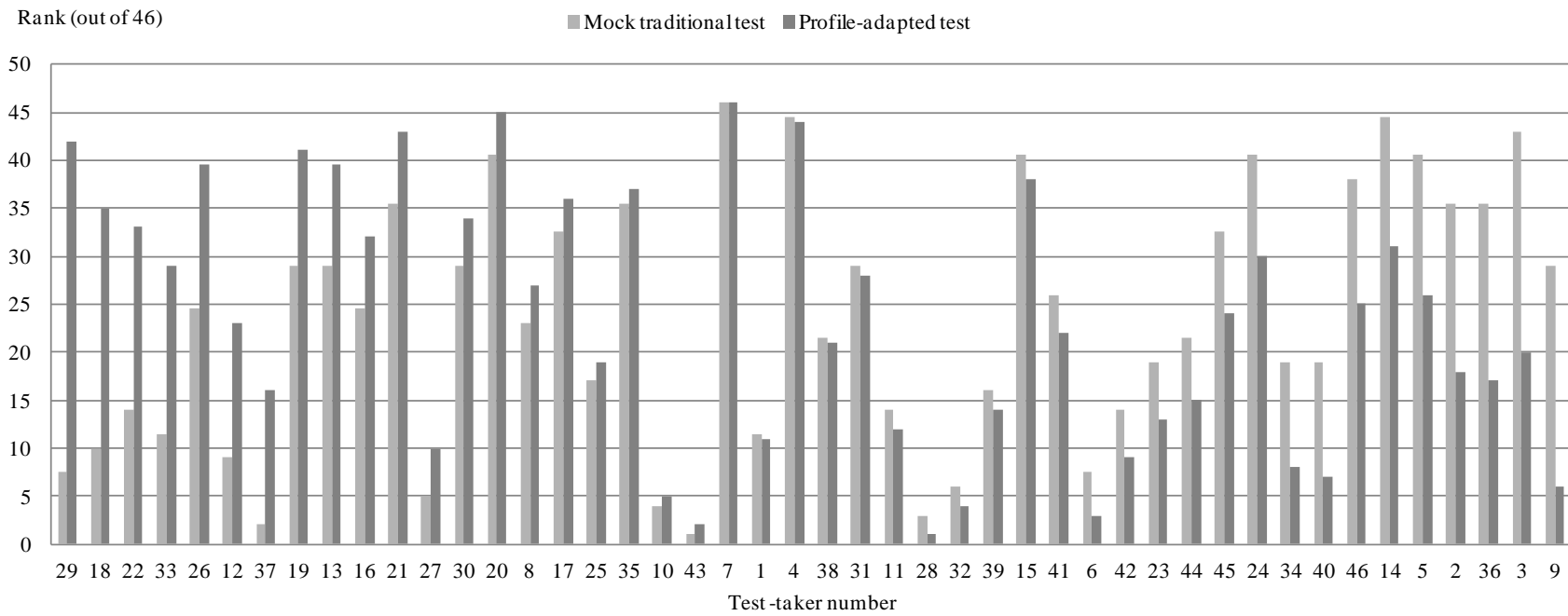
The influence of different aspects of the tests and test-taker variables on the test results are explored in the next section.

Figure 11.4. Scores on the mock traditional text (Test One) vs. scores on the profile-adapted test (Test Two)



The pass mark for the profile-adapted test is 70%  
The pass mark for the mock traditional test is 60%

Figure 11.5. Ranks on the mock traditional text (Test One) vs. ranks on the profile-adapted test (Test Two)



## 11.4 The influence of test-taker variables

The analysis of the influence of test-taker variables was limited by the sizes of some of the sample subgroups, but some inferences could be made. As in the previous section, the comparisons are based on the intuitive scores and a 60% pass mark for Test One and the composite multi-trait scores and a 70% pass mark for Test Two.

### 11.4.1 The training institution

Did students at one institution perform notably better or worse than students at the other institutions? Table 11.11 shows how scores and rankings varied from one institution to another. The results for the IGO translators are included as a point of reference. Mean and median scores and ranks are presented. The mean is of greater interest in the case of scores and the median in the case of ranks. One population (Uni-3) was much smaller than the other. It comprised four test-takers compared with 7-10 in the other groups. This should be taken into account, especially when comparing the ranges.

Table 11.11. Test One and Test Two: scores and ranks, by institution

	Ranks			Scores		
	Mean	Median	Range	Mean	Median	Range
Test One						
Uni-1	25.9	26.0	5-43	62.4	60.0	50-85
Uni-2	22.8	23.0	7.5-40.5	63.5	61.7	52.5-80
Uni-3	33.1	34.0	29-35.5	56.0	55.8	55-57.5
Uni-4	38.2	40.5	19-46	50.0	52.5	32.5-65
Uni-5	19.2	17.0	4-38	67.0	67.5	53.3-86.7
IGO	8.3	6.8	1-19	81.3	81.3	65-97.5
Test Two						
Uni-1	25.4	23.5	10-45	56.9	57.9	31.5-77.1
Uni-2	26.3	27.5	6-42	56.6	54.7	41.3-73.7
Uni-3	29.6	30.5	18-39.5	53.0	53.2	41.3-64.2
Uni-4	33.3	36	8-46	47.3	47.3	27-77.8
Uni-5	20.8	22	5-41	60.0	60	40.8-82.8
IGO	9.8	5.5	1-33	83.2	83.2	50.5-95.1

Excluding the IGO translators, the students at Uni-5 as a group outperformed the students at the other universities on both tests, while the students at Uni-4 overall fared worse than the others. The differences are not huge, however, and the means and medians conceal considerable variations within the same group, as shown by the ranges, with an outstanding performance by one student from Uni-4 and a very weak performance by a Uni-5 student being recorded in Test Two, for example. Meanwhile, none of the Uni-3 students seem to have passed either test. One reason for such variations could be the stage of training the students have reached and whether Spanish was their L2 language. The Uni-3 students, for

example, were first-years for whom Spanish was their second foreign language (as is the case for most of the Uni-4 students), while the Uni-5 students were second years whose first foreign language was mostly Spanish. The analysis of the influence of such factors is beyond the scope of this project but could be of interest to both the universities and IGOs.

#### 11.4.2 Age

The means, medians and ranges of the scores and ranks obtained by different age groups in the two tests are given in Table 11.12.

Table 11.12. Test One and Test Two: mean, median and ranges of the scores and ranks obtained, by age group

	Under 26 (N=24)	Over 26 (N=22)	Under 30 (N=33)	Over 30 (N=13)	Under 40 (N=40)	Over 40 (N=6)
<b>Test One scores</b>						
Mean	58.9	70.3	61.3	72.1	61.4	84.2
Median	57.1	67.5	57.5	70.0	57.9	81.4
Range	32.5 to 86.7	52.5 to 97.5	32.5 to 86.7	52.5 to 97.5	32.5 to 86.7	70 to 97.5
<b>Test One ranks</b>						
Mean	29.1	17.4	26.0	17.2	26.0	7.8
Median	30.8	16.5	29.0	14	27.5	6
Range	4 to 46	1 to 40.5	4 to 46	1 to 40.5	4 to 46	1 to 14
<b>Test Two scores</b>						
Mean	58.9	64.0	59.1	63.2	57.7	77.7
Median	30.0	21.0	24.0	16.0	25.5	8.0
Range	5 to 46	1 to 45	3 to 46	1 to 45	3 to 46	1 to 33
<b>Test Two ranks</b>						
Mean	25.8	21.0	24.0	22.2	25.3	11.3
Median	25.5	18.5	24.0	16.0	25.5	8.0
Range	5 to 46	1 to 45	3 to 46	1 to 45	3 to 46	1 to 33

The data in Table 11.12 show that, in general, performance improves with age since, in both tests, the median of the ranking falls and the mean score increases as the cut-off point between the younger and older group is shifted towards the latter. A two-sample t-test for the younger and older halves of the sample (under 26 and over 26) suggests the difference between the two groups' scores are significantly different only in Test One ( $p=0.004$ ), but not in Test Two ( $p=0.164$ ). The range shown in the figures and the tables moreover reveals that several younger test-takers performed as well as their older counterparts, suggesting it would be unwise to exclude them from the candidate pool.

#### 11.4.3 Sex

With only 9 men among the 46 test-takers, making meaningful comparisons is difficult, but the figures in Table 11.13 suggest that sex had no major influence on performance as there are

no large differences between the median ranks or mean scores for men and women, on either test. A two-sample t-test assuming equal variance (based on results of Levene's test) shows that the difference between the sexes is not statistically significant ( $p=0.32$  in Test One and  $0.464$  in the case of Test Two).

Table 11.13. Test One and Test Two: scores and ranks by sex

	Mean	Median	Range
<b>Scores on Test One</b>			
Men	68.5	60	52.5 to 90
Women	63.3	62.5	32.5 to 97.5
<b>Ranks on Test One</b>			
Men	19.4	24.5	3 to 39.5
Women	24.5	23	1 to 46
<b>Scores on Test Two</b>			
Men	64.1	53.3	40.8 to 95.1
Women	59.4	59.9	27 to 92.2
<b>Ranks on Test Two</b>			
Men	22.1	30	1 to 41
Women	23.9	23	2 to 46

#### 11.4.4 Academic background

Table 11.14. Test One and Test Two: scores, ranks and age group of M.A. holders

Test-taker	Test One rank	Test Two rank	Test One score	Test Two score	Age group
5	33	24	56.7	59.3	21-25
18	10	35	75	47.8	21-25
25	17	19	67.5	64.0	21-25
6	7	3	80	88.7	26-30
23	18	13	65	73.7	31-35
29	7	42	80	38.7	31-35
37	2	16	95	65.5	41-45
22	13	33	70	50.5	46-50
32	6	4	82.5	87.5	51-55
43	1	2	97.5	92.2	56-60

Table 11.14 shows the performance of the M.A. holders on the two tests. The four oldest test-takers (37, 22, 32 and 43) who had M.A.s were all IGO translators, as was test-taker 6. Having an M.A. does not seem to be a major advantage given how poorly 45, 25 and 23 did on Test One and the very low rankings of 45, 18, 29 and 22 on Test Two. Test-taker 22 also had the supposed advantage of age (and hence experience), yet came 33<sup>rd</sup> on Test Two. That said, a t-test does show a statistically significant difference between the two group's scores on Test One ( $p=0.001$  for a 95% confidence level), albeit not in Test Two ( $p = 0.12$ ).

### 11.4.5 Professional experience

IGO experience seems to have been a factor of success in Test One and Test Two, as shown in Table 11.15, with the mean scores rising and the median rankings falling with IGO experience in both tests. Other types of professional experience do not seem to have influenced performance much. Dividing the sample into those with and without each type of experience shows that IGO experience was a greater factor than other experience on both tests. Performing two-sample t-tests on the two subgroups shows that the differences in the scores on both tests were statistically significant ( $p < 0.05$ ) in the case of IGO experience, but not in the case of other experience ( $p = 0.37$  in Test One and  $p = 0.31$  in Test Two). A comparison cannot be made between those with only IGO experience and those with only other types of experience because the sample sizes of the subgroups are too small.

Table 11.15. Test One and Test Two: mean, median and range of scores and ranks, by IGO experience and non-IGO experience

	Test One scores			Test One ranks		
	Mean	Median	Range	Mean	Median	Range
<b>IGO experience</b>						
None (N=31)	58.8	57.5	32.5-80	28.8	29.0	7-46
Up to 2 years (N=9)	72.8	72.5	55-86.7	14.4	11.5	4-35.5
3-30 years (N=6)	80.4	80.0	65-97.5	9.7	8.5	1-19
<b>Other experience</b>						
None (N=19)	62.2	57.5	32.5-85	24.5	29	5-46
Up to 2 years (N=24)	65.9	62.9	45-97.5	22.92	22.25	1-44.5
3-30 years (N=3)	65	70	52.5-72.5	22	14	11.5-40.5
	Test Two scores			Test Two ranks		
	Mean	Median	Range	Mean	Median	Range
<b>IGO experience</b>						
None (N=31)	53.5	53.8	27-79.7	28.8	29	6-46
Up to 2 years (N=9)	70.9	76.5	41.3-88.7	15.4	11	3-39.5
3-30 years (N=6)	80.1	77.1	65.5-95.1	8.5	9.5	1-16
<b>Other experience</b>						
None (N=19)	57.2	53.8	27-88.7	25.9	28	3-46
Up to 2 years (N=24)	63.1	62.9	31.5-95.1	21.2	19	1-45
3-30 years (N=3)	56.9	50.5	43.7-76.5	27.3	33	11-38

In terms of pass rates, the different weight of IGO and non-IGO experience is also apparent. Moreover IGO experience seems to have been a greater influence in Test Two than in Test One. Figures 11.6 and 11.7 show how test scores varied by years of IGO translation experience in Test One and Test Two, respectively. The black line marks the definitive pass mark in Test One in Figure 11.6 and the possible pass mark in Test Two in Figure 11.7.

In Test One, all test-takers with over six months' experience in IGO translation passed, and 14/15 of those with any IGO experience whatsoever passed. Just over one third (11/31) of those with no IGO experience passed. Test Two seems to have posed more of a

challenge to those with no IGO experience: only 4 out of the 31 test-takers in that category passed, while 11 out of 15 of those with at least some experience were successful.

Figure 11.6. Scores on Test One by IGO translation experience

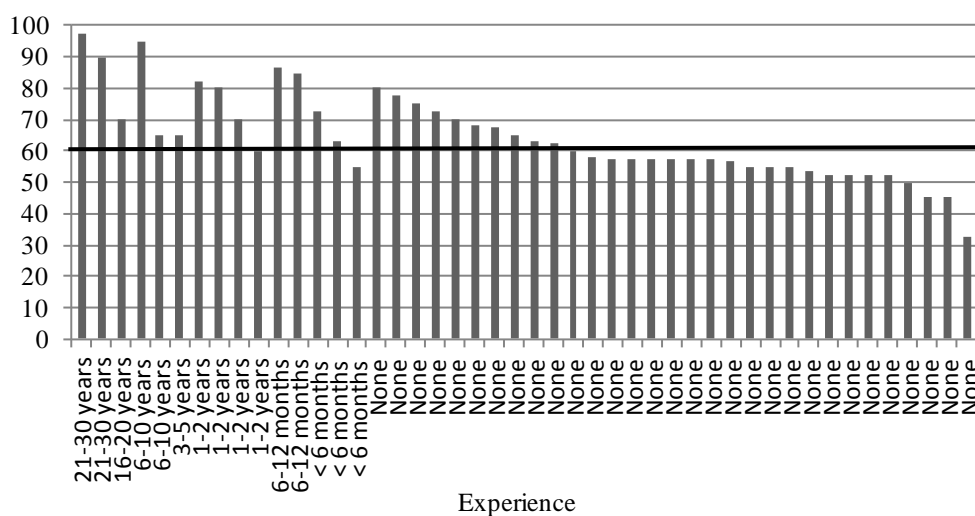
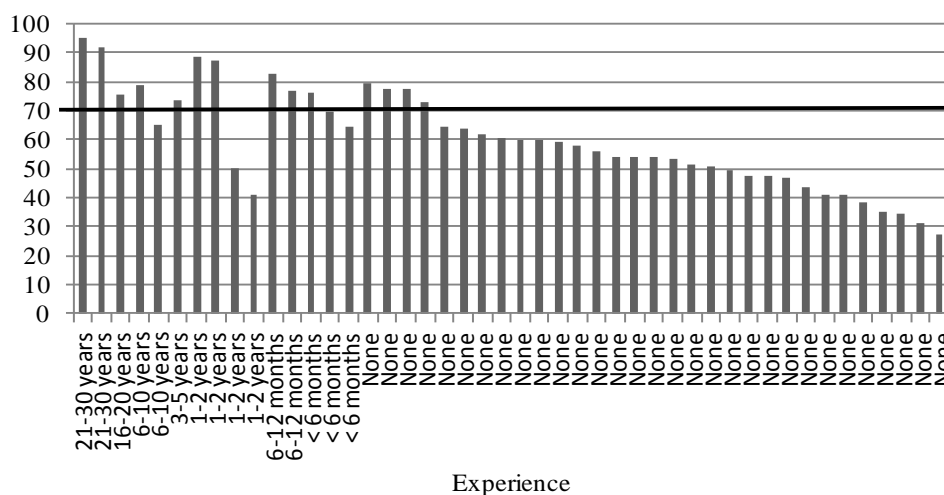


Figure 11.7. Scores on Test Two by IGO translation experience



The data presented in Figures 11.8 and 11.9 on scores in Test One and Test Two by non-IGO experience need to be analyzed in the light of the information on other experience. In Test One (Figure 11.8), for example, the top two scorers among those with less than six months experience actually had 6-10 and 21-30 years' IGO experience, respectively, and three of the five participants with 1-2 years' experience who passed Test One, including the top two, belong to the IGO group. One of those who passed with no non-IGO experience had actually been working in IGOs for 16-20 years. In Test Two (Figure 11.9), the person who scored the highest has less than 6 months' non-IGO experience, but in fact works for an IGO, as do all four of those who passed with 1-2 years non-IGO experience and two of those who passed without any non-IGO experience.



Figure 11.8. Scores on Test One by translation experience outside IGOs

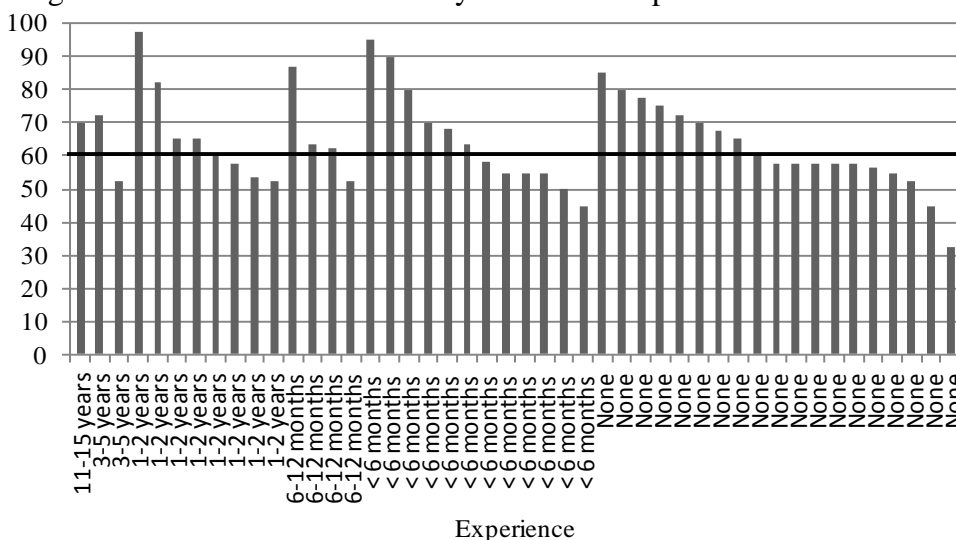
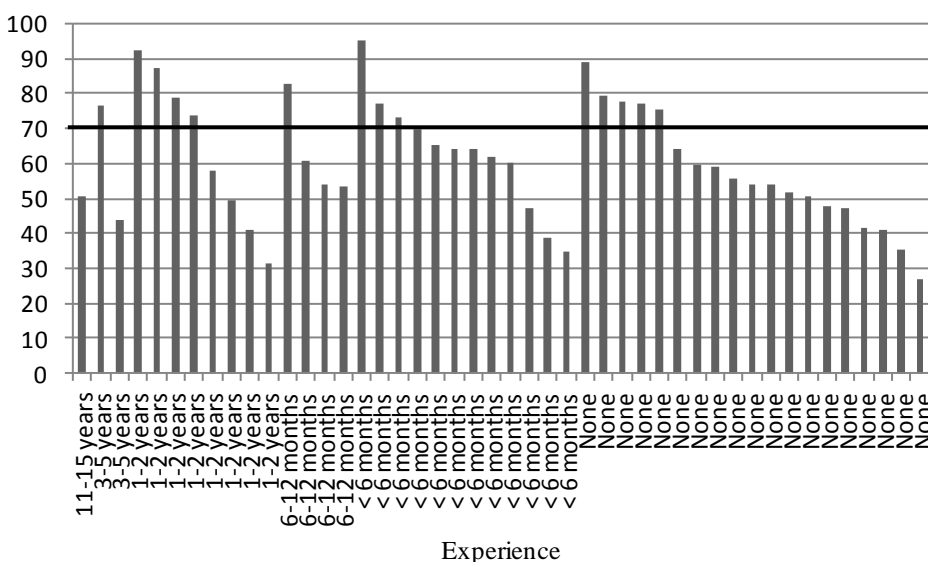


Figure 11.9. Scores on Test Two by translation experience outside IGOs



The changes in the pass rates according to type of experience is summarized in Table 11.16.

Table 11.16. Test One and Test Two: pass rates by type of experience.

Experience	Pass rate Test One	Pass rate Test Two	% change in pass rate between Tests One and Two
No IGO experience	35%	4%	-88%
Some IGO experience	93%	73%	-22%
No other experience	42%	21%	-50%
Some other experience	59%	33%	-44%

The pass rates fall between Test One and Test Two in all cases, but noticeably less in the case of those with at least some IGO experience, suggesting that the profile-adapted test is

possibly better than a traditional test at identifying translators with the skills-knowledge set required for IGO work.

#### 11.4.6 Self-assessment and ranking by teachers

How good were participants at assessing in which test they had fared better? A comparison of how participants rated their performance and whether graders thought they should pass or fail is presented in Table 11.17. Test-takers 9, 29, 8, 25, 12, 37 and 22 realized which test they had performed better on. Candidate 16 was the only one to get it the wrong way around. Such distinctions cannot be verified in all cases as test-takers often rated their performance as the same, then passed or failed both tests as well. Of course, there may be some instances of false modesty, but there is also a certain lack of awareness among some candidates of how poorly they performed. Participants 5, 30, 3, 20, 36, 41, 46 and 14 (20% of test-takers) rated their performances on both tests as satisfactory when they had failed both. The IGO translators seemed to have the greatest self-awareness. In short, self-assessment does not seem a good predictor of performance or useful for criterion validation. It does, however, raise questions about university students' awareness of what is expected of IGO translators.

Table 11.17. Scores and rankings on Test One and Test Two and self-assessment of performance\*

Test-taker	Test One rank	Test Two rank	Test One score	Test Two score	Test One Pass/Fail	Test Two Pass/Fail	Test One Self-assessment	Test Two Self-assessment
Uni-1								
3	43.0	20	50	61.8	Fail	Fail	Satisfactory	Satisfactory
8	23.0	27	62.5	54.1	Pass	Fail	Good	Satisfactory
18	10.0	35	75	47.8	Pass	Fail	Good	Good
20	40.5	45	52.5	31.5	Fail	Fail	Satisfactory	Satisfactory
21	35.5	43	55	35.4	Fail	Fail	Satisfactory	Weak
27	5.0	10	85	77.1	Pass	Pass	Good	Good
36	29.0	28	55	64.4	Fail	Fail	Satisfactory	Satisfactory
39	35.5	17	68.3	73.0	Pass	Pass	Weak	Satisfactory
44	16.0	14	63.3	69.9	Pass	Pass	Good	Satisfactory
31	21.5	15	57.5	53.8	Fail	Fail	Satisfactory	Satisfactory
Uni-2								
5	40.5	26	52.5	55.7	Fail	Fail	Satisfactory	Satisfactory
9	29.0	6	57.5	79.7	Fail	Pass	Satisfactory	Good
23	19.0	13	65	73.7	Pass	Pass	Satisfactory	Satisfactory
29	24.5	39.5	80	38.7	Pass	Fail	Satisfactory	Weak
33	7.5	42	72.5	53.7	Pass	Fail	Satisfactory	Satisfactory
38	29.0	34	63.3	60.6	Pass	Fail	Satisfactory	Satisfactory
30	11.5	29	57.5	49.6	Fail	Fail	Satisfactory	Satisfactory
26	21.5	21	60	41.3	Pass	Fail	Satisfactory	Satisfactory
Uni-3								
2	35.5	18	55	64.2	Fail	Fail	Weak	Poor
13	29.0	39.5	57.5	41.3	Fail	Fail	Satisfactory	Poor
35	35.5	37	55	47.2	Fail	Fail	Satisfactory	Satisfactory
45	32.5	24	56.7	59.3	Fail	Fail	Weak	Poor

Test-taker	Test One rank	Test Two rank	Test One score	Test Two score	Test One Pass/Fail	Test Two Pass/Fail	Test One Self-assessment	Test Two Self-assessment
<b>Uni-4</b>								
7	44.5	44	32.5	27.0	Fail	Fail	Weak	Poor
14	46.0	46	45	51.4	Fail	Fail	Satisfactory	Satisfactory
24	44.5	31	52.5	53.3	Fail	Fail	Satisfactory	Satisfactory
34	40.5	38	65	77.8	Pass	Pass	Satisfactory	Satisfactory
4	32.5	36	45	34.6	Fail	Fail	Weak	Weak
15	40.5	30	52.5	43.7	Fail	Fail	Satisfactory	
17	19.0	8	57.5	47.3	Fail	Fail	Satisfactory	Weak
<b>Uni-5</b>								
25	11.5	11	67.5	64.0	Pass	Fail	Good	Satisfactory
16	4.0	5	60	50.7	Pass	Fail	Weak	Satisfactory
1	9.0	23	72.5	76.5	Pass	Pass	Good	Satisfactory
12	24.5	32	77.5	59.9	Pass	Fail	Good	Satisfactory
41	29.0	41	58.3	60.0	Fail	Fail	Satisfactory	Satisfactory
42	17.0	19	70.0	77.3	Pass	Pass	Satisfactory	Satisfactory
46	26.0	22	53.3	58.1	Fail	Fail	Satisfactory	Satisfactory
10	14.0	9	86.7	82.8	Pass	Pass	Good	Satisfactory
19	38.0	25	57.5	40.8	Fail	Fail	Satisfactory	Satisfactory
<b>IGO</b>								
37	2.0	16	95	65.5	Pass	Fail	Satisfactory	Weak
11	14.0	12	70	75.3	Pass	Pass	Weak	Satisfactory
32	7.5	3	82.5	87.5	Pass	Pass	Good	Satisfactory
6	14.0	33	80	88.7	Pass	Pass	Satisfactory	Satisfactory
22	1.0	2	70	50.5	Pass	Fail	Good	Satisfactory
43	3.0	1	97.5	92.2	Pass	Pass	Good	Good
40	19.0	7	65.0	78.9	Pass	Pass	Satisfactory	Satisfactory
28	6.0	4	90	95.1	Pass	Pass	Good	Good

\*In response to the question: "How would you rate your performance? Poor, weak, satisfactory, good or excellent?"

Another point considered was whether students performed in line with their teachers' expectations. The information provided by Uni-4 was incomplete and is therefore not included in the analysis. Figures 11.10-11.13 show the rankings on both tests for the other student test-takers and their predicted ranking within their group. The Uni-1 ranking was based on recent mean test scores, the others were based on the teachers' direct predictions.

Figure 11.10. Uni-1 participants' rankings on Test One, Test Two and as predicted

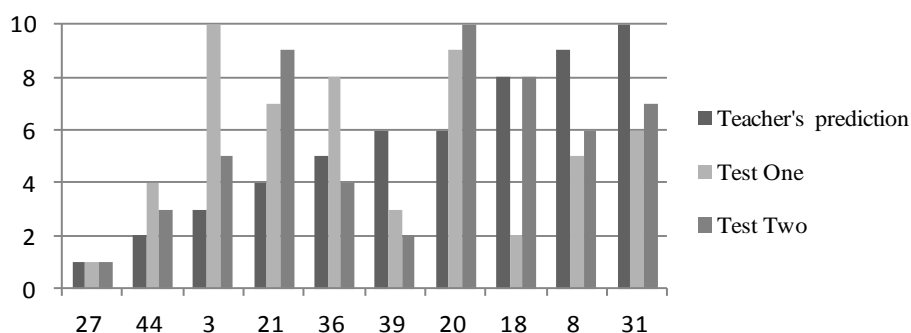


Figure 11.11. Uni-2 participants' rankings on Test One, Test Two and as predicted

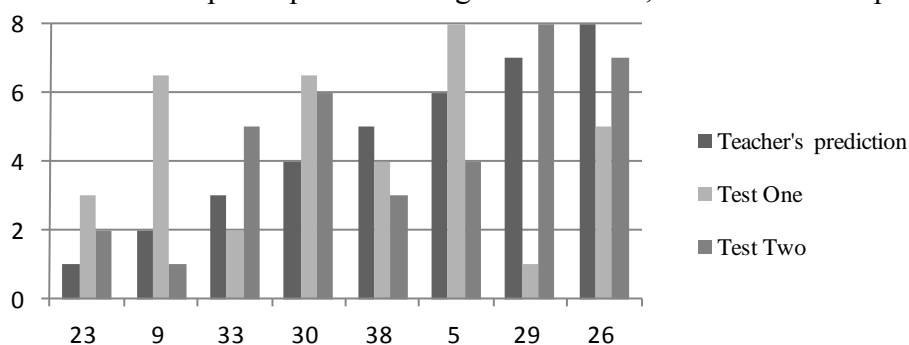
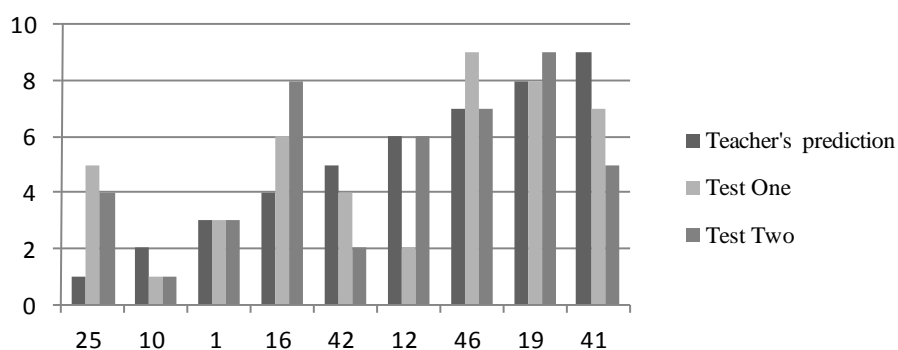


Figure 11.12. Uni-3 participants' rankings on Test One, Test Two and as predicted



Figure 11.13. Uni-5 participants' rankings on Test One, Test Two and as predicted



The predicted and obtained rankings vary considerably for most test-takers in all the groups. The strongest performance was predicted accurately only once (at Uni-1 for both tests), and the weakest not even that. The teachers' predictions were not noticeably closer to the rankings on either test in any group. This is somewhat surprising since Test One, in terms of source text, was thought to be similar to the kinds of tests university students are given. This lack of coincidence between predicted and obtained rankings could suggest several things: that both tests are very different from what the universities give their students, or that predicting performance is very difficult, or that some teachers are better at it than others. This is another area for further research, since it has major implications for both translator training institutions and recruiters. How aware are institutions of what IGO institutions are looking for? And how much can IGO recruiters trust references provided by teachers?

## **11.5 Summary of the analysis of the test trial results**

The analysis suggests that age has some bearing on test performance, but this could be associated with experience in IGO translation since all the older participants belonged in the IGO category. Gender and holding a Master's degree seem to have little influence on performance. The training institutions that participants hailed from might have influenced scores to some extent, but further research would be required to determine whether this had anything to do with the program itself or unrelated aspects of the students' profiles.

Test-taker comments and the pass rates under the intuitive scheme suggest that Test Two was more difficult than Test One. The comparison of the rankings obtained by applying the different scoring methods to both tests suggests that the test task affected performance more than the grading scheme applied. Multi-trait scoring seemed to neutralize the influence of task on scores and to be a more reliable assessment tool than intuitive or points-deduction scoring methods in terms of inter-rater reliability. Both these findings have implications for test design since they suggest that how we test, the task we use and the grading scheme we apply has an impact on performance and needs to be carefully controlled. More importantly, from the viewpoint of this research, the difference between the scores and rankings obtained on the two tests used in the trial is large enough to confirm our fifth hypothesis, namely that a profile-adapted test will produce results that are sufficiently different from those produced by current testing practice as to affect candidate selection in a significant way.

## Chapter 12 Main findings, limitations and suggestions for future research

### 12.1 Main findings

The main findings of the research into the skills-knowledge set required in IGO translation presented at the end of Part I are reproduced here in a more succinct form:

1. IGO translators need far more than language skills: they also need research, computer, analytical and interpersonal skills, as well as extensive general knowledge and, possibly, specialized subject knowledge.
2. Some skills and knowledge types are considered more important than others in terms of their impact on the communicative aims of the organization, thus confirming the hypothesis that:  
*H<sub>1</sub>: Certain skills and knowledge types are more important than others in the context of IGO translation.*
3. Opinion on the relative importance of the components of the skills-knowledge set differs significantly between revisers and translators.
4. Some components of the skills-knowledge set are more often lacking than others, thus confirming the hypothesis that:  
*H<sub>2</sub>: Certain desired skills and knowledge types are more often lacking than others among new recruits at IGOs.*
5. Certain analytical skills are at least as important to find, for example, as rather fundamental components of translation competence, namely knowledge of the source language and target-language writing skills.
6. The importance of certain analytical skills (e.g. detecting inconsistencies and working out obscure passages), research skills (e.g. tracking down resources to check fact) and text-formatting skills (e.g. working with more than basic Word functions) depends on the upstream and downstream quality-control procedures in place in the document production chain.
7. The profile that emerges from the weightings obtained from the cross-referral of the results of the two questionnaires varies from organization to organization and from one language service to another within the same organization.
8. The profile is a dynamic entity that changes with time as the responsibilities of translators, the tools they work with, and the training they receive also change.

A key product of the research presented in Part I was a method for identifying recruitment priorities at IGO translation services. That method was applied to evaluate current text-based testing practice at IGOs and to design a profile-adapted test. The main findings of the research into the testing of candidates for translation work at IGOs are that:

1. Testing practices, in terms of the test task (including test conditions) and grading arrangements, vary considerably from one organization to the next.
2. Not all the important components of the skills-knowledge set identified in the research conducted in Part I are assessed in current recruitment tests, thus confirming the hypothesis that:

*H<sub>3</sub>: Not all the important components of the skills-knowledge set required for IGO translation work are tested in current recruitment examinations.*

The most significant omission is the non-measurement of analytical skills.

3. The analysis of testing practice at two organizations provides evidence to support the hypothesis that:

*H<sub>4</sub>: The weighting of skills and knowledge types in current recruitment examinations does not correspond to that of the profile required in the IGO in question.*

4. Some organizations set domain-relevant test tasks that are done using tools and under conditions similar to those on the job. Others, however, use source texts that bear little resemblance to the documents translated by the IGO, prohibit access to resources in examinations, and have candidates write out their translations by hand.
5. In some organizations, testing practice varies from one language service to the other. The content and difficulty of examinations can thus vary according to the language combinations the candidates offer.
6. No organization seems to work with task specifications beyond a basic text type, a fixed word count for source texts and a time limit for task completion.
7. Many organizations use highly subjective evaluation methods, such as intuitive pass/fail assessments, sometimes carried out by only one person. Others make arrangements for dual, anonymous marking and provide some form of written or oral guidelines. No verification of inter-rater or intra-rater reliability seems to be carried out.

The main findings of the comparison of performance on a traditional test and a profile-adapted were as follows:

1. The rankings obtained by the same group of test-takers on a profile-adapted test, designed in the light of empirically identified priorities and measurement theory, and a mock traditional test were sufficiently different to support the hypothesis that:

*H<sub>5</sub>: A profile-adapted test will produce scores and rankings that are significantly different from those produced by current testing practice.*

2. The nature of the test task seems to affect ranking on translation tests more than the grading scheme used.
3. Multi-trait grading seems to neutralize the influence of the test task and be a less subjective and more reliable scoring method than intuitive or points-deduction methods.
4. IGO experience was the only variable that seemed to influence performance, particularly on the profile-adapted test, suggesting that profile-adapted testing is a more valid way to identify candidates with the skills-knowledge set sought by IGOs.

These findings have several implications for recruitment testing at IGOs. The variation in testing practice across organizations is not necessarily a bad sign, provided that it responds to different recruitment needs; in other words, it can be justified if examinations are tailored to identified recruitment priorities. The lack of standardization, guidelines and specifications, however, suggests that approaches to translation testing are in some, even several, instances rather ad hoc and haphazard, if not arbitrary. In addition to reliability concerns, serious questions about the validity of current recruitment processes are raised by the lack of authenticity of test tasks at many institutions and the finding that examinations often fail to cover all the important components of the skills-knowledge set sought, let alone weight them according to the organization's needs. Profile-adapted testing, based on an empirically identified hierarchy of skills and knowledge types, as demonstrated here, might be a sounder way to select candidates.

## **12.2 Limitations of the research**

Of course, the research has its limitations. Most of these, however, suggest further areas of investigation. Rather surprisingly, the base list of skills and knowledge types presented for rating in the impact and recruits questionnaires was judged to be comprehensive by the survey participants. Only soft skills, such as time management and teamwork, and general knowledge were reported as missing, and then not by a majority of respondents. This may be because the breakdown of translation competence was largely based on the schemes used in



training programs, as derived from the literature review, and hence generally familiar to translators. Considerable thought went into the base list, the goal being to identify and rank components in terms that would be useful as input for the design of text-based tests, but the categorization is ultimately rather arbitrary and does not address the problem of the interrelatedness of the skills and knowledge used in translation. Would the results and implications be very different if translation competence had been broken down into, say, six skills/knowledge areas? As input for text-based testing, such broad categories might not be that useful, but then what level of detail is required? Should the ability to use commas correctly be isolated from punctuation in general? Many revisers would possibly support the move. Ultimately, the measuring instrument can and should be refined and adapted according to individual needs. If used to identify in-house training priorities, for example, the components to be weighted might be quite different, according to the feasibility of their development on the job. This is another area for further research. Apart from being undermined by the relative arbitrariness of the breakdown of translation competence into components, the validity of the measuring instrument is possibly marred by the definitions used. Further research would have to be done to examine how respondents understood some of the concepts presented in the base list.

Although the hierarchies obtained from the survey results are based on expert opinions, the results could benefit from corroboration by other means. Process analysis was dismissed as being unfeasible at the outset of the project, but certainly in narrower contexts, such as an individual language service or a highly specialized organization that translates the same type of text most of the time, process analysis could reveal the skills and knowledge that go into translation work and suggest in what measure they are called into play. The results of that research could be used to confirm or refute the results obtained from the application of the two-questionnaire survey.

The test trial also suffered from several limitations. Practical considerations made it necessary to keep the tests short in terms of source-text length and time allowed for task completion. Several important components of the skills-knowledge set were not assessed, such as general knowledge, specialized knowledge and research skills. One feature of traditional testing, having to write translations out by hand, was not factored into the comparison, and no provisions were made for measuring learning effect. Having to participate as an evaluator myself because of the shortage of available qualified graders possibly compromised the objectivity of the trial and made it more difficult to assess inter-rater reliability. The instructions for one of the grading schemes trialed were misinterpreted and thus managed to highlight one of the problems of traditional testing (inadequate grader

training), but at the expense of the reliability of some of the scores. They should have been worded more clearly. The wording of the descriptors for the rating scales also seems to have been inadequate in a few instances and they could benefit from further trialing and development. As for the test-takers, only a few of them actually matched the profile of typical candidates, who usually have some years' experience before they attempt competitive examinations for IGO translation work. With a sample population of 46, when it came to analysis, the candidate groups were small and heterogeneous, which precluded making very meaningful comparisons regarding the influence of age, experience and training background, for example. These limitations could and should all be addressed if the experiment were to be replicated in the future.

### **12.3. Suggestions for future research**

The analysis of the skills and knowledge required in IGO translation has raised questions that suggest other areas for research: How should the skills that are difficult to assess in text-based tests (e.g. teamwork, organization, time management and interpersonal skills) be measured in recruitment? How effective is the in-house training of new recruits? How can analytical skills be developed? How well are revisers trained to deal with the types of errors most often generated by new recruits? Why do revisers rate the impact of skills and knowledge so differently from translators at the same organization, as suggested by the impact survey findings? Why does the skills-knowledge set required vary from one organization to the next? Is it merely a matter of different task-types and working methods, or are different approaches to translation involved? Why do different language services at the same organization seem to need to find translators with different profiles? Should organizations invest in reference assistants, editing services and text processors or in recruiting translators with research, editing and text-processing skills? All these represent possible fertile areas of enquiry.

The research into IGO recruitment testing has also suggested other lines of investigation. Even though our sample includes the two largest IGO testers and employers of translators, the EU and the UN, the testing survey could be extended to other organizations, and the closer analysis of testing practice could be performed on more translation services. We have identified common features and tendencies in a sample group of just 18 IGO language services. Our considerations have further been limited to the text-based assessment of translators, but there is no reason why oral testing of certain skills or knowledge could not be equally valid or even more appropriate. We have also excluded the interpersonal skills and

personal attributes that contribute to candidate suitability, and this should be factored into the recruitment process. In short, the scope of the research into recruitment testing could be extended in terms of sample size, type of testing and the skills and knowledge studied.

The limitations of our study of test types also suggest many areas for further research. Ideally the profile-adapted test would be tried out on different organizations, with different language combinations, and over several years to see whether it really is a more valid and reliable measuring instrument. Designing the profile-adapted test raised several questions. Does editing a text in L1 draw on the same skills and knowledge as mentally editing a source text while translating? How long should translation tests be, in terms of words per minute and total length? Which aspects of a test task affect performance most: the source text? whether the task is simply translation or translation plus flagging inconsistencies? access to resources? the authenticity of the task? How much does performance vary according to the activation power of the source text? How densely packed with component activators does the text need to be? All these are validity issues that deserve further investigation and could shed new light on our findings and help improve testing practice.

One of the main limitations of the research into testing was the reliability of the scores obtained in the test trial. Scoring translation performance is a major problem. The multi-trait scheme appeared to be the least subjective method and seems to force graders to rate underlying traits, i.e. evidence of skills and knowledge, rather than to assess the quality of the translation produced (revisers probably instinctively do the latter otherwise since it is what they are in the habit of doing). But more research is required to see if this is the case and to examine the inter- and intra-reliability of the scores produced by the multi-trait method, how much evidence is required of each component for valid inferences to be drawn, and whether the benefits outweigh the effort required to implement it. One constraint, even in this project, is finding qualified graders with sufficient time to participate in test trials.

## **12.4 Final remarks**

Identifying ways to improve translation testing would enhance not just recruitment processes, but academic assessment procedures too. The gap between the academic world and the IGO world was highlighted in three ways in the research: in the shortcomings identified among new recruits; in the surprising lack of awareness among students of the standards expected of IGO translators; and in the inability of teachers to predict which students would do best or worst. The empirical results of the impact and recruits questionnaires and the general skills-

knowledge hierarchy identified for IGO translation could be used to begin to address that gap. If trainers use the information to prepare candidates for the IGO profile, there is a greater chance of more suitable candidates graduating from translator training programs and of organizations finding the translators they need. If organizations use the method to identify their recruitment priorities and adjust their candidate selection procedures accordingly, then there is a greater probability of vacancies being filled by the most suitable candidates. The methodology could be applied as often as necessary to update the ideal candidate profile as job descriptions change with advances in technology and the emergence of new organizational arrangements. The practice need not be limited just to IGO translation: the method could easily be adapted to other types of specialized translation, as well as to other fields in which new recruits are mentored, such as in-house interpreting or editing.

The possible value of this research is thus not just the empirical data on the skills and knowledge required in IGO translation and on translation testing practice at IGOs, but the method devised for identifying recruitment priorities and the finding that domain-relevant testing may be more valid, and multi-trait scoring more reliable, than current testing practices. The profile-adapted approach requires considerable investments of time and energy. It is now up to recruiters to decide whether it is worth applying. IGOs are facing a major shortage of translators of the caliber required in the next few years. Hence the pedagogical assistance and internships the UN arranges within the framework of the memorandums of understanding it has signed with select universities around the world, as well as the DGT's support for translator training through the EMT and the Optimale projects. Efforts to improve the training of translators need to be complemented by the improved selection of recruits by IGOs, however. Projected demand is such that, even with improvements in pre-recruitment translator training, IGOs will be forced to employ translators who do not meet the standards set in the past. They need to be sure they can accurately identify those who come closest. This is not just for the sake of productivity, even though hiring mistakes are very expensive and improving selection processes will generate savings down the line. IGOs need to optimize their recruitment practices for ethical reasons too. Recruitment decisions have major impacts on individuals and systems, and, as Bachman notes, “[i]f the potential costs of error are great, it would be unethical not to make every effort to achieve the highest levels of reliability and validity possible” (1990: 57). As the investigation of importance at the outset of this research showed, international organizations need to protect their reputation by ensuring they recruit the best and do so fairly. That means ensuring the validity, reliability and fairness of their procedures for testing and selecting translators.



## References

- Adab, Beverly. 2000. "Evaluating Translation Competence." In *Developing Translation Competence*, Schäffner, C. and Adab, B. (eds.). Amsterdam and Philadelphia: John Benjamins. 215-228.
- Alderson, J. Charles, Caroline Clapham, and Diane Wall. 1995. *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Angelelli, Claudia V. 2009. "Using a rubric to assess translation ability. Defining the construct." In *Testing and Assessment in Translation and Interpreting Studies. A call for dialogue between research and practice*, Angelelli, C. V. and Jacobson, H. E. (eds.). American Translators Association Scholarly Monograph Series 14 (6). Amsterdam and Philadelphia: John Benjamins. 13-47.
- Arango-Keeth, Fanny and Geoffrey S. Koby. 2003. "Assessing assessment. Translation training evaluation and the needs of industry quality assessment." In *Beyond the Ivory Tower: Rethinking Translation Pedagogy*, Baer, B.J. and Koby, G.S. (eds.). Amsterdam and Philadelphia: John Benjamins. 117-134.
- Bachman, Lyle. 1990. *Fundamental Considerations in Language Testing*. Oxford and New York: Oxford University Press.
- Bachman, Lyle and Adrian Palmer. 1996. *Language Testing in Practice*. Oxford and New York: Oxford University Press.
- Baer, Brian J. and Geoffrey S. Koby (eds.). 2003. *Beyond the Ivory Tower: Rethinking Translation Pedagogy*. Amsterdam and Philadelphia: John Benjamins.
- Beeby, Allison. 2000. "Evaluating the Development of Translation Competence". In *Developing Translation Competence*, Schäffner, C. and Adab, B. (eds.). Amsterdam and Philadelphia: John Benjamins. 186-198.
- Beeby, Allison. 2001. "Choosing an Empirical-Experimental Model for Investigating Translation Competence". In *Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects*, Olohan, M. (ed.). Manchester: St Jerome. 29-42.
- Bell, Roger T. 1991. *Translation and Translating - Theory and Practice*. London: Longman.
- Campbell, Stuart J. 1991. "Towards a Model of Translation Competence". *Meta* 36 (2/3): 329-343.
- Campbell, Stuart J. and Sandra Hale. 1999. *What Makes a Text Difficult to Translate?* Refereed Proceedings of the 23rd Annual ALAA Congress of 19 April 1999. <http://iop.al.net/forum/showthread.php?t=2503>. Visited January 2011.

- Campbell, Stuart J. and Sandra Hale. 2003. "Translation and interpreting Assessment in the Context of Educational Measurement". In *Translation Today. Trends and Perspectives*, Anderman, G. and Rogers, M. (eds.). Clevedon: Multilingual Matters. 205-224.
- Cao, Deborah. 1996. "Towards a Model of Translation Proficiency". *Target* 8 (2): 325-340.
- Cao, Deborah and Xingmin Zhao. 2008. "Translation at the United Nations as Specialized Translation". *The Journal of Specialised Translation* 9: 39-54.
- Castellan, N. John and Siegel, Sidney. 1970. *Estadística no paramétrica aplicada a las ciencias de la conducta*. Mexico City: Editorial Trillas.
- Chapelle, Carol. 1998. "Construct Definition and validity inquiry in SLA research." In *Interfaces between Second Language Acquisition and Language Testing research*, Bachman, L.F. and Cohen, A.D. (eds.). Cambridge: Cambridge University Press, 50-166. Reported in Fulcher and Davidson.
- Chesterman, Andrew. 1997. *Memes of Translation*. Amsterdam and Philadelphia: John Benjamins.
- Chesterman, Andrew and Emma Wagner. 2002. *Can Theory Help Translators? A Dialogue between the Ivory Tower and the Wordface*. Manchester: St Jerome.
- Colina, Sonia 2003. "Empirically-based translation pedagogy". In *Beyond the Ivory Tower. Rethinking Translation Pedagogy*, Baer, B. J. and Koby, G. S. (eds.). Amsterdam and Philadelphia: John Benjamins. 29-59.
- Colina, Sonia. 2008. "Translation Quality Evaluation. Empirical Evidence for a Functionalist Approach". *The Translator* 14 (1): 97-134.
- Cushing Weigle, Sarah. 2002. *Assessing Writing*. Cambridge: Cambridge University Press.
- Delisle, Jean. 1980. *Analyse du discours comme méthode de traduction Initiation à la traduction française de textes pragmatiques anglais. Théorie et pratique*. English translation of Part I by Patricia Logan and Monica Creery. 1988. *Translation: An Interpretive Approach*. Ottawa: University of Ottawa Press.
- Delisle, Jean, Hannelore Lee-Jahnke and Monique C. Cormier. 1999. *Translation Terminology*. FIT Monograph Series. Amsterdam and Philadelphia: John Benjamins.
- DGACM, United Nations. 2011a. Language Outreach Portal of the United Nations <http://www.unlanguage.org/LE/Overview/Translators/default.aspx>. Visited October 2011.
- DGACM, United Nations. 2011b. Contractual Translation Unit. Available online at: <http://www.un.org/depts/DGACM/ctu>. Last visited September 2011. Last update 23 May 2011.

- DGT, European Commission. 2008. Translating for a multilingual community. European Commission. Available online at:  
[http://ec.europa.eu/dgs/translation/faq/index\\_en.htm#6](http://ec.europa.eu/dgs/translation/faq/index_en.htm#6). Last visited October 2011.
- EMT. 2009. *Competences for professional translators, experts in multilingual and multimedia communication*.  
[http://ec.europa.eu/dgs/translation/programmes/emt/key\\_documents/emt\\_competences\\_translators\\_en.pdf](http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf). Last visited October 2010.
- Englund Dimitrova, Birgitta. 2005. *Expertise and Explicitation in the Translation Process*. Amsterdam and Philadelphia: John Benjamins.
- European Parliament Council. 2008. "Recommendation of the European Parliament and of the Council of 23 April 2008". *Official Journal of the European Union 2008/C 111/01*. Available online at:  
<http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:111:0001:0007:EN:PDF>. Accessed March 2011.
- Fulcher, Glenn, and Fred Davidson. 2007. *Language Testing and Assessment: An Advanced Resource Book*. London and New York: Routledge.
- Gile, Daniel. 1995. *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam and Philadelphia: John Benjamins.
- Hale, Susan, and Stuart Campbell. 2002. "The Interaction between text difficulty and translation accuracy". *Babel* 48(1): 14-33.
- Hamp-Lyons, Liz. 1991. "Scoring procedures for ESL contexts." In *Assessing Second Language Writing in Academic Contexts*, Hamp-Lyons, L. (ed.). Norwood, NJ: Ablex. 241-276.
- Hatim, Basil, and Ian Mason. 1997. *The Translator as Communicator*. London and New York: Routledge.
- Hindle, William H. 1965/1984. *A Guide to Writing for the United Nations*. New York. Department of Conference Services, United Nations. Internal document (ST/DCS/3) also available at:  
[http://www.un.org/depts/OHRM/sds/lcp/English/docs/a\\_guide\\_to\\_writing\\_for\\_the\\_united\\_nations.pdf](http://www.un.org/depts/OHRM/sds/lcp/English/docs/a_guide_to_writing_for_the_united_nations.pdf). Last visited October 2010.
- House, Juliane. 1998. "Quality in Translation". In *Encyclopedia of Translation Studies*. Baker, M. and K. Malmkjaer (eds.). London and New York: Routledge. 197-200.
- Howard, Diane. 2011. "Failure analysis in a professional translation setting" In *Translation Research Projects 3*, Pym, A. (ed.). Tarragona: Intercultural Studies Group. 15-29. Also available online at:



[http://isg.urv.es/publicity/isg/publications/trp\\_3\\_2011/index.htm](http://isg.urv.es/publicity/isg/publications/trp_3_2011/index.htm)

- Hurtado Albir, Amparo. 1995. "La didáctica de la traducción. Evolución y estado actual" in *Perspectivas de la traducción*, Fernández, P. and Bravo, J. M. (eds.). Valladolid: Universidad de Valladolid. 49-74.
- Hurtado Albir, Amparo. 1996. "La enseñanza de la traducción directa 'general'. Objetivos de aprendizaje y metodología". In *La enseñanza de la traducción*. Castelló: Universitat Jaume I. 31-56.
- Kelly, Dorothy A. 2000. "Text Selection for Developing Translator Competence: Why Texts from the Tourist Sector Constitute Suitable Material". In *Developing Translation Competence*, Schäffner, C. and Adab, B. (eds.). Amsterdam and Philadelphia: John Benjamins. 158-166.
- Kelly, Dorothy A. 2005. *A Handbook for Translator Trainers: A Guide to Reflective Practice*. Manchester: St. Jerome.
- Kiraly, Donald C. 1995. *Pathways to translation: pedagogy and process*. Kent, Ohio: Kent State University Press.
- Kiraly, Donald C. 2000. *A Social-Constructivist Approach to Translator Education*. Manchester: St Jerome.
- Knoke, David and George W. Bohrnstedt. 1994. *Statistics for Social Data Analysis*. Itasca, Ill.: F.E. Peacock.
- Koller, Werner. 1979. *Einführung in die Übersetzungswissenschaft*. Heidelberg. Quell & Meyer.
- Koskinen, Kaisa. 2000. "Institutional Illusions. Translating in the EU Commission: A Finnish Perspective". *The Translator* 6(1): 49-65.
- Koskinen, Kaisa. 2008. *Translating Institutions. An Ethnographic Study of EU Translation*. Manchester: St Jerome.
- Kussmaul, Paul. 1995. *Training the Translator*. Amsterdam and Philadelphia: John Benjamins.
- Mossop, Brian. 1988. "Translating Institutions: a missing factor in translation theory". *TTR : traduction, terminologie, redaction* 1(2): 65-71.
- Mossop, Brian. 2001. *Revising and Editing for Translators*. Manchester: St Jerome.
- Munday, Jeremy. 2001. *Introducing Translation Studies*. London and New York: Routledge.
- Neubert, Albert. 2000. "Competence in Language, Languages, and in Translation". In *Developing Translation Competence*, Schäffner, C. and Adab, B. (eds.). Amsterdam and Philadelphia: John Benjamins. 3-18:

- Nord, Christiane. 2005. *Text Analysis in Translation: Theory, Method, and Didactic Application of a Model for Translation-Oriented Text Analysis*. Amsterdam and New York: Rodopi.
- Orozco, Mariana. 2000. "Building a measuring instrument for the acquisition of translation competence in trainee translators." In *Developing Translation Competence*, Schäffner, C. and Adab, B. (eds.). Amsterdam and Philadelphia: John Benjamins. 199-214.
- Orozco, Mariana and Amparo Hurtado Albir. 2002. "Measuring Translation Competence Acquisition". *Meta* 47 (3): 375-402.
- PACTE. 2000. "Acquiring Translation Competence: Hypotheses and Methodological Problems in a Research Project". In *Investigating Translation*, Beeby, A., Ensinger, D. and Presas, M. (eds.). Amsterdam and Philadelphia: John Benjamins. 99-106.
- PACTE. 2003. "Building a Translation Competence Model". In *Triangulating Translation: Perspectives in Process Oriented Research*, Alves, F. (ed.). Amsterdam and Philadelphia: John Benjamins. 43-66.
- PACTE. 2008. "First Results of Translation Competence Experiment: 'Knowledge of Translation' and 'Efficacy of the Translation Process'". In *Translator and Interpreting Training: Issues Methods and Debates*, Kearns, J. (ed.). London: Continuum. 104-126.
- PACTE. 2011. "Results of the Validation of the PACTE Translation Competence Model: Translation Project and Dynamic Translation Index." In *IATIS Yearbook 2010*, O'Brien, S. (ed.). London: Continuum (forthcoming), also available at [http://grupsderecerca.uab.cat/pacte/sites/grupsderecerca.uab.cat/pacte/files/2011\\_PACTE\\_Continuum.pdf](http://grupsderecerca.uab.cat/pacte/sites/grupsderecerca.uab.cat/pacte/files/2011_PACTE_Continuum.pdf). Accessed March 2011.
- Presas, Marisa. 2000. "Bilingual competence and translation competence." In *Developing Translation Competence*, Schäffner, C. and Adab, B. (eds.). Amsterdam and Philadelphia: John Benjamins. 20-31.
- Pym, Anthony. 2003. "Redefining Translation Competence in an Electronic Age. In Defence of a Minimalist Approach". *Meta* 48 (4): 481-497.
- Pym, Anthony. 2010. *Exploring Translation Theories*. London and New York: Routledge.
- Reiss, Katarina. 1982. "Como averiguar o grau de dificuldade de uma tradução?". In *Letras de Hoje* (2): 7-19. Reported in Hale and Campbell.
- Reiss, Katarina. 1977/89. "Text types, translation types and translation assessment". In *Readings in Translation Theory*. 1989. Chesterman, A. (ed.). Helsinki: Finn Lectura. 105-115.

- Roberts, Roda. 1984. "Compétence du nouveau diplômé en traduction." In *Traduction et Qualité de Langue. Actes du Colloque Société des traducteurs du Québec/ Conseil de la langue française*. Quebec: Editeur Officiel du Québec. 172-184.
- Robinson, Douglas. 2003. *Becoming a Translator. An Introduction to the Theory and Practice of Translation*. London and New York: Routledge.
- Schäffner, Christina. 2007. "Politics and Translation". In *A Companion to Translation Studies, Topics in Translation 34*, Kuhlwiczak, P. and Littau, K. (eds.). Clevedon: Multilingual Matters. 134-147.
- Sekel, Stephen. 2008. *United Nations Translators and Interpreters - silent partners in the diplomatic process*. Keynote Address delivered by Mr. Yohannes Mengesha, Assistant Secretary-General of the United Nations, at the 18th FIT World Congress, 4-7 August 2008 Shanghai, China. Posted on the intranet (staff-only) website of the United Nations Secretariat in September 2008.
- Toury, Gideon. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam and Philadelphia: John Benjamins.
- Tusler, Robert. 1996. *An Overview of Project Risk Management*.  
<http://www.netcomuk.co.uk~rusteler/project/elements>. Visited April 2011.
- Verleysen, Piet. 2010. Opening address given at the Jiamcatt - International Annual Meeting on Computer-Assisted Translation and Terminology, Luxembourg, 28 April 2010.  
[http://ec.europa.eu/dgs/translation/publications/presentations/speeches/20100428\\_jiamcatt\\_en.pdf](http://ec.europa.eu/dgs/translation/publications/presentations/speeches/20100428_jiamcatt_en.pdf)
- Vermeer, Hans. 1989/2000. "Skopos and commission in translational action". In *The Translation Studies Reader*, Venuti, L. (ed.). London and New York: Routledge. 221-232.
- Waddington, Christopher. 1999. *Estudio comparativo de diferentes métodos de evaluación de traducción general (Inglés-Español)*. Doctoral thesis No. 17/1999. Madrid: Publicaciones de la Universidad Pontificia Comillas.
- Waddington, Christopher. 2001a. "Different Methods of Evaluating Student Translation: The Question of Validity". *Meta* 46 (2): 311-325.
- Waddington, Christopher. 2001b. "Should translations be assessed holistically or through error analysis". *Hermes, Journal of Linguistics* 26: 15-38.
- Wagner, Emma, Svend Bech, and Jesús M. Martínez. 2002. *Translating for the European Union Institutions*. Manchester: St Jerome.

- Way, Catherine. 2008. "Systematic Assessment of Translator Competence: In Search of Achilles' Heel". In *Translator and Interpreting Training: Issues Methods and Debates*, Kearns, J. (ed.). London. Continuum. 88-103.
- Williams, Malcolm. 1989. "The Assessment of Professional Translation Quality: Creating Credibility out of Chaos". *TTR : traduction, terminologie, redaction* 2 (2): 13-33.
- Wilss, Wolfram. 1982. *The Science of Translation: Problems and Methods*. Tübingen: Gunter Narr.



## Appendixes

### Appendix 1. The impact questionnaire used in the first pilot

#### THE SKILLS AND KNOWLEDGE NEEDED BY TRANSLATORS AT YOUR ORGANIZATION

##### A. WHOSE WORK YOU REVISE

Do you revise the work of:

temporary and/or permanent in-house translators?  external translators?  both?

If both, which type of translators will you base your responses on? (Please restrict your answers to just the one)

in-house translators  external translators

##### B. THE SKILLS AND KNOWLEDGE REQUIRED

(i) KNOWLEDGE: How much knowledge do translators at your organization need in the following areas?

1	Knowledge of the source language (vocabulary, expressions, rhetorical devices)	<input type="checkbox"/> None <input type="checkbox"/> Minimal <input type="checkbox"/> Some <input type="checkbox"/> Sound <input type="checkbox"/> Extensive
2	Knowledge of the source-language culture(s) (including sensitive issues, social, economic and political situations)	<input type="checkbox"/> None <input type="checkbox"/> Minimal <input type="checkbox"/> Some <input type="checkbox"/> Sound <input type="checkbox"/> Extensive
3	Knowledge of the subject	<input type="checkbox"/> None <input type="checkbox"/> Minimal <input type="checkbox"/> Some <input type="checkbox"/> Sound <input type="checkbox"/> Extensive
4	Knowledge of the organization and how it works	<input type="checkbox"/> None <input type="checkbox"/> Minimal <input type="checkbox"/> Some <input type="checkbox"/> Sound <input type="checkbox"/> Extensive

(ii) COMPREHENSION AND ANALYTICAL SKILLS: To what extent are the following skills important?

The ability to...

5	understand complex subjects	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
6	master new subjects quickly	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
7	work out the meaning of obscure passages in the source text	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
8	detect inconsistencies, contradictions, nonsense, unintended ambiguities, etc. in the source text	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
9	detect logical, mathematical or factual errors in the source text (eye for detail)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential

(iii) THE BASICS: How important is it that translations at your organization do not contain...?

10	slight meaning errors (inaccuracies)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
11	spelling mistakes	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
12	incorrect grammar	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
13	punctuation errors	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
14	formatting errors	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
15	unwarranted omissions	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
16	unwarranted additions	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
17	errors of fact (e.g. names that have not been checked)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
18	unidiomatic (unnatural-sounding) language	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
19	poor word choices (inexact terms, inappropriate language)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
20	stylistic errors (e.g. dangling modifiers, faulty parallelism, poorly constructed sentences, excessive use of the same high-impact words, mixed metaphors)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential

(iv) OTHER ASPECTS OF TRANSLATION WORK. To what extent are the following skills important?

The ability to...

21	write and convey the source-text message clearly	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
22	tailor language to the targeted reader	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential

23	capture every nuance of the source text in the translation	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
24	bring translations into line with new versions, i.e. input changes smoothly making any modifications necessary to surrounding text	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
25	achieve the right tone and register	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
26	improve on the style of the source text (e.g. remove tedious repetitions, pare down convoluted sentences, eliminate verbosity)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
27	produce translations that flow smoothly even when the source text does not	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
28	improve on the content of the source text (remove, correct or at least point out redundancies, unintended ambiguities, contradictions, unexplained acronyms, misleading headings, etc.)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential

(v) RESEARCH SKILLS. To what extent are the following skills important?

The ability to...

29	find the correct terminology	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
30	track down sources of relevant background information (mine information)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
31	work with reference documents (to find accepted phrasing and terminology)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential

(vi) COMPUTER SKILLS. To what extent are the following skills important?

The ability to...

32	type accurately and fairly fast	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
33	reproduce the formatting of the source text	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
34	work with more than basic Word functions (macros, track changes, tables, autocorrect, etc.)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
35	work with translation memory software	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
36	work with Excel documents	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
37	work with PowerPoint presentations	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential

(vii) OTHER SKILLS. To what extent are the following skills important?

The ability to...

38	meet tight deadlines (translate fast and under pressure)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
39	explain their translation decisions or problems they have with the source text (e.g. to authors, users or revisers)	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
40	follow complicated instructions about what needs to be done with a text (additions that need translating, parts that need relocating, patching together, revising against new versions,	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential

	etc.)	
41	adhere to in-house style guidelines	<input type="checkbox"/> Irrelevant <input type="checkbox"/> Not very important <input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential

(vii) Please list any other skills or knowledge required and rate their importance using the scale below

1		<input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
2		<input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
3		<input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential
4		<input type="checkbox"/> Desirable <input type="checkbox"/> Very important <input type="checkbox"/> Essential

Thank you so much for your input! I will send you a summary of the findings as soon as I can. Please now complete your profile and return this form by e-mail to .....

Once again, I assure you total anonymity.

#### YOUR PROFILE

1. Organization \_\_\_\_\_
2. Years with the organization \_\_\_\_\_
3. Most recent position in the organization:
  - Head of department/unit/service
  - Reviser
  - Translator
  - Other (please specify) \_\_\_\_\_
4. Years working as a translator or reviser for international organizations \_\_\_\_\_
5. Years working in translation in general \_\_\_\_\_



## Appendix 2. The recruits questionnaire used in the first pilot

### A. WHOSE WORK YOU REVISE

Do you revise the work of:

- a. in-house permanent staff translators?  b. in-house short-term translators?  
 c. external translators?

### B. THE SKILLS AND KNOWLEDGE THAT NEW RECRUITS LACK

Please use the following scale:

1=Never 2=Rarely 3=Sometimes 4=Often 5=Always NR=Skill/Knowledge not required

Also please indicate whether the skill or knowledge is expected to be acquired before or after recruitment

(i) Knowledge: How often do new recruits lack the following...?

1=Never 2=Rarely 3=Sometimes 4=Often 5=Always NR=Skill/Knowledge not required

1	Sufficient knowledge of the source language (vocabulary, expressions, rhetorical devices)	1 2 3 4 5 NR	pre-recruitment	post-recruitment
2	Sufficient knowledge of the source-language culture(s) (including sensitive issues, social, economic and political situations)	1 2 3 4 5 NR	pre-recruitment	post-recruitment
3	Sufficient knowledge of the subject	1 2 3 4 5 NR	pre-recruitment	post-recruitment
4	Sufficient knowledge of the organization and how it works	1 2 3 4 5 NR	pre-recruitment	post-recruitment

(ii) Comprehension and analytical skills: How often do new recruits lack the ability to...?

1=Never 2=Rarely 3=Sometimes 4=Often 5=Always NR=Skill/Knowledge not required

5	understand complex subjects	1 2 3 4 5 NR	pre-recruitment	post-recruitment
6	master new subjects quickly	1 2 3 4 5 NR	pre-recruitment	post-recruitment
7	work out the meaning of obscure passages in the source text	1 2 3 4 5 NR	pre-recruitment	post-recruitment
8	detect inconsistencies, contradictions, nonsense, unintended ambiguities, etc. in the source text	1 2 3 4 5 NR	pre-recruitment	post-recruitment
9	detect logical, mathematical or factual errors in the source text (eye for detail)	1 2 3 4 5 NR	pre-recruitment	post-recruitment

(iii) The translations that new recruits produce. How often do the translations of new recruits contain...?

1=Never 2=Rarely 3=Sometimes 4=Often 5=Always N/A=Not applicable

10	meaning errors (inaccuracies)	1 2 3 4 5 NR
11	spelling mistakes	1 2 3 4 5 NR
12	incorrect grammar	1 2 3 4 5 NR
13	punctuation errors	1 2 3 4 5 NR
14	formatting errors	1 2 3 4 5 NR
15	unwarranted omissions	1 2 3 4 5 NR
16	unwarranted additions	1 2 3 4 5 NR
17	errors of fact (e.g. names that have not been checked)	1 2 3 4 5 NR
18	unidiomatic (unnatural-sounding) language	1 2 3 4 5 NR
19	poor word choices (inexact terms, inappropriate language)	1 2 3 4 5 NR
20	stylistic errors (e.g. dangling modifiers, faulty parallelism, tedious repetition, over-use of the same high-impact words, mixed metaphors)	1 2 3 4 5 NR

(iv) Other aspects of translation work. How often do new recruits lack the ability to...?

1=Never 2=Rarely 3=Sometimes 4=Often 5=Always NR=Skill/Knowledge not required

21	write and convey the source-text message clearly	1 2 3 4 5 NR	pre-recruitment	post-recruitment
22	tailor language to the targeted reader	1 2 3 4 5 NR	pre-recruitment	post-recruitment
23	capture every nuance of the source text in the translation	1 2 3 4 5 NR	pre-recruitment	post-recruitment
24	bring translations into line with new versions, i.e. input changes smoothly making any modifications necessary to surrounding text	1 2 3 4 5 NR	pre-recruitment	post-recruitment
25	achieve the right tone and register	1 2 3 4 5 NR	pre-recruitment	post-recruitment

26	improve on the style of the source text (e.g. remove tedious repetitions, pare down convoluted sentences, eliminate verbosity)	1 2 3 4 5 NR	pre-recruitment	post-recruitment
27	produce translations that flow smoothly even when the source text does not	1 2 3 4 5 NR	pre-recruitment	post-recruitment
28	improve on the content of the source text (remove, correct or at least point out redundancies, unintended ambiguities, contradictions, unexplained acronyms, misleading headings, etc.)	1 2 3 4 5 NR	pre-recruitment	post-recruitment

(v) Research skills. How often do new recruits lack the ability to...?

1=Never 2=Rarely 3=Sometimes 4=Often 5=Always

NR=Skill/Knowledge not required

29	find the correct terminology	1 2 3 4 5 NR	pre-recruitment	post-recruitment
30	track down useful sources of background information (mine information)	1 2 3 4 5 NR	pre-recruitment	post-recruitment
31	work with reference documents (to find accepted phrasing and terminology)	1 2 3 4 5 NR	pre-recruitment	post-recruitment

(vi) Computer skills. How often do new recruits lack the ability to...?

1=Never 2=Rarely 3=Sometimes 4=Often 5=Always

NR=Skill/Knowledge not required

32	type accurately and fairly fast	1 2 3 4 5 NR	pre-recruitment	post-recruitment
33	reproduce the formatting of the source text	1 2 3 4 5 NR	pre-recruitment	post-recruitment
34	work with more than basic Word functions (macros, track changes, tables, autocorrect, etc.)	1 2 3 4 5 NR	pre-recruitment	post-recruitment
35	work with translation memory software	1 2 3 4 5 NR	pre-recruitment	post-recruitment
36	work with Excel documents	1 2 3 4 5 NR	pre-recruitment	post-recruitment
37	work with PowerPoint presentations	1 2 3 4 5 NR	pre-recruitment	post-recruitment

(vii) Other skills. How often do new recruits lack the ability to...?

1=Never 2=Rarely 3=Sometimes 4=Often 5=Always

NR=Skill/Knowledge not required

38	meet tight deadlines (translate fast and under pressure)	1 2 3 4 5 NR	pre-recruitment	post-recruitment
39	explain their translation decisions or problems they have with the source text (e.g. to authors, users or revisers)	1 2 3 4 5 NR	pre-recruitment	post-recruitment
40	follow complicated instructions about what needs to be done with a text (additions that need translating, parts that need relocating, patching together, revising against new versions, etc.)	1 2 3 4 5 NR	pre-recruitment	post-recruitment
41	adhere to in-house style guidelines	1 2 3 4 5 NR	pre-recruitment	post-recruitment

(vii) Please list any other skills or knowledge that new recruits lack and rate their frequency using the same scale and indicating when they are expected to be acquired

2=Rarely 3=Sometimes 4=Often 5=Always

1		1 2 3 4 5 NR	pre-recruitment	post-recruitment
2		1 2 3 4 5 NR	pre-recruitment	post-recruitment
3		1 2 3 4 5 NR	pre-recruitment	post-recruitment
4		1 2 3 4 5 NR	pre-recruitment	post-recruitment

## C. IN-HOUSE TRAINING

What in-house training is provided for new recruits? (Multiple answers possible)

 Formal training (e.g. courses or workshops)

Please describe which skills and knowledge are covered and duration of training

 Mentoring by more experienced in-house translators Other, please specify \_\_\_\_\_

## D. COMMENTS OR OBSERVATIONS ON THE SKILL SET REQUIRED AT YOUR ORGANIZATION

Thank you so much for your input! I will send you a summary of the findings as soon as I can. Please now complete your profile and return this form by e-mail to .....

## YOUR PROFILE

1. Organization \_\_\_\_\_
2. Years with the organization \_\_\_\_\_
3. Most recent position in the organization:
  - Head of department/unit/service
  - Reviser
  - Translator
  - Other (please specify) \_\_\_\_\_
4. Years working as a reviser for international organizations \_\_\_\_\_
5. Year working in translation in general \_\_\_\_\_

### Appendix 3. How revisers become aware of translators' weaknesses

Selected snippets of interaction between revisers and translators during feedback sessions observed at an IGO between 10 January and 15 March 2009.<sup>38</sup>

Reviser: Why did you use “assets” for “acciones” here?

Translator: I thought that was what *acciones* were.

Reviser: Well, could they have been assets?

Translator: Ugh, I don't know...

Reviser: Have you ever heard of territorial ordering?

Translator: No, but that is what it says and I couldn't think of anything better.

Reviser: You used “increase” three times in this paragraph. It was getting a bit repetitive.

Translator: Well, they use “aumentó” three times in the Spanish.

Reviser: We talk about “official development assistance” not “aid”.

Reviser: We use “older adults” not “the elderly”.

Translator: Wow, I didn't know that.

Reviser: Well, it was in the background document I sent you. Did you check it?

Reviser: Watch it! You can't use a high-impact word like “trigger” twice in one sentence.

Reviser: I had to completely rewrite this section. It wasn't very clear. Did you have problems understanding the original?

Translator: No, I understood it ok, I just didn't know how to get all the ideas across in one sentence.

Reviser: Well, you could have split the sentence in two....

Translator: Is it always “compared with” not “compared to”? I see you changed it here...

Reviser: Well it depends if you are comparing like elements or not...

Reviser: You should really have checked whether there was an official translation for this organization.

Translator: I couldn't find one.

Reviser: Well it's on the official ECOSOC list.

Translator: Oh, I see you changed “which” to “that”. Can you explain why?

Reviser: You should really use the same translation for this throughout, you know.

Translator: Oh yes, sorry, I didn't notice.

Reviser: You might want to check the punctuation rules, there are good explanations in ...

---

<sup>38</sup> Anonymity was requested and granted.

## Appendix 4. The impact questionnaire used in the second pilot

### Section A

1. Which organization are you currently working for?
2. What is your current position in the organization?
  - Head of Department/Unit/Service
  - Reviser
  - Translator
3. Please base your answers in the following sections on the work of either in-house translators or external translators (but not both). Which will you base your answers on?
  - The work of in-house translators
  - The work of external translators
4. Which languages do you translate from in your current post?
  - Arabic  Chinese  English  French  German  Portuguese  Russian  Spanish
  - Others (please specify)
5. Which languages do you translate into in your current post?
  - Arabic  Chinese  English  French  German  Portuguese  Russian  Spanish
  - Others (please specify)

### Section B

For the purposes of this survey, effective translations are those that achieve the communicative aims of the organization and protect its image.

How large is the impact of the following skills and knowledge types on the effectiveness of translations at your organization? 1= minimal; 5 = enormous; N/A = not applicable because the skill or knowledge is not required.

1. Knowledge of the source language (vocabulary, expressions, rhetorical devices)
2. Knowledge of the different varieties of the source language
3. Knowledge of the source-language culture(s) (history, geography, economic and political situation, customs, value-laden concepts, sensitive issues, etc.)
4. Knowledge of the subject (technical knowledge, e.g. of economics, international law, science, technology)
5. Knowledge of the organization and how it works
6. The ability to understand complex topics
7. The ability to master new subjects quickly (i.e. gain more than a layperson's knowledge)
8. The ability to work out the meaning of obscure passages in the source text
9. The ability to detect inconsistencies, contradictions, nonsense, unintended ambiguities, misleading headings, etc. in the source text
10. The ability to detect mathematical errors in the source text
11. An extensive vocabulary in the target language
12. Knowledge of spelling rules in the target language
13. Knowledge of the finer points of grammar of the target language
14. Knowledge of punctuation rules in the target language
15. The ability to produce idiomatic (natural-sounding) language in the target text
16. The ability to produce translations that flow smoothly even when the source text does not
17. The ability to find the most appropriate word or word combination
18. The ability to recast sentences in the target language (to say the same thing in a different ways)
19. The ability to write elegantly regardless of the style of the original
20. The ability to write and convey the source-text message clearly
21. The ability to capture every nuance of the source text in the translation
22. The ability to achieve the appropriate tone and register in the target text
23. Knowledge of target-language varieties
24. Knowledge of target-language cultures (knowledge of history, geography, economic and political situation, customs, traditions, belief systems, value-laden concepts, sensitive issues, etc.)
25. The ability to tailor the language of the target text to the readers' needs
26. The ability to adhere to in-house style conventions
27. The ability to ensure the completeness of the target text (i.e. no unwarranted omissions)
28. The ability to ensure the coherence of the target text (e.g. consistent terminology use, no contradictions, logical connections of ideas)
29. The ability to track down sources of information to check facts
30. The ability to track down sources to obtain a better grasp of the thematic aspects of a text (understand the topic)
31. The ability to mine reference material for accepted phrasing and terminology (those used by the organization or in a specialized field)
32. The ability to judge the reliability of information sources

33. The ability to type accurately and fast
34. The ability to translate fast and under pressure (to meet tight deadlines)
35. The ability to explain translation decisions or translation problems posed by the source text (e.g. to authors, users or revisers)
36. The ability to follow complicated instructions about what needs to be done with a text (additions that need translating, parts that need relocating, patching together, revising against new versions, etc.)
37. The ability to make effective use of translation memory software
38. The ability to make effective use of electronic terminology tools
39. The ability to work with more than basic Word functions (formatting, macros, track changes, tables, autocorrect, etc.)
40. The ability to work with Excel documents and/or PowerPoint presentations

### Section C

1. Are new recruits expected to have all the skills and types of knowledge you have rated from 1-5 before they join the organization?  
No (please specify which knowledge or skills are expected to be acquired after recruitment)
2. Please list any other skills or types of knowledge that have an impact on the effectiveness of translations at your organization and rate their impact using the same scale as above (1 = minimal impact; 5 = enormous impact).E.g. Knowledge of translation theory: 3
3. Please use the space below to make any observations about the skills and knowledge that translators need at your organization or to comment on the questionnaire. Your input is much appreciated.
- 4.. Thank you so much for taking the time to answer the questionnaire. Which e-mail address would you like the summary of the survey findings to be sent to?

## Appendix 5. The recruits questionnaire used in the second pilot

### Section A

1. Which organization are you currently working for?
2. What is your current position in the organization?
  - Head of Department/Unit/Service
  - Reviser
  - Translator
3. Please base your answers in the following sections on the work of either in-house translators or external translators (but not both). Which will you base your answers on?
  - The work of in-house translators
  - The work of external translators
4. Which languages do you translate from in your current post?
  - Arabic  Chinese  English  French  German  Portuguese  Russian  Spanish
  - Others (please specify)
5. Which languages do you translate into in your current post?
  - Arabic  Chinese  English  French  German  Portuguese  Russian  Spanish
  - Others (please specify)

### Section B

For the purposes of this questionnaire, new recruits are translators who have been working with the organization for less than 12 months.

Please think about the mistakes you usually correct when going over translations by new recruits. How often do you think those mistakes are due to a lack of the following things?

1 = Almost never; 5 = Almost always; N/A = Not applicable because the skill or knowledge is not required of new recruits.

1. Knowledge of the source language (vocabulary, expressions, rhetorical devices)
2. Knowledge of the different varieties of the source language
3. Knowledge of the source-language culture(s) (history, geography, economic and political situation, customs, value-laden concepts, sensitive issues, etc.)
4. Knowledge of the subject (technical knowledge, e.g. of economics, international law, science, technology)
5. Knowledge of the organization and how it works
6. The ability to understand complex topics
7. The ability to master new subjects quickly (i.e. gain more than a layperson's knowledge)
8. The ability to work out the meaning of obscure passages in the source text
9. The ability to detect inconsistencies, contradictions, nonsense, unintended ambiguities, misleading headings, etc. in the source text
10. The ability to detect mathematical errors in the source text
11. An extensive vocabulary in the target language
12. Knowledge of spelling rules in the target language
13. Knowledge of the finer points of grammar of the target language
14. Knowledge of punctuation rules in the target language
15. The ability to produce idiomatic (natural-sounding) language in the target text
16. The ability to produce translations that flow smoothly even when the source text does not
17. The ability to find the most appropriate word or word combination
18. The ability to recast sentences in the target language (to say the same thing in a different ways)
19. The ability to write elegantly regardless of the style of the original
20. The ability to write and convey the source-text message clearly
21. The ability to capture every nuance of the source text in the translation
22. The ability to achieve the appropriate tone and register in the target text
23. Knowledge of target-language varieties
24. Knowledge of target-language cultures (knowledge of history, geography, economic and political situation, customs, traditions, belief systems, value-laden concepts, sensitive issues, etc.)
25. The ability to tailor the language of the target text to the readers' needs
26. The ability to adhere to in-house style conventions
27. The ability to ensure the completeness of the target text (i.e. no unwarranted omissions)
28. The ability to ensure the coherence of the target text (e.g. consistent terminology use, no contradictions, logical connections of ideas)
29. The ability to track down sources of information to check facts
30. The ability to track down sources to obtain a better grasp of the thematic aspects of a text (understand the topic)

31. The ability to mine reference material for accepted phrasing and terminology (those used by the organization or in a specialized field)
32. The ability to judge the reliability of information sources
33. The ability to type accurately and fast
34. The ability to translate fast and under pressure (to meet tight deadlines)
35. The ability to explain translation decisions or translation problems posed by the source text (e.g. to authors, users or revisers)
36. The ability to follow complicated instructions about what needs to be done with a text (additions that need translating, parts that need relocating, patching together, revising against new versions, etc.)
37. The ability to make effective use of translation memory software
38. The ability to make effective use of electronic terminology tools
39. The ability to work with more than basic Word functions (formatting, macros, track changes, tables, autocorrect, etc.)
40. The ability to work with Excel documents and/or PowerPoint presentations

## Section C

1. In order to contextualize the findings, please answer the following questions about document processing at your organization using the scale provided. (Never Rarely Sometimes Often Always)  
How often are source texts edited (for style and/or content) before they are translated?  
How often are source texts pretranslated before being sent to the translator (previously translated text is highlighted or provided)?  
How often are source texts referenced before being sent to the translator (names are checked, official translations of terms given, background documents and sources identified, etc.)?  
How often in practice are translations fully revised against the source text?  
How often are translations (whether revised or not) sent to the author or requester for approval?
2. In your translation unit, are translators responsible for the following aspects of the translated text? (Yes, No)  
Formatting  
Spelling  
Adhering to in-house style conventions  
Observations
3. Please use the space below to make any observations about the skills and knowledge that new translators need and lack at your organization or to comment on the questionnaire. Your input is much appreciated.
4. Thank you so much for taking the time to answer the questionnaire. Which e-mail address you would like the summary of the survey findings to be sent to?



## Appendix 6. Final version of the impact questionnaire

### Section A

1. Which organization are you currently working for?
2. What is your current position in the organization?
  - Head of Department/Unit/Service
  - Reviser
  - Translator
3. Please base your answers in the following sections on the work of either in-house translators or external translators (but not both). Which will you base your answers on?
  - The work of in-house translators
  - The work of external translators
4. Which languages do you translate from in your current post?  
 Arabic  Chinese  English  French  German  Portuguese  Russian  Spanish  
Others (please specify)
5. Which languages do you translate into in your current post?  
 Arabic  Chinese  English  French  German  Portuguese  Russian  Spanish  
Others (please specify)

### Section B

For the purposes of this survey, effective translations are those that achieve the communicative aims of the organization and protect its image.

How large is the impact of the following skills and knowledge types on the effectiveness of translations at your organization? 1= minimal; 5 = enormous; N/A = not applicable because the skill or knowledge is not required.

1 = Almost never; 5 = Almost always; N/A = Not applicable because the skill or knowledge is not required of new recruits.

1. Knowledge of the source language (vocabulary, expressions, rhetorical devices)
2. Knowledge of the different varieties of the source language
3. Knowledge of the source-language culture(s) (history, geography, economic and political situation, customs, value-laden concepts, sensitive issues, etc.)
4. Knowledge of the subject (technical knowledge, e.g. of economics, international law, science, technology)
5. Knowledge of the organization and how it works
6. The ability to understand complex topics
7. The ability to master new subjects quickly (i.e. gain more than a layperson's knowledge)
8. The ability to work out the meaning of obscure passages in the source text
9. The ability to detect inconsistencies, contradictions, nonsense, unintended ambiguities, misleading headings, etc. in the source text
10. The ability to detect mathematical errors in the source text
11. An extensive vocabulary in the target language
12. Knowledge of spelling rules in the target language
13. Knowledge of the finer points of grammar of the target language
14. Knowledge of punctuation rules in the target language
15. The ability to produce idiomatic (natural-sounding) language in the target text
16. The ability to produce translations that flow smoothly even when the source text does not
17. The ability to select and combine words in the target language to capture the exact and detailed meanings (nuances) of the source text
18. The ability to recast sentences in the target language (to say the same thing in different ways)
19. The ability to produce an elegantly written target text regardless of how elegantly written the source text is
20. The ability to convey the source-text message clearly
21. The ability to convey the intended effect of the source text
22. The ability to achieve the appropriate tone and register in the target text
23. Knowledge of target-language varieties
24. Knowledge of target-language cultures (knowledge of history, geography, economic and political situation, customs, traditions, belief systems, value-laden concepts, sensitive issues, etc.)
25. The ability to tailor the language of the target text to the readers' needs
26. The ability to adhere to in-house style conventions
27. The ability to ensure the completeness of the target text (i.e. no unwarranted omissions)
28. The ability to ensure the coherence of the target text (e.g. consistent terminology use, no contradictions, logical connections of ideas)
29. The ability to track down sources of information to check facts

30. The ability to track down sources to obtain a better grasp of the thematic aspects of a text (understand the topic)
31. The ability to mine reference material for accepted phrasing and terminology (those used by the organization or in a specialized field)
32. The ability to judge the reliability of information sources
33. The ability to type accurately and fast
34. The ability to maintain quality even when translating under time pressure
35. The ability to justify translation decisions and explain translation problems posed by the source text (e.g. to authors, users or revisers)
36. The ability to follow complicated instructions about what needs to be done with a text (additions that need translating, parts that need relocating, patching together, revising against new versions, etc.)
37. The ability to make effective use of translation memory software
38. The ability to make effective use of electronic terminology tools
39. The ability to work with more than basic Word functions (formatting, macros, track changes, tables, autocorrect, etc.)
40. The ability to work with Excel documents and/or PowerPoint presentations

### Section C

1. Are new recruits expected to have all the skills and types of knowledge you have rated from 1-5 before they join the organization?  
No (please specify which knowledge or skills are expected to be acquired after recruitment)
2. Please list any other skills or types of knowledge that have an impact on the effectiveness of translations at your organization and rate their impact using the same scale as above (1 = extremely small impact; 5 = extremely large impact). E.g. Knowledge of translation theory: 3
3. Please use the space below to make any observations about the skills and knowledge that translators need at your organization or to comment on the questionnaire. Your input is much appreciated.
4. Thank you so much for taking the time to answer the questionnaire. Which e-mail address would you like the summary of the survey findings to be sent to?

## Appendix 7. Final version of the recruits questionnaire

### Section A

1. Which organization are you currently working for?
2. What is your current position in the organization?
  - Head of Department/Unit/Service
  - Reviser
  - Translator
3. Please base your answers in the following sections on the work of either in-house translators or external translators (but not both). Which will you base your answers on?
  - The work of in-house translators
  - The work of external translators
4. Which languages do you translate from in your current post?
  - Arabic  Chinese  English  French  German  Portuguese  Russian
  - Spanish
  - Others (please specify)
5. Which languages do you translate into in your current post?
  - Arabic  Chinese  English  French  German  Portuguese  Russian
  - Spanish
  - Others (please specify)

### Section B

For the purposes of this questionnaire, new recruits are translators who have been working with the organization for less than 12 months.

Please think about the mistakes you usually correct when going over translations by new recruits. How often do you think those mistakes are due to a lack of the following things?

1 = Almost never; 5 = Almost always; N/A = Not applicable because the skill or knowledge is not required of new recruits.

1. Knowledge of the source language (vocabulary, expressions, rhetorical devices)
2. Knowledge of the different varieties of the source language
3. Knowledge of the source-language culture(s) (history, geography, economic and political situation, customs, value-laden concepts, sensitive issues, etc.)
4. Knowledge of the subject (technical knowledge, e.g. of economics, international law, science, technology)
5. Knowledge of the organization and how it works
6. The ability to understand complex topics
7. The ability to master new subjects quickly (i.e. gain more than a layperson's knowledge)
8. The ability to work out the meaning of obscure passages in the source text
9. The ability to detect inconsistencies, contradictions, nonsense, unintended ambiguities, misleading headings, etc. in the source text
10. The ability to detect mathematical errors in the source text
11. An extensive vocabulary in the target language
12. Knowledge of spelling rules in the target language
13. Knowledge of the finer points of grammar of the target language
14. Knowledge of punctuation rules in the target language
15. The ability to produce idiomatic (natural-sounding) language in the target text
16. The ability to produce translations that flow smoothly even when the source text does not
17. The ability to select and combine words in the target language to capture the exact and detailed meanings (nuances) of the source text
18. The ability to recast sentences in the target language (to say the same thing in different ways)
19. The ability to produce an elegantly written target text regardless of how elegantly written the source text is
20. The ability to convey the source-text message clearly
21. The ability to convey the intended effect of the source text
22. The ability to achieve the appropriate tone and register in the target text
23. Knowledge of target-language varieties
24. Knowledge of target-language cultures (knowledge of history, geography, economic and political situation, customs, traditions, belief systems, value-laden concepts, sensitive issues, etc.)
25. The ability to tailor the language of the target text to the readers' needs
26. The ability to adhere to in-house style conventions
27. The ability to ensure the completeness of the target text (i.e. no unwarranted omissions)
28. The ability to ensure the coherence of the target text (e.g. consistent terminology use, no contradictions, logical connections of ideas)
29. The ability to track down sources of information to check facts

30. The ability to track down sources to obtain a better grasp of the thematic aspects of a text (understand the topic)
31. The ability to mine reference material for accepted phrasing and terminology (those used by the organization or in a specialized field)
32. The ability to judge the reliability of information sources
33. The ability to type accurately and fast
34. The ability to maintain quality even when translating under time pressure
35. The ability to justify translation decisions and explain translation problems posed by the source text (e.g. to authors, users or revisers)
36. The ability to follow complicated instructions about what needs to be done with a text (additions that need translating, parts that need relocating, patching together, revising against new versions, etc.)
37. The ability to make effective use of translation memory software
38. The ability to make effective use of electronic terminology tools
39. The ability to work with more than basic Word functions (formatting, macros, track changes, tables, autocorrect, etc.)
40. The ability to work with Excel documents and/or PowerPoint presentations

### Section C

1. In order to contextualize the findings, please answer the following questions about document processing at your organization using the scale provided. (Never Rarely Sometimes Often Always)
  - (a) How often are source texts edited (for style and/or content) before they are translated?
  - (b) How often are source texts pretranslated before being sent to the translator (previously translated text is highlighted or provided)?
  - (c) How often are source texts referenced before being sent to the translator (names are checked, official translations of terms given, background documents and sources identified, etc.)?
  - (d) How often in practice are translations fully revised against the source text?
  - (e) How often are translations (whether revised or not) sent to the author or requester for approval?
2. In your translation unit, are translators responsible for the following aspects of the translated text? (Yes, No)
  - (a) Formatting
  - (b) Spelling
  - (c) Adhering to in-house style conventionsObservations:
3. Please use the space below to make any observations about the skills and knowledge that new translators need and lack at your organization or to comment on the questionnaire. Your input is much appreciated.
4. Thank you so much for taking the time to answer the questionnaire. Which e-mail address you would like the summary of the survey findings to be sent to?

## Appendix 8. Summary of the comments made in the impact and recruits questionnaires on IGO translation

### *Comments made in the impact questionnaire*

Complaints were made about the restrictions imposed by tradition and conventions. One respondent commented that “Complying with the in-house rules and conventions is sometimes more important than the ability to reproduce a text elegantly or even accurately”, and another stated that “translators have to ‘forget’ their general knowledge and use the terms agreed on at the inter-institutional level even if they don’t agree with them or have a better suggestion.”

There were several laments about having to work under time pressure and to meet productivity targets and about the resulting loss of quality: “Quantity is often more important than quality, unfortunately” remarked one translator; and “Quality is of the essence and should not be put aside because of time constraints” insisted another. Regrets were also voiced about the lack of time to maintain and improve one’s language skills (both source languages and the mother tongue) and to broaden knowledge in supposed areas of specialization.

### *Comments made in the recruits questionnaire*

One reviser stated that less and less work was being revised as the deadlines were such that revisers were being asked to translate rather than to revise, which was having a negative effect on the quality of translations and the mentoring of new recruits.

The situation, of course, varies from one organization to another: as one reviser pointed out, new recruits in many organizations are not novice translators at all; on the contrary, they are deliberately recruited because of their experience. Another noted that “Only for freelance translators who are in their early twenties and are just entering the profession have I seen major problems of lack of skills or knowledge. Such problems should be corrected over time before any organization recruits them to regular staff posts”. This comment reflects a clear belief that the onus is on training schools and translators themselves to get themselves up to scratch.

Similarly the comments about working with poorly written originals, which require the “ability to understand, navigate through and decode the various types of ‘pidgin’ English [...which] is becoming more and more usual within international organizations”, came from revisers working at organizations in which documents are only rarely sent for editing prior to translation.

In some organizations, there seems to be a generational divide inasmuch as some revisers refer to “old-fashioned” and “traditional” skills, while others warn that “New translators in this organization need to be able to step into a time warp. We have some translators who have worked in the private sector and know all about timeliness and keeping the customer satisfied: they have trouble hiding their shock when they arrive. Though the salaries we are paid soon help them to see the benefits of an older, more sedate system, where we still use red ink on paper.”

There was also recognition that the demands of the job have increased, especially in terms of turnaround times and the number and wide range of new subjects being incorporated into the international agenda (and in the case of the EU, the need to master several new languages). One reviser said, “I can only fully subscribe to the generic definition of future duties carried out by translators working at EU institutions”.<sup>39</sup> After quoting the long list of

---

<sup>39</sup> “The institutions are seeking to recruit highly qualified university graduates with a degree in languages and/or in subjects such as law, economics, auditing and finance, natural sciences, political and social sciences, and technology who are capable of adapting throughout their career. Successful candidates will have to be able to translate texts into their main language from at least two languages. The texts are mainly political, legal, economic, financial, scientific or technical in nature, are frequently demanding and encompass all the spheres of

expectations, he added the remark “Good luck!”. Whether this was directed towards candidates (because they are unlikely to ever be able to meet the rather ambitious expectations) or the EU (because it will be impossible to find translators who match the profile), or both, remains unclear, but it does reveal the general feeling that translators are being asked to do more and more. As another reviser, at a specialized agency of the UN, wrote: “The focus on number of words or pages per day is such that specific skills and knowledge upgrades get lost in the process, leaving some colleagues on the sideway. This is an area your research may take a look at as more and more is expected of translators within shorter deadlines and ongoing multiple translation and non-translation assignments/jobs are expected of them.”

The questionnaire inspired many comments about training needs. There were calls for: standardized in-house training modules for new recruits; closer cooperation between organizations and translation schools; more, and more intense, mentoring by senior revisers; clearer instruction or training on in-house style; the appointment of an experienced translator or reviser with teaching skills to supervise new recruits; and, in translator training programmes, the use of more typical source texts, intensive mother tongue development, and thorough grounding in the basics of economics, law, international relations, etc. (as opposed to just terminology). Some respondents noted that it takes time (months) to become familiar with in-house style and texts and much longer (one estimated 2-3 years) to become a fully competent, reliable and productive translator in an organization. Others wondered whether some of the required skills could be taught at all: “As to making sense of badly written texts, I’m not sure it’s a skill that can ever be taught or learned”, noted one reviser with regret; “Spelling conventions and the use of translation software can be taught, but the ability to reproduce the logic of an argument (aside from understanding the technical subjects) seems to be more innate”, commented another.

---

activity of the European Union. This work requires candidates to be fully conversant with computers and office technology. Your ‘main language’ is your mother tongue or a language of which you have an equivalent command. The European Union institutions attach particular importance to the ability to grasp problems of all kinds, often complex in nature, to react rapidly to changing circumstances and to communicate effectively. You will have to show initiative and imagination and be highly motivated. You should be able to work frequently under pressure, both on your own and in a team, and adjust to a multicultural working environment. You will also be expected to develop your professional skills throughout your career. (<http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:025A:0001:0014:EN:PDF> ).”

## Appendix 9. Questions asked in the questionnaires on ad hoc tests and competitive examinations.

1. Which organization are you answering for?\*
2. Which type of posts or work are ad hoc tests used to recruit for? (Multiple answers possible)
  - (a) Permanent staff positions
  - (b) Short-term in-house work
  - (c) External freelance or contractual work
  - (d) Other (please specify in space below)
3. How are applicants screened before taking your organization's ad hoc test? (Multiple answers possible)
  - (a) They aren't: anyone can take the test.
  - (b) They are invited to sit the test on the basis of their CV.
  - (c) References are obtained from previous employers.
  - (d) They have to pass other tests first (e.g. general knowledge, writing, psychometric tests, etc.)
 If you checked (d) or use any other screening procedure, please elaborate in the space below:
4. Language directionality. What is tested? (Multiple answers possible)
  - (a) Translation into Language 1 (mother tongue or main language of education)
  - (b) Translation out of Language 1
 Comments or explanations
5. Do candidates have to do anything other than translate (e.g. summarize or edit a text, revise a translation)?
6. How many translations are done in the test? (Do not include translation of isolated terms or phrases, if applicable)
7. Please state how many words normally have to be translated in how much time in each translation (up to a maximum of 5). For example, if usually about 300 words have to be translated in 90 minutes, put "300 in 90".
8. Where do test texts come from?
  - (a) An internal source (a text produced by or for the organization)
  - (b) An external source (e.g. the press, another organization, literature)
  - (c) Both internal and external sources
 If you checked (b) or (c), what kind of external sources are normally used (e.g. newspapers, scientific journals, etc.)?
9. What type of texts are candidates asked to translate?
  - (a) General texts only
  - (b) Specialized texts only
  - (c) Both general and specialized texts
    - (i) Which of these descriptions applies to the specialized test texts used for this testing? (multiple answers possible): They cover a specialized or technical subject handled by the organization.
      - They cover a little-known specialized subject.
      - They include technical language a layperson would not know.
      - They contain in-house terminology and acronyms.
10. Test-text features
  - i. How often do the texts selected for use as translation test papers have the following features? (Please rate each one) Scale: never-rarely-sometimes-often-always
    - a. In terms of content, the text is typical of the documents translated by the translation service.
    - b. In terms of style and register, the text is typical of the documents translated by the translation service.

- c. The text is well written (coherent, cohesive, consistent use of terms, etc.).
- d. The text is poorly written in parts (sloppy style, grammar, inconsistent use of terms, etc.).
- e. The text has a particular rhetorical style (e.g. speeches, letters, laws).
- f. The text contains stylistic features unique to the source language or culture.
- g. The text contains only straightforward vocabulary.
- h. The text contains difficult vocabulary.
- i. The text is noticeably written in a particular regional variety of the source language.
- j. The text contains in-house terminology and acronyms.
- k. The text contains sections that are difficult to translate into gender-neutral or politically correct language.
- l. The text contains culture-specific concepts that require explanation in the target language.
- m. The text contains logic errors, contradictions or inconsistencies.
- n. The text contains passages in which the meaning is obscure.
- o. The text contains subtleties that could easily be missed.
- p. The text contains sections that require considerable reformulation in the target language.

Please describe any other content-related features of the test texts used in the space below.

11. Are the texts wholly authentic or are they modified to meet certain criteria?

- (a) Authentic (reproduced exactly as found)
- (b) Modified

If they are modified, please explain in what way.

12. Are candidates given (real or fictitious) instructions or background information? (e.g. information about the origins of the source text, the requester, the intended reader, or the purpose of the translation)

13. Are candidates required to complete the test on a computer?

- (a) Yes
- (b) No

Computer-based examinations : please complete the table. Scale: never-rarely-sometimes-often-always

- Are test texts in Word?
- Are test texts in Excel?
- Are test texts in PowerPoint?
- Are test texts in Pdf?
- Do test texts include graphics?
- Are candidates asked to use the track changes function?

Please specify any other document formatting skills tested.

14. Are candidates provided with any material by the examiners other than the test papers (e.g. glossaries, dictionaries, background documents, editorial style guidelines, suggested websites to consult)?

15. Are candidates allowed to bring any resources with them to the examination? (e.g. glossaries or dictionaries)

16. Are candidates tested... Scale: never-rarely-sometimes-often-always

- (a) at a specific location under supervision?
- (b) remotely without supervision?
- (c) under another type of arrangement? If you ever use other types of arrangements (c), please describe them in the space below.

17. Are any measures taken to ensure that candidates cannot cheat?

Please describe any measures taken

18. Grading

i. Are there written guidelines for grading test papers?



(a) Yes

(b) No

Please explain how test papers are graded. Alternatively provide your name and e-mail address so that I may contact you for a copy of the guidelines.

ii. How long does it take, on average, for one person to grade one candidate's test answers (all the translation test papers together)?

iii. Do you evaluate the quality of translations only, or do you also evaluate the way the candidate translates?

(a) Quality of translation only

(b) Quality of translation and how candidate translates

Feel free to comment on this point in the space below

19. How would the following errors/omissions in a candidate's test answers be viewed in your assessment process? Please use the following 1-5 scale, where: 1= an irrelevant error (will not affect whether the candidate passes or fails); 5 = a fatal error (instant disqualification/elimination); and N/A = Not applicable due to the way the test is designed.

- a. Basic distortion of the text's overall meaning(wrong message conveyed)
- b. Loss of nuance, shift in emphasis, slight distortions
- c. Lack of lexical precision (individual words mistranslated)
- d. Failure to use the correct technical term
- e. Failure to use official/correct names for organizations, programmes, countries, or acronyms
- f. Inappropriate tone or register
- g. Lack of flow, naturalness, readability
- h. Internal inconsistencies
- i. Failure to complete the assignment in the allotted time
- j. Failure to adhere to in-house style guidelines
- k. Failure to detect, correct or point out inconsistencies or slips in logic
- l. Grammatical errors
- m. Punctuation errors
- n. Spelling errors
- o. Typing errors
- p. Formatting errors
- q. Unwarranted omissions
- r. Unwarranted additions

What other error types, if any, are taken into consideration in your assessment process? Please rate their relative importance using the same scale as above.

20. What features of a translation (other than an absence of errors, if applicable) could improve a candidate's chances of passing at your organization? E.g. added explanations or glosses, strikingly apt expressions, solutions of difficult problems, fast completion of the task, omission of redundancies, correction of inconsistencies, evidence of thorough research, evidence of technical or cultural knowledge, etc. Please list them in order of importance.

21. Please use the space below for comments or observations about ad hoc testing at your organization, existing constraints (e.g. budget or logistics), perceived strengths or weaknesses of current or past practice, etc. Please also feel free to comment on the questionnaire. Any input is much appreciated.

Thank you so much for taking the time to answer the questionnaire. Please indicate in the space below which e-mail address you would like the summary of the survey findings to be sent to.

## Appendix 10. Graders survey

As some of you may know, I am researching translation test construction. I am currently looking at grading schemes and assessment criteria and I would really appreciate your input.

If you have a moment, could you please indicate how heavily you weighted different skills and knowledge in your most recent assessment of examination papers by completing the table below.

Notes:

- ❖ Please base your answers on the last set of papers you assessed.
- ❖ Please read through the whole list before beginning so that you rate the types of evidence in relation to each other.
- ❖ Please add any other evidence that you take into consideration and rate it using the same scale.

*Please circle a number on the scale provided, where 1=not taken into consideration and 5=the most heavily weighted. If there was no such evidence to be weighted, please put "0".*

Evidence of knowledge of the source language	1 2 3 4 5 - 0
Evidence of general knowledge	1 2 3 4 5 - 0
Evidence of ability to capture exact and detailed meanings (nuances)	1 2 3 4 5 - 0
Evidence of ability to write clearly	1 2 3 4 5 - 0
Evidence of ability to write correctly (e.g. grammar, spelling, capitalization)	1 2 3 4 5 - 0
Evidence of ability to produce target texts that flow smoothly	1 2 3 4 5 - 0
Evidence of the ability to produce a coherent target text (no internal inconsistencies, logical connection of ideas)	1 2 3 4 5 - 0
Evidence of ability to convey the intended effect of the source text	1 2 3 4 5 - 0
Evidence of the ability to achieve the right tone and register for the text	1 2 3 4 5 - 0
Other evidence taken into consideration and how heavily it is weighted	

Are there any instant disqualifiers, i.e. things that will result in automatic failure? (e.g. a grammar mistake, failure to follow instructions)

Any comments, explanations, observations, etc., would be most welcome. Please feel free to use the back of the sheet as well if you need the space.

## Appendix 11. The abbreviations used to refer to the skills and knowledge types rated in the second pilot and the changes made to the base list after the second pilot.

Full description and number in questionnaires	Abbreviated form used in the analysis of the results
1. Knowledge of the source language (vocabulary, expressions, rhetorical devices)	Knowledge of the SL
2. Knowledge of the different varieties of the source language	Knowledge of SL varieties
3. Knowledge of the source-language culture(s) (history, geography, economic and political situation, customs, value-laden concepts, sensitive issues, etc.)	Knowledge of SL culture(s)
4. Knowledge of the subject (technical knowledge, e.g. of economics, international law, science, technology)	Knowledge of the subject
5. Knowledge of the organization and how it works	Knowledge of the organization
6. The ability to understand complex topics	Understand complex topics
7. The ability to master new subjects quickly (i.e. gain more than a layperson's knowledge)	Master new subjects quickly
8. The ability to work out the meaning of obscure passages in the source text	Work out the meaning of obscure ST
9. The ability to detect inconsistencies, contradictions, nonsense, unintended ambiguities, misleading headings, etc. in the source text	Detect inconsistencies, etc.
10. The ability to detect mathematical errors in the source text	Detect mathematical errors in the ST
11. An extensive vocabulary in the target language	An extensive TL vocabulary
12. Knowledge of spelling rules in the target language	Knowledge of TL spelling rules
13. Knowledge of the finer points of grammar of the target language	Knowledge of TL grammar
14. Knowledge of punctuation rules in the target language	Knowledge of TL punctuation rules
15. The ability to produce idiomatic (natural-sounding) language in the target text	Produce idiomatic translations
16. The ability to produce translations that flow smoothly even when the source text does not	Produce translations that flow smoothly
17. The ability to find the most appropriate word or word combination	Find the best words
18. The ability to recast sentences in the target language (to say the same thing in a different ways)	Recast sentences in the TL
19. The ability to write elegantly regardless of the style of the original	Write elegantly regardless of the ST
20. The ability to write and convey the source-text message clearly	Convey the ST message clearly
21. The ability to capture every nuance of the source text in the translation	Capture every nuance of the ST
22. The ability to achieve the appropriate tone and register in the target text	Achieve the right tone and register
23. Knowledge of target-language varieties	Knowledge of TL varieties
24. Knowledge of target-language cultures (knowledge of history, geography, economic and political situation, customs, traditions, belief systems, value-laden concepts, sensitive issues, etc.)	Knowledge of TL cultures
25. The ability to tailor the language of the target text to the readers' needs	Tailor language to the readers' needs
26. The ability to adhere to in-house style conventions	Adhere to in-house style conventions
27. The ability to ensure the completeness of the target text (i.e. no unwarranted omissions)	Ensure the completeness of the TT
28. The ability to ensure the coherence of the target text (e.g. consistent terminology use, no contradictions, logical connections of ideas)	Ensure the coherence of the TT
29. The ability to track down sources of information to check facts	Track down sources to check facts
30. The ability to track down sources to obtain a better grasp of the thematic aspects of a text (understand the topic)	Track down sources to understand topic
31. The ability to mine reference material for accepted phrasing and terminology (those used by the organization or in a specialized field)	Mine reference material for phrasing
32. The ability to judge the reliability of information sources	Judge the reliability of sources
33. The ability to type accurately and fast	Type accurately and fast
34. The ability to translate fast and under pressure (to meet tight deadlines)	Translate fast and under pressure
35. The ability to explain translation decisions or translation problems	Explain translation decisions or

<p>posed by the source text (e.g. to authors, users or revisers)</p> <p>36. The ability to follow complicated instructions about what needs to be done with a text (additions that need translating, parts that need relocating, patching together, revising against new versions, etc.)</p> <p>37. The ability to make effective use of translation memory software</p> <p>38. The ability to make effective use of electronic terminology tools</p> <p>39. The ability to work with more than basic Word functions (formatting, macros, track changes, tables, autocorrect, etc.)</p> <p>40. The ability to work with Excel documents and/or PowerPoint presentations</p>	<p>problems</p> <p>Follow complicated instructions</p> <p>Work with translation memory software</p> <p>Work with electronic terminology tools</p> <p>Handle more than basic Word functions</p> <p>Work with Excel and/or PowerPoint</p>
---	---

Numbers refer to the number of the question on the skill or knowledge type in the impact and recruits questionnaires used in the second pilot. The only changes made after the second pilot to the base list of skills and knowledge types presented for rating in the questionnaires were as follows:

<b>Full description and number in questionnaires</b>	<b>Abbreviated form used in the analysis of the results</b>
17. The ability to select and combine words in the target language to capture the exact and detailed meanings (nuances) of the source text. Abbreviation	Capture nuances of ST
21. The ability to convey the intended effect of the source text	Convey the intended effect of the ST
34. The ability to maintain quality even when translating under time pressure	Maintain quality even under time pressure

## Appendix 12. Source text used in Test One

The text was taken from a Uruguayan newspaper.<sup>40</sup> Deletions are marked with struck through text, and additions are marked in italics. The words and phrases that were expected to not necessarily be known to all test-takers but which they were expected to be able to guess from context have been underlined.

### Aprendizaje frustrado

Nuestro país está pasando por una ~~racha de~~ bonanza económica. Dicha situación da pie a comentarios, consejos y presagios de variada calidad. Proviene de los cuatro puntos cardinales de la cátedra, la política y el periodismo y *la opinión pública*.

Algunos hablan de medidas contracíclicas; son los economistas: tienen sus estadísticas y sus libretos. Si fueran menos técnicos dirían: ojo, ~~que la alegría va por barrios~~ *en tiempo de vacas gordas*... (viene a ser lo mismo). Los vinculados a la función de gobierno -porque lo integran o integraron alguna vez- ofrecen cálculos respecto a ~~cuánta piola se puede dar a la cometa~~ *cuánta rienda suelta se puede dar a la situación* antes de que se dispare la inflación. Un tercer grupo son los pesimistas: esto se va a acabar, no puede durar, la burbuja está a punto de pincharse.

Yo no soy economista, no tengo elementos para emitir una opinión fundada. En cuanto a la futurología, ella me atrae poco. Por temperamento (y por edad) tiendo a ser cauteloso. Además más allá de lo económico quiero encaminar mi reflexión hacia otro terreno del cual se podría derivar otro tipo de beneficio.

Es evidente que el país vive un período de vigorosa expansión económica ~~y que ésta ya lleva varios años: hace un par de~~ *zafras que la gente viene con el riñón cubierto*. El año que recién termina arrojó una venta de automóviles cero kilómetro nunca vista: más de 40.000 unidades. Un escalón más abajo, la venta de motos y ciclomotores fue récord. El que tiene bolsillo pequeño gastó un poco más que los años pasados y el de bolsillo grande gastó mucho más. En términos que hacen justicia al famoso olfato del presidente Mujica para captar y describir la realidad nacional, *los adinerados abarrotaron los centros comerciales para comprar zapatillas de marca* ~~oligarcas abarrotaron los shopping centers para comprar championes~~ o cambiar el televisor.

Pero a lo que voy es que ese fenómeno indiscutible es analizado y comentado generalmente sólo en términos cuantitativos. Quiero señalar -y ese es el punto de este artículo- que es esencialmente diferente el progreso económico (~~el enriquecimiento~~) que se gesta como resultado o bien de más esfuerzo, o del ingenio aplicado o bien de mayor tecnificación o de una más inteligente organización, que el enriquecimiento que sobreviene sin que hayamos hecho nada diferente de lo que veníamos haciendo siempre. Y eso es lo que está pasando. Es una sorpresa y no una conquista.

Un progreso económico que sea producto de un mayor esfuerzo nacional (trabajo) se convierte, además, en un aprendizaje. El fenómeno económico adquiere una explicación: se progresó porque se acometieron tales y cuales tareas o porque se eliminaron tales o cuales rémoras; aparece un camino, indicado y confirmado por el éxito palpable.

Cuando el enriquecimiento sólo sobreviene produce sorpresa; una linda sorpresa, pero no una lección, no enseña nada, no muestra caminos, no confirma ni desmiente nada. Al revés: fomenta la aversión al trabajo y ~~huele a batacazo de timbero~~ *la creencia en las buenas rachas de fortuna*. Y al Uruguay le hace mal cualquier evento que le alimente su inveterada sospecha de que el progreso económico sólo se puede esperar del amparo del gobierno o de un golpe de suerte proveniente del exterior: nunca del trabajo propio, del ingenio, de la habilidad y del tesón ~~aplicados con conocimiento y cimentados en autoestima~~.

<sup>40</sup> Posadas, Juan Martín. 2011. "Aprendizaje frustrado". In *El País*, edition of 13 March 2011. Available online at: <http://www.elpais.com.uy/110313/predit-553004/juanmartinposadas/aprendizaje-frustrado/>

## Appendix 13. Base texts used to develop the source text translated in the profile-adapted test

1. Speech given by Alicia Bárcena, Executive Secretary of the Economic Commission for Latin America and the Caribbean in Guatemala, 11 October 2010. available online at: [http://www.eclac.cl/prensa/noticias/discursossecretaria/6/41166/Discursoguatemala\\_111010.pdf](http://www.eclac.cl/prensa/noticias/discursossecretaria/6/41166/Discursoguatemala_111010.pdf)

[...] en estos últimos 40 años los costos acumulados son de aproximadamente 127 mil millones de dólares. De este total, la economía se ha visto afectada en un 33%, el desarrollo social un 35%, y la infraestructura en un 25%. Cuánto se podría haber logrado invertir productivamente con estos \$127 mil millones de dólares o bien si tuviéramos una buena estrategia de prevención de riesgos no estaríamos contando con esta cifra tan alta. Vemos con pesar que los países y regiones afectadas desgraciadamente son también países con alta vulnerabilidad social.

2. Text taken from *DESASTRES Y DESARROLLO: EL IMPACTO EN 2010 (Cifras preliminares)* Unidad de Desastres, CEPAL, Coordinated by Ricardo Zapata. Available online at [http://www.eclac.cl/desastres/noticias/noticias/2/42102/Desastres2010\\_WEB.pdf](http://www.eclac.cl/desastres/noticias/noticias/2/42102/Desastres2010_WEB.pdf). Page 3.

### DESASTRES Y DESARROLLO: EL IMPACTO EN 2010

En lo corrido de 2010 se han producido una serie de eventos climáticos extremos, particularmente en el arco continental de la cuenca del Caribe (desde México a Colombia y Venezuela en la cuenca del Caribe) y en la región andina (particularmente en Ecuador y Bolivia), sobrepuestos a una reducida capacidad de las cuencas y ecosistemas frágiles: Laderas inestables y humedales y napas saturadas para poder asimilar el impacto excesivo y repetido de las lluvias y tormentas en la región. Es decir que se suma variabilidad y cambio climático a degradación ambiental para potenciar el riesgo y ocasiona los daños y pérdidas observados. Esta situación de riesgo incrementado —presente en otras regiones en desarrollo particularmente en el sudeste asiático y en la zona subsaheliana de África— ha llevado a la aplicación de la metodología de evaluación de desastres en numerosos casos en el marco de la asociación del sistema de Naciones Unidas con el Banco Mundial y la Unión Europea.

### CUADRO 1.

#### RESUMEN PRELIMINAR DE IMPACTO DE DESASTRES EN AMÉRICA LATINA Y EL CARIBE EN 2010

Número	Tipo de desastre	Muertes	Población afectada	Costo Millones de \$US
98	TOTAL	225,684	13,868,359	49,188
13	Epidemiológico	1,211	33,470	565
79	Climatológico	1,380	9,318,685	9,840
6	Geofísico	223,093	4,214,934	38,783

Hay evidencia científica creciente que confirma que el cambio climático ha incrementado la intensidad de los ciclones y de las tormentas tropicales principalmente a través de su efecto sobre la elevación de la temperatura atmosférica y de la superficie del mar<sup>1</sup>. Si este comportamiento se agudiza en las próximas décadas, los países de la región y muy particularmente Centroamérica, la región Andina y el Caribe, potencialmente enfrentarán eventos climáticos más devastadores.

Este incremento de las amenazas pone de relieve las vulnerabilidades existentes en estas sociedades en los ámbitos social, económico y ambiental, ampliando substancialmente el riesgo asociado a desastres naturales. Nótese que las causas de la mayor intensidad de estos fenómenos tuvieron su origen en procesos de industrialización de las naciones hoy desarrolladas, y a los que las naciones de nuestra región en su mayor parte han contribuido de manera mínima. Son, sin embargo, como Guatemala, muy amenazados por el propio cambio climático y muy vulnerables debido a insuficiencias en la institucionalidad para un desarrollo resiliente. Para internalizar las externalidades provocadas por el cambio climático por quienes las han causado. Nos urge la realización de foros y procesos de negociación ambiciosos y exitosos en el ámbito global.

**Appendix 14. Source text used in Test Two (the profile-adapted test)**

The base texts can be found in Appendix 13. Some of the challenges are identified in italics, and changes made after the pilot to reduce the word count are indicated in a smaller font.

En lo transcurrido de 2010 se han producido [*correct tense must be used here- the text was written in 2010*] una serie de eventos climáticos extremos particularmente en la cuenca del Caribe (desde México a Venezuela en [“el arco continental de” removed] la cuenca del Caribe [*repetition unnecessary*]) y en los Andes [previously “en la región andina”] (particularmente en Ecuador y Bolivia), y tan solo los daños económicos han alcanzado más de 49 mil millones de dólares. Por el monto de los daños y pérdidas y por el número de personas fallecidas el impacto es disímil entre países, siendo Chile, Haití y Brasil los que más sufrieron en la región, si bien el costo económico es muy considerable en términos absolutos para México, Brasil o Colombia (si bien el efecto completo en este último país no se podrá tener hasta que termine la estación lluviosa que promete extenderse hasta inicios del próximo año) y, en términos relativos en Santa Lucía por el tamaño de ese país isleño [*long sentence needs to be drafted carefully to make meaning clear*].

Asimismo el evento más dramático es sin duda el terremoto de Haití que causó la muerte de más de 220.000 personas y desplazó 1.760.000 fuera de sus lugares de vida normal dejándoles hasta el día de hoy en refugios temporales con graves riesgos (cólera, inseguridad alimentaria [*possibly unfamiliar term*], servicios básicos, agua, saneamiento) [*will need reformulation, not all of these are risks*].

En el cuadro arriba [*it is below*] se resume el impacto [“de los desastres en la región” removed] a lo largo de las últimas cuatro décadas.

**CUADRO 1. IMPACTO ACUMULADO DE DESASTRES EN LA REGIÓN** [*previously “América latina y el Caribe”*], **SEGÚN TIPO DE DESASTRES Y SECTORES AFECTADOS (1970-2000)** [*should be 2010, said four decades and the report is for 2010 also next paragraph shows must be 2010; commas need changing to points and vice versa*]

[“sectores” removed] (millones de dólares de 2008)

	<b>Climatológico</b>	<b>Geofísico</b>	<b>Epidemiológico</b>	<b>Total</b>	<b>Porcentaje</b>
<b>Económicos</b>	137.064,7	34.363,8	3.105,6	174.534,1	48,9%
<b>Sociales</b>	22.829,8	59.551,6	1.133,3	83.514,8	23,4%
<b>Infraestructura</b>	49.361,2	33.899,9	2.996,2	86.257,3	24,2%
<b>Medio ambiente</b>	1.582,4	257,2	44,6	1.884,2	0,5%
<b>Otros</b>	1.916,1	2.526,1	52,3	4.494,5	1,3%
<b>Remoción escombros y limpieza</b>	649,0	2.587,4	0,0	3.236,4	0,9%
<b>Gastos emergencias</b>	1.460,5	1.289,7	0,9	2.751,1	0,8%
<b>Total</b>	214.863,8	134.475,8	7.332,8	356.672,4	1000% [ <i>should be 100%</i> ]

Llama a la reflexión cuánto se podría haber logrado invertir productivamente con el monto del impacto acumulado o bien si tuviéramos una buena estrategia de prevención de riesgos no estaríamos contando con cifras tan altas [*obscure meaning here, will require strong analytical skills. Use of “we” perspective typical of Spanish discourse but usually not reflected in IGO translations unless a political speech*]. Mientras tanto la distribución sectorial del impacto en el 2010 ha diferido notablemente de la acumulada: la economía se ha visto afectada en un 33%, el desarrollo social en un 35%, y la infraestructura en un 25%, 49%, 23% y 24%, respectivamente), mostrándose una caída importante [*“importante” here means “large”*] en el protagonismo de los daños económicos [*the table will help make sense of this part. “Protagonism” will not work in English*].

De otra parte la heterogeneidad regional queda en evidencia al analizar, evaluar y copilar los datos para [*“medir el impacto de los desastres en” removed*] las distintas subregiones de la región [*repetition of region/subregion to be avoided*], evidenciando, en general, consecuencias más severas en las [*here refers to subregions*] con países de menor tamaño, menor desarrollo relativo, mayor dependencia del medio ambiente para su producción y menor diversificación y tecnificación de la misma [*some vocabulary problems: possibly “technification”; translators will need to make sure they know what “misma” refers to*].

Asimismo hay evidencia científica creciente que confirma que gracias [*inappropriate use of “thanks to”*] al cambio climático los países de la región y desgraciadamente sobre todo los menos [*should be more not less*] vulnerables antes descritos, potencialmente enfrentarán eventos climáticos más devastadores en las próximas décadas. Nótese que [*typical Spanish structure, needs reformulation in English*] las causas de la mayor intensidad de estos fenómenos tuvieron su origen en los procesos de industrialización de los EEUU y las naciones hoy desarrolladas [*must put in “other” or they will imply the US is not developed*], a los que [*“las naciones de” removed*] nuestra región en su mayor parte han contribuido de manera mínima. De allí que para asegurar que sean internalizadas las externalidades provocadas por el cambio climático por quienes las han causado, nos urge la realización de foros y procesos de negociación ambiciosos y exitosos en el ámbito global [*challenge to make this sound natural*].



## Appendix 15. Questions included in the feedback questionnaires for test trial participants

### Test One

Please answer all questions. Otherwise you will not be able to move on to the next section.

Thank you for your feedback, it is highly appreciated.

1. Your name (please indicate the name you used to identify your translation. It is required only to cross-refer results and will not appear in any publication)
2. Your age group: 21-25; 26-30; 31-35; 36-40; 41-4;5 46-50; 51-55; 56-6;0 >60
3. Sex: Male Female
4. Academic qualifications already obtained. B.A./B.Sc.; M.A./M.Sc.; Other (please specify under subject studied)
5. Subjects studied
6. University at which you are pursuing an M.A./ Organization(s) where you currently work
7. Total experience working as a translator for international organizations. Please report time actually spent working, i.e. years of experience, not the time since you began, in case they are different). None; < 6 months; 6-12 months; 1-2 years; 3-5 years; 6-10 years; 11-15 years; 16-20 years; 21-30 years; over 30 years.
8. Total experience working as a translator outside international organizations (please report time actually spent working, i.e. years of experience, not the time since you began, in case they are different). None; < 6 months; 6-12 months; 1-2 years; 3-5 years; 6-10 years; 11-15 years; 16-20 years; 21-30 years; over 30 years

Please answer the following questions as fully as possible in the spaces provided. You may wish to use the line numbering of the source text for reference.

1. Were the instructions clear? Yes; No; If not, please explain why not.
2. How long did it take you to produce a first version? To complete the translation? Please comment. Did you feel rushed? If you did not finish, how much more time would you have liked?
3. Which words, phrases or parts, if any, did you have problems understanding at first?
4. Terminology problems aside, which were the most difficult parts to phrase in English? What was difficult? Please identify the part(s) and explain. Capturing the detailed level of meaning; Making the text read well in English; Something else
5. What type of non-linguistic knowledge do you think the text called for, if any? General knowledge; Subject knowledge (e.g. economics); Knowledge of the source culture; None; Other (please specify)
6. Overall, on a scale of 1 - 5, how difficult do you think the translation task was? (1= extremely easy, and 5= extremely difficult)
7. How would you rate your performance? Poor; Weak; Satisfactory; Good; Excellent
8. Please feel free to comment below on the translation test. Is it what you expected? Do you feel you were you able to show how well you can translate? Do you think it was fair?
9. Would you like feedback on your performance? Yes; No

### Test Two

The questions were the same as in the feedback questionnaire for Test One, except that the open-ended comments section invited respondents to make a comparison with Test One:

“Please feel free to comment below on the translation test. Is it what you expected? Do you feel you were you able to show how well you can translate? Do you think it was fair? Please make comparisons with Test One.”

**Appendix 16. Test One: scores and rankings obtained using grading scheme A  
(the intuitive and the points-deduction methods)**

Test taker	Institution	Intuitive score Team 1	Intuitive score Team 2	Points balance Team 1	Points balance Team 2	Team 2 grader
1	Uni-5	15	14	-2.0	-1.0	H
2	Uni-3	11	11	-21.0	-7.5	H
3	Uni-1	10	10	-32.5	-10.5	H
4	Uni-4	9	9	-44.0	-28.0	H
5	Uni-2	11	10	-26.0	-14.0	H
6	IGO	17	15	-16.0	-19.5	B
7	Uni-4	6	7	-59.0	-50.5	B
8	Uni-1	13	12	-3.0	-12.0	B
9	Uni-2	12	11	-17.0	-16.0	B
10	Uni-5	19	15	-7.0	-3.5	B
11	IGO	14	14	-6.0	-6.0	B
12	Uni-5	16	15	-4.0	-23.5	F
13	Uni-3	11	12	-23.0	-48.5	F
14	Uni-4	8	10	-45.5	-53.5	F
15	Uni-4	10	11	-31.0	-39.0	F
16	Uni-5	11	13	-22.5	-25.0	F
17	Uni-4	10	13	-25.5	-23.5	F
18	Uni-1	14	16	-11.5	-19.5	F
19	Uni-5	11	12	-17.0	-29.0	F
20	Uni-1	10	11	-27.5	-40.5	F
21	Uni-1	10	12	-17.0	-29.0	F
22	IGO	14	14	-2.5	-11.0	F
23	Uni-2	14	12	-7.0	-26.0	F
24	Uni-4	10	11	-27.0	-25.0	D
25	Uni-5	13	14	-2.0	-3.0	D
26	Uni-2	12	12	-24.0	-14.5	D
27	Uni-1	18	16	-1.0	-1.5	D
28	IGO	19	17	0.0	1.0	D
29	Uni-2	17	15	-13.0	-13.0	D
30	Uni-2	12	11	-16.5	-19.0	D
31	Uni-1	12	11	-37.0	-38.0	D
32	IGO	17	16	0.0	-2.0	D
33	Uni-2	16	13	-12.0	-21.0	D
34	Uni-4	13	13	-25.5	-13.0	D
35	Uni-3	10	12	-40.0	-20.5	C
36	Uni-1	9	15	-30.5	-12.5	C
37	IGO	19	19	-4.0	-1.0	C
38	Uni-2	11	16	-21.0	-9.5	C
39	Uni-1	12	17	-21.5	-5.0	C
40	IGO	10	16	-23.0	-5.5	C
41	Uni-5	11	14	-12.5	-9.0	C
42	Uni-5	13	15	-20.5	-7.0	C
43	IGO	20	19	0.5	-1.0	C
44	Uni-1	12	16	-15.0	-5.5	C
45	Uni-3	11	14	-30.0	-12.5	C
46	Uni-5	10	13	-21.5	-13.5	C

## Appendix 17. Test Two: scores and rankings, obtained using grading scheme B (multi-trait scoring)

Test taker	Institution	Flow (Grading Team 1)	Clarity 1 (Grading Team 1)	Tone (Grading Team 1)	Source Language (Grading Team 1)	Accuracy (Grading Team 1)	TL vocab (Grading Team 1)	Idiomatcity (Grading Team 1)	Coherence (Grading Team 1)	Correctness (Grading Team 1)	Completeness (Grading Team 1)	Detection (Grading Team 1)	Pass/Fail (Grading Team 1)	Flow (Grading Team 2)	Clarity (Grading Team 2)	Tone (Grading Team 2)	SL (Grading Team 2)	Accuracy (Grading Team 2)	TL vocab (Grading Team 2)	Idiomatcity (Grading Team 2)	Coherence (Grading Team 2)	Correctness (Grading Team 2)	Completeness (Grading Team 2)	Detection (Grading Team 2)	Pass/Fail (Grading Team 2)	Grader from Grading Team 2
4	Uni-4	2	2	3	1	1	2	2	2	3	2	2	f	1	1	1	1	1	1	1	1	3	2	2	f	A
5	Uni-2	4	2	3	3	2	2	3	1	3	3	5	f	3	3	2	3	3	2	3	1	3	3	5	f	A
7	Uni-4	1	1	1	1	1	1	1	1	3	2	f	1	1	1	1	1	1	1	1	1	3	2	f	A	
8	Uni-1	3	3	3	2	3	2	3	1	3	3	5	f	3	3	2	3	3	2	2	1	3	2	5	f	A
9	Uni-2	5	5	3	3	3	3	4	3	5	3	5	p	4	5	3	4	4	3	4	3	4	3	5	p	A
10	Uni-5	3	4	3	5	4	3	4	4	5	3	7	p	4	4	3	5	4	3	4	3	4	3	7	p	A
12	Uni-5	3	3	2	3	3	3	2	3	2	3	4	f	3	3	2	4	3	2	3	2	4	3	4	f	A
13	Uni-3	3	2	2	2	2	2	2	1	3	2	3	f	2	2	2	2	2	1	2	1	3	2	3	f	A
14	Uni-4	2	3	2	2	2	3	2	2	4	3	4	f	2	2	3	2	2	2	2	1	3	3	4	f	A
15	Uni-4	2	3	2	1	1	2	2	2	1	3	5	f	1	2	1	1	1	1	2	1	1	3	5	f	A
16	Uni-5	3	2	2	3	3	2	2	3	2	3	4	f	2	2	1	3	2	1	2	2	2	3	4	f	A
43	IGO	5	4	3	5	5	5	5	4	4	3	6	p	5	5	3	5	5	5	5	5	3	3	6	p	A
17	Uni-4	2	3	2	2	3	2	3	2	3	3	4	f	2	3	1	1	2	1	2	1	2	2	4	f	B
18	Uni-1	3	2	3	3	3	2	1	2	2	3	4	f	2	2	2	2	2	1	1	2	2	3	4	f	B
19	Uni-5	2	2	2	1	1	2	1	2	2	3	5	f	1	1	1	1	1	1	1	1	3	5	f	B	
20	Uni-1	1	1	2	1	2	1	1	2	2	3	2	f	1	1	1	1	1	1	1	1	2	3	2	f	B
21	Uni-1	1	2	2	2	1	2	1	1	3	3	2	f	1	2	2	1	1	1	1	1	2	3	2	f	B
22	IGO	4	2	3	3	2	2	3	2	3	3	4	f	3	2	2	3	1	1	2	2	1	2	4	f	B
24	Uni-4	3	3	3	2	3	2	3	2	3	3	5	f	2	2	3	2	1	2	2	2	2	2	5	f	B
25	Uni-5	4	4	3	3	3	3	4	3	3	3	3	f	3	3	3	3	3	2	3	3	3	3	3	f	B
26	Uni-2	3	2	3	2	1	3	2	1	1	2	4	f	2	2	2	2	1	1	1	1	1	2	4	f	B
27	Uni-1	4	5	3	3	3	3	3	3	3	3	7	p	4	4	3	3	4	3	2	2	4	3	7	p	B
29	Uni-2	2	2	2	2	1	1	1	2	3	3	3	f	1	2	1	2	1	1	1	1	3	3	3	f	B
30	Uni-2	3	2	3	3	2	1	2	2	1	3	6	f	2	2	1	2	1	1	2	1	1	3	6	f	B
31	Uni-1	3	3	1	3	2	3	2	2	2	3	5	f	2	2	2	2	1	4	2	2	2	3	5	f	B
32	IGO	5	5	3	4	3	3	5	5	5	3	6	p	4	4	3	3	4	4	5	4	4	3	6	p	B
33	Uni-2	2	2	2	2	2	3	4	2	2	3	6	f	3	2	2	2	1	2	3	1	2	3	6	f	B
1	Uni-5	4	4	3	3	3	3	4	3	3	2	6	f	5	4	3	3	4	3	2	4	4	3	6	p	G
2	Uni-3	4	3	3	1	5	2	3	3	3	3	4	f	3	3	3	1	1	4	2	4	3	3	4	f	G
3	Uni-1	3	4	3	2	3	2	2	3	3	3	3	f	4	3	2	3	4	3	2	4	3	3	3	f	G
6	IGO	4	4	3	5	3	4	5	5	4	3	7	p	5	4	3	5	2	5	4	5	3	3	7	p	G
11	IGO	5	4	3	3	2	5	4	4	3	2	5	f	4	4	3	3	3	4	4	4	2	2	5	f	G
23	Uni-2	4	4	2	2	3	3	3	4	3	3	6	f	4	4	2	2	2	3	3	4	3	3	6	f	G
34	Uni-4	3	3	3	4	3	3	3	3	3	3	7	f	3	4	2	5	4	4	4	4	3	3	7	p	G
35	Uni-3	4	3	3	3	2	3	2	1	3	3	3	f	3	2	2	2	1	2	1	1	2	3	3	f	G
36	Uni-1	3	3	3	4	4	3	3	2	3	3	5	f	4	3	3	3	2	3	3	2	3	3	5	f	G
37	IGO	4	3	3	3	3	4	3	3	3	3	5	p	4	3	3	2	2	4	4	2	3	3	5	f	G
38	Uni-2	4	3	3	3	3	2	3	3	3	3	4	f	3	3	3	3	3	3	2	2	3	3	4	f	G
39	Uni-1	4	4	3	4	3	3	4	3	3	3	4	f	5	4	3	5	4	5	4	2	4	3	4	p	G
40	IGO	5	5	3	4	3	3	5	3	4	3	4	p	5	5	3	4	2	5	5	3	4	3	4	p	G
41	Uni-5	3	3	3	3	2	3	3	2	3	3	4	f	4	4	3	4	2	3	4	1	3	3	4	f	G
42	Uni-5	4	4	3	5	3	3	2	4	3	3	5	f	4	4	3	4	2	4	2	5	3	3	5	?	G
44	Uni-1	4	4	2	4	3	3	3	3	3	3	5	f	4	4	3	3	3	4	4	2	2	3	5	p	G
45	Uni-3	4	4	3	2	2	3	2	2	3	3	3	f	3	3	3	2	3	4	3	2	4	3	3	p	G
46	Uni-5	3	2	2	3	1	3	3	2	3	2	6	f	3	3	3	2	2	3	3	2	2	2	6	f	G
28	IGO	5	5	3	5	4	4	5	5	5	3	6	p	5	5	3	5	4	4	5	5	5	3	6	p	K

### Appendix 18. Test One: scores obtained from the application of scheme B

Test taker	Institution	Flow (Grading Team 1)	Clarity 1 (Grading Team 1)	Tone (Grading Team 1)	Source Language (Grading Team 1)	Accuracy (Grading Team 1)	TL vocab (Grading Team 1)	Idiomatycity (Grading Team 1)	Coherence (Grading Team 1)	Correctness (Grading Team 1)	Completeness (Grading Team 1)	Pass/Fail (Grading Team 1)	Flow (Grading Team 2)	Clarity (Grading Team 2)	Tone (Grading Team 2)	SL (Grading Team 2)	Accuracy (Grading Team 2)	TL vocab (Grading Team 2)	Idiomatycity (Grading Team 2)	Coherence (Grading Team 2)	Correctness (Grading Team 2)	Completeness (Grading Team 2)	Pass/Fail (Grading Team 2)	Grader from Grading Team 2
32	IGO	5	5	3	5	4	5	5	5	5	3	p	5	5	3	5	5	5	5	5	5	3	p	A
43	IGO	5	5	3	5	5	4	5	5	5	3	p	5	5	3	5	5	4	5	5	4	3	p	A
37	IGO	5	4	3	3	3	5	4	2	3	3	p	4	3	3	3	2	3	3	3	3	3	p	A
33	Uni-2	5	5	3	3	3	4	5	4	3	3	p	4	4	3	3	3	4	4	3	3	2	p	A
36	Uni-1	4	4	3	2	3	3	3	2	4	3	f	3	3	2	3	2	2	2	2	3	3	f	A
40	IGO	4	4	3	4	4	4	5	5	4	3	p	4	4	3	4	4	4	4	5	3	2	p	A
42	Uni-5	3	4	3	4	3	4	3	4	5	3	p	4	4	3	3	4	2	2	5	4	3	p	A
44	Uni-1	3	4	2	3	2	3	3	4	3	3	f	3	3	3	2	2	2	3	4	3	3	f	A
34	Uni-4	3	3	3	3	2	2	4	2	3	3	p	3	3	3	3	3	1	2	4	3	3	f	A
41	Uni-5	3	3	2	3	3	1	2	3	3	3	f	2	3	2	3	3	1	2	3	3	3	f	A
35	Uni-3	3	2	3	2	1	3	2	2	2	3	f	3	2	3	3	2	1	2	2	3	3	f	A
38	Uni-2	3	3	3	3	3	2	3	3	3	2	f	3	3	2	3	3	2	3	3	3	2	f	A
31	Uni-1	3	3	1	2	1	2	2	3	3	2	f	2	3	1	2	1	1	2	3	3	2	f	A
45	Uni-3	3	3	3	3	2	3	2	2	3	1	f	2	3	2	2	2	2	2	2	3	2	f	A
46	Uni-5	3	4	2	2	1	2	3	2	2	2	f	2	3	1	2	1	1	2	2	3	2	f	A
39	Uni-1	2	3	3	4	2	2	2	3	3	3	p	2	3	3	3	2	2	2	3	3	3	f	A
22	IGO	5	5	3	5	3	4	5	5	4	3	p	5	5	3	5	4	5	4	5	5	2	p	B
23	Uni-2	5	4	3	4	3	4	5	4	4	3	p	4	3	3	4	4	4	4	5	4	3	p	B
1	Uni-5	5	4	3	4	4	3	3	4	3	3	p	4	3	3	3	3	1	2	3	2	3	p	B
10	Uni-5	4	4	3	4	3	5	4	5	5	3	p	3	3	3	3	3	4	4	4	5	3	p	B
11	IGO	4	5	3	4	3	4	4	3	4	3	p	4	4	3	4	3	4	4	2	3	3	f	B
15	Uni-4	4	2	2	2	2	2	1	4	3	f	3	1	2	1	1	1	2	1	3	3	f	B	
12	Uni-5	4	4	3	4	4	4	5	4	3	2	p	4	3	3	4	4	3	4	3	3	2	f	B
5	Uni-2	4	3	3	2	1	3	3	1	3	3	f	3	2	2	1	1	2	3	1	2	3	f	B
13	Uni-3	4	3	3	2	1	2	3	2	3	3	f	4	3	2	2	1	1	2	1	2	3	f	B
8	Uni-1	4	2	3	4	2	3	3	4	3	3	p	4	2	2	3	2	2	2	1	1	3	f	B
19	Uni-5	3	3	3	2	1	2	3	2	5	2	f	3	3	2	2	1	2	2	3	5	2	f	B
18	Uni-1	3	3	3	4	3	2	2	3	4	3	p	2	2	2	4	2	1	2	2	4	3	f	B
4	Uni-4	3	2	3	2	1	2	2	1	3	3	f	2	1	2	1	1	4	3	1	3	3	f	B
20	Uni-1	3	3	2	2	2	3	3	2	3	3	f	4	3	2	1	1	2	2	1	3	3	f	B
21	Uni-1	3	3	3	3	2	2	2	3	4	1	f	3	3	2	2	2	2	2	3	3	1	f	B
24	Uni-4	3	3	3	3	2	3	3	3	2	2	f	4	3	2	4	1	2	3	2	3	1	f	B
6	IGO	3	3	3	3	1	4	3	3	4	3	p	3	2	2	2	1	3	3	2	2	3	f	B
3	Uni-1	3	3	3	2	2	3	2	2	3	3	f	2	2	2	1	1	1	2	1	2	3	f	B
16	Uni-5	3	3	2	2	3	3	3	4	2	2	p	4	4	2	3	4	5	4	3	2	2	p	B
2	Uni-3	3	3	3	3	2	4	3	2	3	2	f	2	2	2	2	1	4	2	2	1	2	f	B
9	Uni-2	3	3	2	1	1	1	2	1	2	2	p	3	2	2	1	1	1	2	1	1	2	f	B
17	Uni-4	2	2	2	4	2	1	1	1	1	3	f	2	2	1	1	1	2	2	1	2	3	f	B
7	Uni-4	2	1	1	1	1	1	1	2	2	3	f	2	1	1	1	1	1	2	1	2	3	f	B
14	Uni-4	2	1	2	1	1	2	2	1	2	3	f	1	1	2	1	1	1	1	1	2	3	f	B
28	IGO	5	5	3	4	4	5	5	5	5	3	p	5	5	3	5	4	5	5	5	4	3	p	G
25	Uni-5	5	4	2	4	3	3	4	5	5	3	p	4	4	3	5	4	4	5	5	4	3	p	G
27	Uni-1	5	5	3	4	3	4	4	4	3	3	p	4	5	3	5	4	5	4	5	3	3	p	G
30	Uni-2	5	4	3	2	2	4	4	3	4	3	p	4	4	3	3	3	4	4	2	3	3	f	G
29	Uni-2	4	4	3	3	3	3	4	3	5	3	p	4	4	3	3	4	5	4	3	4	3	f	G
26	Uni-2	3	4	2	3	3	3	4	3	4	3	p	4	3	2	3	4	4	3	4	3	3	f	G

**Appendix 19. Test Two: scores obtained from the application of scheme A**

Test taker	Institution	Intuitive score awarded by Grading Team 1	Intuitive score awarded by Grading Team 2	Points balance Grading Team 1	Points balance Grading Team 2	Grader from Grading Team 2
43	IGO	19	18	-3.5	11	E
40	IGO	15	16	-10	-6	E
37	IGO	15	12	-5	-7	E
44	Uni-1	12	16	-13	-10.5	E
39	Uni-1	11	11	-14	-12	E
45	Uni-3	10	14	-24	-13.5	E
38	Uni-2	9	11	-28	-17.5	E
36	Uni-1	11	12	-15	-19	E
42	Uni-5	12	13	-18	-19.5	E
46	Uni-5	10	12	-31	-24.5	E
41	Uni-5	10	10	-27	-26.5	E
34	Uni-4	12	12	-24	-28.5	E
11	IGO	18	17	2	-3.5	I
6	IGO	18	16	6	-4	I
10	Uni-5	11	16	-15	-4.5	I
8	Uni-1	11	14	-23	-7	I
5	Uni-2	11	11	-19	-9	I
3	Uni-1	10	11	-25	-9	I
2	Uni-3	8	10	-31	-10	I
9	Uni-2	14	8	-11	-11.5	I
1	Uni-5	11	10	-21	-14	I
4	Uni-4	4	6	-44	-16.5	I
7	Uni-4	6	2	-53	-19	I
28	IGO	19	19	5	-0.5	J
25	Uni-5	12	18	-8	-1.5	J
23	Uni-2	13	17	-23	-4	J
27	Uni-1	18	16	-7.5	-4.5	J
32	IGO	17	14	-8.5	-5	J
18	Uni-1	11	15	-19	-5	J
22	IGO	11	15	-13	-6	J
20	Uni-1	11	12	-25	-7	J
12	Uni-5	10	13	-29	-7.5	J
21	Uni-1	8	14	-32	-7.5	J
29	Uni-2	9	11	-41	-7.5	J
13	Uni-3	8	14	-36	-8	J
30	Uni-2	8	12	-37	-8	J
24	Uni-4	12	11	-18	-8.5	J
19	Uni-5	10	13	-31	-8.5	J
15	Uni-4	10	12	-30	-10	J
16	Uni-5	11	10	-20	-10.5	J
17	Uni-4	10	11	-26	-11	J
14	Uni-4	9	10	-29	-11.5	J
26	Uni-2	10	10	-36	-11.5	J
33	Uni-2	10	9	-28	-12	J
31	Uni-1	8	8	-37	-15	J
35	Uni-3	9	10	-41	-30	K

**Appendix 20. Test One and Test Two: mean intuitive scores**

Test taker	Institution	Test One					Test Two						
		Intuitive score Team 1	Intuitive score Team 2	Intuitive score Team 3	Mean intuitive score (20)	Mean intuitive score (100)	Intuitive rank (1-46)	Intuitive score Team 1	Intuitive score Team 2	Intuitive score Team 3	Mean intuitive score (20)	Mean intuitive score (100)	Intuitive rank (1-46)
1	Uni-5	15	14		14.5	72.5	11.5	11	10		10.5	52.5	30.5
2	Uni-3	11	11		11.0	55.0	35.5	8	10		9.0	45.0	43.0
3	Uni-1	10	10		10.0	50.0	43.0	10	11		10.5	52.5	30.5
4	Uni-4	9	9		9.0	45.0	44.5	4	6		5.0	25.0	45.0
5	Uni-2	11	10		10.5	52.5	40.5	11	11		11.0	55.0	25.5
6	IGO	17	15		16.0	80.0	7.5	18	16		17.0	85.0	4.5
7	Uni-4	6	7		6.5	32.5	46.0	6	2	4	4.0	20.0	46.0
8	Uni-1	13	12		12.5	62.5	23.0	11	14	10	11.7	58.3	16.0
9	Uni-2	12	11		11.5	57.5	29.0	14	8	11	11.0	55.0	25.5
10	Uni-5	19	15	18	17.3	86.7	4.0	11	16	12	13.0	65.0	10.0
11	IGO	14	14		14.0	70.0	14.0	18	17		17.5	87.5	3.0
12	Uni-5	16	15		15.5	77.5	9.0	10	13	10	11.0	55.0	25.5
13	Uni-3	11	12		11.5	57.5	29.0	8	14	8	10.0	50.0	36.5
14	Uni-4	8	10		9.0	45.0	44.5	9	10		9.5	47.5	41.0
15	Uni-4	10	11		10.5	52.5	40.5	10	12		11.0	55.0	25.5
16	Uni-5	11	13		12.0	60.0	24.5	11	10		10.5	52.5	30.5
17	Uni-4	10	13	11	11.3	56.7	32.5	10	11		10.5	52.5	30.5
18	Uni-1	14	16		15.0	75.0	10.0	11	15	10	11.7	58.3	17.0
19	Uni-5	11	12		11.5	57.5	29.0	10	13	9	11.5	57.5	20.5
20	Uni-1	10	11		10.5	52.5	40.5	11	12		11.5	57.5	20.5
21	Uni-1	10	12		11.0	55.0	35.5	8	14	9	10.3	51.7	33.0
22	IGO	14	14		14.0	70.0	14.0	11	15	11	12.3	61.7	13.5
23	Uni-2	14	12		13.0	65.0	19.0	13	17	11	13.7	68.3	9.0
24	Uni-4	10	11		10.5	52.5	40.5	12	11		11.5	57.5	20.5
25	Uni-5	13	14		13.5	67.5	17.0	12	18	15	15.0	75.0	8.0
26	Uni-2	12	12		12.0	60.0	24.5	10	10		10.0	50.0	36.5
27	Uni-1	18	16		17.0	85.0	5.0	18	16		17.0	85.0	4.5
28	IGO	19	17		18.0	90.0	3.0	19	19		19.0	95.0	1.0
29	Uni-2	17	15		16.0	80.0	7.5	9	11		10.0	50.0	36.5
30	Uni-2	12	11		11.5	57.5	29.0	8	12	10	10.0	50.0	36.5
31	Uni-1	12	11		11.5	57.5	29.0	8	8		8.0	40.0	44.0
32	IGO	17	16		16.5	82.5	6.0	17	14	17	16.0	80.0	6.0
33	Uni-2	16	13		14.5	72.5	11.5	10	9		9.5	47.5	41.0
34	Uni-4	13	13		13.0	65.0	19.0	12	12		12.0	60.0	15.0
35	Uni-3	10	12		11.0	55.0	35.5	9	10		9.5	47.5	41.0
36	Uni-1	9	15	9	11.0	55.0	35.5	11	12		11.5	57.5	20.5
37	IGO	19	19		19.0	95.0	2.0	15	12	10	12.3	61.7	13.5
38	Uni-2	11	16	11	12.7	63.3	21.5	9	11		10.0	50.0	36.5
39	Uni-1	12	17	12	13.7	68.3	16.0	11	11		11.0	55.0	25.5
40	IGO	10	16	13	13.0	65.0	19.0	15	16		15.5	77.5	7.0
41	Uni-5	11	14	10	11.7	58.3	26.0	10	10		10.0	50.0	36.5
42	Uni-5	13	15		14.0	70.0	14.0	12	13		12.5	62.5	12.0
43	IGO	20	19		19.5	97.5	1.0	19	18		18.5	92.5	2.0
44	Uni-1	12	16	10	12.7	63.3	21.5	12	16	10	12.7	63.3	11.0
45	Uni-3	11	14	9	11.3	56.7	32.5	10	14	11	11.7	58.3	18.0
46	Uni-5	10	13	9	10.7	53.3	38.0	10	12		11.0	55.0	25.5

## Appendix 21. Test One and Test Two: mean component scores

Test taker	Institution	Test One									Test Two											
		Flow (5)	Clarity (5)	Tone (3)	SL knowledge (5)	Accuracy (5)	TL vocab (5)	Idiomatity (5)	Coherence (5)	Correctness (5)	Completeness (3)	Flow (5)	Clarity (5)	Tone (3)	SL knowledge (5)	Accuracy (5)	TL vocab (5)	Idiomatity (5)	Coherence (5)	Correctness (5)	Completeness (3)	Detection (7)
1	Uni-5	4.5	3.5	3	3.5	3.5	2	2.5	3.5	2.5	3	4.5	4	3	3	3.5	3	3	3.5	3.5	2.5	6
2	Uni-3	2.5	2.5	2.5	2.5	1.5	4	2.5	2	2	2	3.5	3	3	1	1	4.5	2	4	3	3	4
3	Uni-1	2.5	2.5	2.5	1.5	1.5	2	2	1.5	2.5	3	3.5	3.5	2.5	2.5	3.5	2	3.5	3	3	2.5	
4	Uni-4	2.5	1.5	2.5	1.5	1	3	2.5	1	3	3	1.5	1.5	2	1	1	1.5	1.5	1.5	3	2	2
5	Uni-2	3.5	2.5	2.5	1.5	1	2.5	3	1	2.5	3	3.5	2.5	2.5	3	2.5	2	3	1	3	3	5
6	IGO	3	2.5	2.5	2.5	1	3.5	3	2.5	3	3	4.5	4	3	5	2.5	4.5	4.5	5	3.5	3	6.5
7	Uni-4	2	1	1	1	1	1	1.5	1.5	2	3	1	1	1	1	1	1	1	1	1	3	2
8	Uni-1	4	2	2.5	3.5	2	2.5	2.5	2.5	2	3	3	3	2.5	2.5	3	2	2.5	1	3	2.5	4.5
9	Uni-2	3	2.5	2	1	1	1	2	1	1.5	2	4.5	5	3	3.5	3.5	3	4	3	4.5	3	5
10	Uni-5	3.5	3.5	3	3.5	3	4.5	4	4.5	5	3	3.5	4	3	5	4	3	4	3.5	4.5	3	6.5
11	IGO	4	4.5	3	4	3	4	4	2.5	3.5	3	4.5	4	3	3	2.5	4.5	4	4	2.5	2	5
12	Uni-5	4	3.5	3	4	4	3.5	4.5	3.5	3	2	3	3	2	3.5	3	2.5	2.5	2.5	3	3	4
13	Uni-3	4	3	2.5	2	1	1.5	2.5	1.5	2.5	3	2.5	2	2	2	1.5	2	1	3	2	3	
14	Uni-4	1.5	1	2	1	1	1.5	1.5	1	2	3	2	2.5	2.5	2	2	2.5	2	1.5	3.5	3	4
15	Uni-4	3.5	1.5	2	1.5	1.5	1.5	2	1	3.5	3	1.5	2.5	1.5	1	1	1.5	2	1.5	1	3	4.5
16	Uni-5	3.5	3.5	2	2.5	3.5	4	3.5	3.5	2	2	2.5	2	1.5	3	2.5	1.5	2	2.5	2	3	4
17	Uni-4	2	2	1.5	2.5	1.5	1.5	1.5	1	1.5	3	2	3	1.5	1.5	2.5	1.5	2.5	1.5	2.5	2.5	3.5
18	Uni-1	2.5	2.5	2.5	4	2.5	1.5	2	2.5	4	3	2.5	2	2.5	2.5	1.5	1	2	2	3	3.5	
19	Uni-5	3	3	2.5	2	1	2	2.5	2.5	5	2	1.5	1.5	1.5	1	1	1.5	1	1.5	1.5	3	5
20	Uni-1	3.5	3	2	1.5	1.5	2.5	2.5	1.5	3	3	1	1	1.5	1	1.5	1	1	1.5	2	3	2
21	Uni-1	3	3	2.5	2.5	2	2	2	3	3.5	1	1	2	2	1.5	1	1.5	1	1	2.5	3	2
22	IGO	5	5	3	5	3.5	4.5	4.5	5	4.5	2.5	3.5	2	2.5	3	1.5	1.5	2.5	2	2	2.5	4
23	Uni-2	4.5	3.5	3	4	3.5	4	4.5	4.5	4	3	4	4	2	2	2.5	3	3	4	3	3	6
24	Uni-4	3.5	3	2.5	3.5	1.5	2.5	3	2.5	2.5	1.5	2.5	2.5	3	2	1.5	2.5	2	2	2.5	2.5	4.5
25	Uni-5	4.5	4	2.5	4.5	3.5	3.5	4.5	5	4.5	3	3.5	3.5	3	3	3	3	3	3	3	3	3
26	Uni-2	3.5	3.5	2	3	3.5	3.5	3.5	3.5	3.5	3	2.5	2	2.5	2	1	2	1.5	1	1	2	4
27	Uni-1	4.5	5	3	4.5	3.5	4.5	4	4.5	3	3	4	4.5	3	3	3.5	3	2.5	2.5	3.5	3	7
28	IGO	5	5	3	4.5	4	5	5	5	4.5	3	5	5	3	5	4	4	5	5	5	3	6
29	Uni-2	4	4	3	3	3.5	4	4	3	4.5	3	1.5	2	1.5	2	1	1	1	1.5	3	3	2.5
30	Uni-2	4.5	4	3	2.5	2.5	4	4	2.5	3.5	3	2.5	2	2	2.5	1.5	1	2	1.5	1	3	6
31	Uni-1	2.5	3	1	2	1	1.5	2	3	3	2	2.5	2.5	1.5	2.5	1.5	3.5	2	2	2	3	5
32	IGO	5	5	3	5	4.5	5	5	5	5	3	4.5	4.5	3	3.5	3.5	3.5	5	4.5	4.5	3	6
33	Uni-2	4.5	4.5	3	3	3	4	4.5	3.5	3	2.5	2.5	2	2	2	1.5	2.5	3.5	1.5	2	3	6
34	Uni-4	3	3	3	3	3	1.5	2	4	2.5	3	3	3.5	2.5	4.5	3.5	3.5	3.5	3.5	3	3	7
35	Uni-3	3	2	3	2.5	1.5	2	2	2	2.5	3	3.5	2.5	2.5	2.5	1.5	2.5	1.5	1	2.5	3	3
36	Uni-1	3.5	3.5	2.5	2.5	2.5	2.5	2.5	2	3.5	3	3.5	3	3	3.5	3	3	3	2	3	3	5
37	IGO	4.5	3.5	3	3	3	3.5	3.5	2.5	3	3	4	3	3	2.5	2.5	3.5	4	2.5	3	3	4.5
38	Uni-2	3	3	2.5	3	3	2	3	3	3	2	3.5	3	3	3	3	3	2	2.5	3	3	3.5
39	Uni-1	2	3	3	3.5	2	2	2	3	3	3	4.5	4	3	4.5	3.5	4	4	2.5	3.5	3	4
40	IGO	4	4	3	4	4	4	4.5	5	3.5	2.5	5	5	3	4	2.5	4	5	3	4	3	4
41	Uni-5	2.5	3	2	3	3	1	2	3	3	3	3.5	3.5	3	3.5	2	3	3.5	1.5	3	3	3.5
42	Uni-5	3.5	4	3	3.5	3.5	3	2.5	4.5	4.5	3	4	4	3	4.5	2.5	3.5	2	4.5	3	3	5
43	IGO	5	5	3	5	5	4	5	5	4.5	3	5	4.5	3	5	5	5	5	4.5	3.5	3	6
44	Uni-1	3	3.5	2.5	2.5	2	2.5	3	4	3	3	4	4	2.5	3.5	3	3.5	3.5	2.5	2.5	3	5
45	Uni-3	2.5	3	2.5	2.5	2	2.5	2	2	3	1.5	3.5	3.5	3	2	2.5	3.5	2.5	2	3.5	3	3
46	Uni-5	2.5	3.5	1.5	2	1	1.5	2.5	2	2.5	2	3	2.5	2.5	2.5	1.5	3	3	2	2.5	2	6

## Appendix 22. Test Two: scores and rankings by weighting of A components and Non-A components

Test taker	Institution	Unweighted scores	Unweighted ranking	Scores Weighted 3:2	Rank weighted 3:2	Scores weighted 3:1	Rank weighted 3:1	Scores weighted 6:1	Rank weighted 6:1
28	IGO	95.1	1.0	95.1	1.0	95.1	1.0	95.0	1.0
43	IGO	94.1	2.0	93.5	2.0	92.2	2.0	90.8	2.0
6	IGO	87.5	3.0	87.9	3.0	88.7	3.0	89.3	3.0
32	IGO	86.9	4.0	87.1	4.0	87.5	4.0	87.7	4.0
10	Uni-5	83.9	5.0	83.5	5.0	82.8	5.0	82.0	5.0
40	IGO	82.5	6.0	81.3	6.0	78.9	7.0	76.5	11.0
9	Uni-2	81.0	7.0	80.6	7.0	79.7	6.0	78.6	6.0
39	Uni-1	78.8	8.0	76.9	9.0	73.0	14.0	69.2	14.0
34	Uni-4	76.7	9.0	77.1	8.0	77.8	8.0	78.4	7.0
42	Uni-5	75.6	10.0	76.2	10.0	77.3	9.0	78.2	8.0
1	Uni-5	75.4	11.5	75.8	12.0	76.5	11.0	77.1	10.0
27	Uni-1	75.4	11.5	76.0	11.0	77.1	10.0	77.9	9.0
11	IGO	74.4	13.0	74.7	13.0	75.3	12.0	75.8	13.0
44	Uni-1	71.3	14.0	70.8	14.5	69.9	15.0	68.8	15.0
37	IGO	69.5	15.0	68.1	16.0	65.5	16.0	62.8	17.0
23	Uni-2	69.3	16.0	70.8	14.5	73.7	13.0	76.4	12.0
36	Uni-1	68.3	17.0	67.0	17.0	64.4	17.0	61.8	18.0
25	Uni-5	67.5	18.0	66.3	18.0	64.0	19.0	61.7	19.0
41	Uni-5	65.4	19.0	63.6	20.0	60.0	22.0	56.4	24.0
38	Uni-2	64.5	20.0	63.2	21.0	60.6	21.0	58.0	23.0
45	Uni-3	63.9	21.0	62.3	23.0	59.3	24.0	56.3	25.0
3	Uni-1	63.5	22.0	63.0	22.0	61.8	20.0	60.6	20.0
2	Uni-3	63.4	23.0	63.7	19.0	64.2	18.0	64.6	16.0
12	Uni-5	62.2	24.0	61.4	24.0	59.9	23.0	58.3	21.0
5	Uni-2	60.4	25.0	58.8	25.0	55.7	26.0	52.7	28.0
46	Uni-5	57.8	26.0	57.9	26.0	58.1	25.0	58.1	22.0
8	Uni-1	57.4	27.0	56.2	27.0	54.1	27.0	51.9	30.0
33	Uni-2	54.8	28.0	54.4	28.0	53.7	29.0	53.0	27.0
14	Uni-4	54.6	29.0	53.5	31.0	51.4	31.0	49.4	33.0
24	Uni-4	54.3	30.0	54.0	29.0	53.3	30.0	52.6	29.0
31	Uni-1	53.8	31.0	53.8	30.0	53.8	28.0	53.7	26.0
22	IGO	53.1	32.0	52.2	32.0	50.5	33.0	48.7	34.0
35	Uni-3	52.4	33.0	50.6	34.0	47.2	37.0	43.8	38.0
16	Uni-5	51.6	34.0	51.3	33.0	50.7	32.0	50.0	32.0
18	Uni-1	50.3	35.0	49.5	35.0	47.8	35.0	46.2	36.0
30	Uni-2	48.4	36.0	48.8	36.0	49.6	34.0	50.3	31.0
17	Uni-4	47.6	37.0	47.5	37.0	47.3	36.0	47.0	35.0
13	Uni-3	45.1	38.0	43.8	38.0	41.3	39.5	38.8	41.0
26	Uni-2	42.5	39.0	42.1	40.0	41.3	39.5	40.4	40.0
15	Uni-4	41.3	40.0	42.1	39.0	43.7	38.0	45.1	37.0
29	Uni-2	40.5	41.0	39.9	41.0	38.7	42.0	37.4	42.0
19	Uni-5	39.2	42.0	39.8	42.0	40.8	41.0	41.8	39.0
21	Uni-1	38.7	43.0	37.6	43.0	35.4	43.0	33.3	43.0
4	Uni-4	37.4	44.0	36.5	44.0	34.6	44.0	32.8	44.0
20	Uni-1	34.4	45.0	33.4	45.0	31.5	45.0	29.6	45.0
7	Uni-4	29.3	46.0	28.5	46.0	27.0	46.0	25.5	46.0



**Appendix 23. Test Two: composite scores obtained by weighting components according to impact and frequency ratings (profile weighting) and by weighting A components and non-A components 3:1**

Test taker	Institution	Profile-adapted scores (100)	Scores weighting A components 3:1 (100)	Profile-adapted rank	Rank using 3:1 weighting
28	IGO	95.0	95.1	1.0	1.0
43	IGO	94.6	92.2	2.0	2.0
6	IGO	87.3	88.7	3.0	3.0
32	IGO	87.0	87.5	4.0	4.0
10	Uni-5	83.5	82.8	5.0	5.0
9	Uni-2	81.0	79.7	7.0	6.0
40	IGO	82.2	78.9	6.0	7.0
34	Uni-4	76.4	77.8	9.0	8.0
42	Uni-5	75.3	77.3	12.0	9.0
27	Uni-1	75.7	77.1	11.0	10.0
1	Uni-5	76.0	76.5	10.0	11.0
11	IGO	75.0	75.3	13.0	12.0
23	Uni-2	69.5	73.7	15.0	13.0
39	Uni-1	78.5	73.0	8.0	14.0
44	Uni-1	71.4	69.9	14.0	15.0
37	IGO	69.4	65.5	16.0	16.0
36	Uni-1	68.0	64.4	17.0	17.0
2	Uni-3	62.6	64.2	23.0	18.0
25	Uni-5	67.3	64.0	18.0	19.0
3	Uni-1	63.5	61.8	21.0	20.0
38	Uni-2	64.2	60.6	20.0	21.0
41	Uni-5	64.8	60.0	19.0	22.0
12	Uni-5	61.6	59.9	24.0	23.0
45	Uni-3	63.2	59.3	22.0	24.0
46	Uni-5	57.7	58.1	26.0	25.0
5	Uni-2	60.2	55.7	25.0	26.0
8	Uni-1	57.3	54.1	27.0	27.0
31	Uni-1	52.8	53.8	32.0	28.0
33	Uni-2	54.4	53.7	28.0	29.0
24	Uni-4	53.9	53.3	29.0	30.0
14	Uni-4	53.3	51.4	30.0	31.0
16	Uni-5	51.3	50.7	34.0	32.0
22	IGO	53.2	50.5	31.0	33.0
30	Uni-2	48.6	49.6	36.0	34.0
18	Uni-1	50.0	47.8	35.0	35.0
17	Uni-4	47.3	47.3	37.0	36.0
35	Uni-3	51.7	47.2	33.0	37.0
15	Uni-4	40.9	43.7	40.0	38.0
13	Uni-3	44.7	41.3	38.0	39.5
26	Uni-2	42.6	41.3	39.0	39.5
19	Uni-5	38.5	40.8	42.0	41.0
29	Uni-2	39.0	38.7	41.0	42.0
21	Uni-1	37.1	35.4	43.0	43.0
4	Uni-4	36.5	34.6	44.0	44.0
20	Uni-1	33.3	31.5	45.0	45.0
7	Uni-4	28.3	27.0	46.0	46.0

**Appendix 24. Test One: composite scores obtained  
 by weighting A and non-A components 3:1**

Test taker	Institution	Score weighting A components 3:1 (100)	Rank based on 3:1 weighting (1-46)
32	IGO	98.44	1.5
43	IGO	98.44	1.5
28	IGO	96.17	3.0
22	IGO	92.69	4.0
27	Uni-1	87.55	5.0
40	IGO	86.34	6.0
25	Uni-5	85.74	7.0
23	Uni-2	84.83	8.0
33	Uni-2	80.89	9.0
10	Uni-5	80.3	10.0
12	Uni-5	79.23	11.0
29	Uni-2	78.94	12.0
11	IGO	77.58	13.0
42	Uni-5	77.12	14.0
30	Uni-2	74.85	15.0
37	IGO	73.04	16.5
1	Uni-5	73.03	16.5
26	Uni-2	70.32	18.0
16	Uni-5	67.44	19.0
44	Uni-1	65.32	20.5
34	Uni-4	65.32	20.5
38	Uni-2	62.45	22.0
36	Uni-1	61.7	23.0
39	Uni-1	58.53	24.5
8	Uni-1	58.52	24.5
24	Uni-4	58.06	26.0
18	Uni-1	57.17	27.0
6	IGO	56.71	28.5
41	Uni-5	56.71	28.5
21	Uni-1	55.49	30.0
19	Uni-5	55.2	31.0
13	Uni-3	53.98	32.0
35	Uni-3	53.08	33.0
20	Uni-1	52.63	34.0
45	Uni-3	52.17	35.0
2	Uni-3	52.02	36.0
5	Uni-2	51.72	37.0
3	Uni-1	48.54	38.0
46	Uni-5	46.57	39.0
4	Uni-4	45.83	40.0
15	Uni-4	45.37	41.0
31	Uni-1	44.76	42.0
9	Uni-2	41.13	43.0
17	Uni-4	38.12	44.0
14	Uni-4	34.03	45.0
7	Uni-4	31.76	46.0