

Unclassified

EDU/IMHE/AHELO/GNE(2008)13/FINAL

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development

22-Nov-2010

English - Or. English

DIRECTORATE FOR EDUCATION
INSTITUTIONAL MANAGEMENT IN HIGHER EDUCATION GOVERNING BOARD

Group of National Experts on the AHELO Feasibility Study

TRANSLATION, ADAPTATION AND VERIFICATION OF TEST MATERIAL IN OECD
INTERNATIONAL SURVEYS

Paris, 17-18 December 2008

This document was prepared by Béatrice Halleux-Monseur. This document is a truncated version of "Anticipating Potential Translation Problems When Writing Items - Memo for Test Developers".

This document is available in PDF format only.

Contact: Karine Tremblay, Directorate for Education
[Tel: +33 1 45 24 91 82; Email: Karine.Tremblay@oecd.org]

JT03292948

Document complet disponible sur OLIS dans son format d'origine
Complete document available on OLIS in its original format



EDU/IMHE/AHELO/GNE(2008)13/FINAL
Unclassified

English - Or. English

TABLE OF CONTENTS

1. Introduction	3
1.1 The OECD/PISA Study and its ERA option	4
1.2 The OECD/PIAAC study	5
2. Item Development – Instrument Preparation	5
2.1 OECD/PISA.....	6
2.2 OECD/PISA 2009/ERA Option	8
2.3 OECD/PIAAC	9
3. Translation and Adaptation Procedures	9
3.1 OECD/PISA.....	10
3.2. OECD/PISA/ERA	13
3.3. OECD/PIAAC	13
4. Verification procedures	13
5. Summary and points for discussion in the AHELO context	15
5.1 AHELO.....	16
REFERENCES	19
APPENDIX 1: ANTICIPATING TRANSLATION PROBLEMS WHEN WRITING ITEMS. MEMO FOR TEST DEVELOPERS.....	21
APPENDIX 2: PISA 2009 TRANSLATION AND ADAPTATIONS GUIDELINES.....	22

TRANSLATION, ADAPTATION AND VERIFICATION OF TEST MATERIAL IN OECD INTERNATIONAL SURVEYS

Béatrice Halleux-Monseur

1. Introduction

1. This working document has been prepared for the first National Experts Meeting for the AHELO feasibility study. It explores and describes translation and adaptation procedures that are of use in some of the OECD International Surveys. It is primarily based on the guidelines elaborated for the PISA National Project Managers, the working documents prepared for the PISA item developers and the PISA Technical Advisory Group, the OECD Technical Report for PISA 2006 (OECD, 2008), the PISA/ERA NPM documents and the PIAAC NPM documents.

2. Literature on empirical comparative research refers to translation issues as one of the most frequent problems in Cross-Cultural surveys. Translation errors are much more frequent than other problems, such as clearly identified discrepancies due to cultural biases or curricular differences (Harkness, Van de Vijver and Mohler, 2003; and Hambleton, Merenda and Spielberger, 2005).

3. Since the early 60's, concerns have been raised as to whether assessment instruments translated in multiple languages and implemented in countries with different cultures were indeed comparable. Criticisms first mostly focused on translation difficulties and equivalence issues for reading assessments while, for mathematics and science assessments, criticisms concentrated more on curricular equivalence issues (except for the adaptation of unfamiliar terms). For questionnaire translation, the only area of concern at that time was the adaptation of the terminology related to school organisation. There was no other specific recommendation for the translation and adaptation of these instruments.

4. This idea that translation and adaptation guidelines would vary according to the type of instruments and their related domains survived through decades. In the late 80's and the early 90's, however, important changes began to occur. The models used for detecting gender or socio-economic DIFs (Differential Item Functioning) in national assessments started to be applied for detecting item/country interactions in multilingual studies, encouraging much more systematic investigation of translation errors. In 1991, the International Test Commission (ITC) led by Ronald K. Hambleton started to work on the development of test translation and adaptation guidelines. In 1994, they produced a set of 22 *Guidelines for Test Adaptation* (Hambleton, 1994). Virtually, all international surveys have since then developed their translation and adaptation procedures on the basis of the ITC guidelines. One of the important outcomes of this work was to demonstrate that linguistic and cultural issues were not limited to reading, but also applied to other domains of assessment such as mathematics and science as well as questionnaires.

5. In a working document that was reviewing the linguistic and cultural appropriateness in international assessments, A. Grisay explains: *"In 1995, an incident took place when France withdrew their data from the International Adult Literacy Study (IALS), claiming that the IALS tests were culturally biased. The hot debate that ensued resulted in extensive additional analyses undertaken by the agencies involved in IALS (ETS and STATCAN) to flesh up their technical report with detailed documentation on all cultural and linguistic points criticized by the French authorities. One of the direct consequences of the*

incident has been to encourage the development of much stricter quality control procedures in the translation and adaptation activities included in all following international surveys – not only the second round of IALS, but also PISA, TIMSS and PIRLS. »

6. If a survey is done merely to rank countries or students, this problem can somewhat be avoided in identifying, then dropping, the most unstable items. The aim of most international surveys, however, is to develop descriptive scales and, in this case, translation errors are of greater concern. The interpretation of a scale can be severely biased by unstable item characteristics from one country to another. One of the important responsibilities of any international survey is therefore to ensure that the instruments used for the assessment provide reliable and fully comparable information.

7. For the purpose of this document, OECD International Surveys that will be considered will be the OECD/PISA programme, its international option, Electronic Reading Assessment (ERA), and, to a lesser extent, the OECD/PIAAC programme.

1.1 The OECD/PISA Study and its ERA option

8. The Programme for International Student Assessment (PISA) is an internationally standardised assessment that was jointly developed by participating countries and administered to 15-year-old students. As explained in various OECD publications, *“PISA assesses how far students near the end of compulsory education have acquired some of the knowledge and skills that are essential for full participation in society. In all cycles, the domains of reading, mathematical and scientific literacy are covered not merely in terms of mastery of the school curriculum, but in terms of important knowledge and skills needed in adult life. ...Pisa focuses on young people abilities to use their knowledge and skills to meet real-life challenges.”*

9. Each 3-year cycle of PISA assesses reading, mathematics and science literacy. PISA 2000 focused on reading literacy (which means that an extensive set of tasks in this domain was used). In PISA 2003, the main focus was on mathematics literacy and on an additional domain, problem solving. In PISA 2006, the emphasis was on science literacy and a “Computer Based Assessment of Science” (CBAS) was proposed as an international option. PISA 2009 will assess again reading as the major domain, and has developed an “Electronic Reading Assessment” (ERA) as an international option.

10. Contextual questionnaires are developed for each cycle. Some of them are compulsory: these are the Student and the School Questionnaires, others are optional: these are the Educational Career, the Information Technology and the Parent Questionnaires. The information gathered through these questionnaires is very wide, from student backgrounds to the school system, through family and institutional factors and student learning attitudes.

11. The Electronic Assessment is included in the PISA 2009 Framework about Reading Literacy. The current draft of this framework recognises that *“While many of the skills required for print and electronic reading are similar, electronic reading demands some new emphases and strategies to be added to the repertoire of readers. Gathering information on the Internet requires skimming and scanning through large amounts of material and immediately evaluating its credibility. Critical thinking, therefore, has become more important than ever in reading literacy* (Halpern, 1989; Shetzer & Warschauer, 2000; Warschauer, 1999). Warschauer concludes that *“overcoming the “digital divide” is not only a matter of achieving online access, but also of enhancing people's abilities to integrate, evaluate and communicate information.”* ERA has thus developed test items in a simulated “Web setting”.

1.2 The OECD/PIAAC study

12. The Programme for the International Assessment of Adult Competencies will assess the level and distribution of adult skills across countries. It will focus on the key cognitive and workplace skills that are required for successful participation in the economy and society of the 21st century. The aim is to assess 16-to-65-year old adults in the domains of literacy and reading components, numeracy, and problem solving in technology-rich environments. It will be administered for the first time in 2011.

13. PIAAC will also gather a range of other information including the antecedents and outcomes of skills, as well as information on usage of information technology and literacy and numeracy practices generally (<http://www.oecd.org>). These background questionnaires will therefore be used for collecting data on the adults' backgrounds that might help to explain the differences in achievement observed among groups of adults within and across countries.

14. The different sections of this document will focus on:

- Item development: the impact of multi-cultural and multi-languages assessments on item development, how it works in some OECD studies.
- Translation and adaptation procedures and guidelines used in PISA, PISA/ERA and PIAAC.
- International verification: aims and rationale.
- Summary of the different issues to be discussed considering the specific settings and timeline of AHELO.

2. Item Development – Instrument Preparation

15. Developing instruments for an international study has constraints that would not exist if the test and/or questionnaire were developed for a national study, in only one language. On top of considering the curriculum and cultural issues that may appear between countries, the fact that these instruments will need to be translated in various languages also influences the way the items should be written and/or prepared.

16. One of the recurring criticisms about international studies used to be that most of material was developed in English only, by people from Anglo-Saxon culture and thus did not represent the different types of cultures in the study. In PISA, PISA/ERA and PIAAC (and other OECD International Studies), substantial efforts have been made to try and ensure that the tests are both suitable for translation and representative of different cultures.

17. The different procedures that were set in order to achieve this goal are summarised in the following paragraphs, for each of the three OECD studies. PISA, PISA/ERA and PIAAC all developed cognitive instruments and their coding guides, contextual questionnaires and manuals designed to reduce potential sources of bias in field operations and test administration procedures. The cognitive instruments consist of “units”, each of them including a stimulus followed by one or more items (questions) related to the stimulus. Depending of the assessment, coding instructions are given either after each of the items inside the unit, then as separated and consolidated coding guides (paper-based assessment) or as separate instructions for each unit in a MS Word document (computer-based assessment).

2.1 OECD/PISA

18. Part of the new test material is prepared by the Consortium test development teams on the basis of the submissions received from participating countries. For example, in PISA 2006, 21 countries submitted items in English or in their national language(s). Prior to this, countries had received described item guidelines to follow when writing these items. The other part of the test material is developed by the development teams themselves. In PISA 2009, the item development teams were based in Australia, Belgium, Denmark, Germany, Japan, and Norway. Each team first developed its items in its own language, tested them in cognitive laboratories before translating them into English. Item were then reviewed and rated by:

- the International PISA Reading, Math or Science Forum;
- the participating countries (the National Centre, the National Expert Groups); and
- the Reading, Math, Science and Questionnaire Expert Groups.

19. The background questionnaires (Student, School, Parent, Educational Career and IT familiarity Questionnaires) are developed by one team only. However, as for the cognitive items, all questionnaire items are reviewed at the national level through different consultations and, at the international level, through the Questionnaire Expert Group. The PISA Governing Board also makes strong recommendations and the OECD secretariat defines what needs to be included/added.

20. The item development teams receive specific information/training about “How to anticipate potential translation and cultural issues”. The document that was prepared for that purpose was mainly based on the experience gained during the previous PISA cycles (see Appendix 1). Issues to be considered are at different levels:

- Syntactic issues.
- Vocabulary issues.
- Graphic issues (mainly dependent on the settings used).
- Cultural issues.

21. One of the particular characteristics of PISA is that it provides two source versions: one in English and one in French. The French version is developed at an early stage through a double-translation of the English material, followed by a reconciliation (made in two steps by a translator specialised in educational assessment and by the translation referee) so that any comments from the translation team could, along with the comments received from the Expert Groups and the NPMs, be used in the finalisation of both source versions. Developing a second source version has quite a number of advantages:

- Early detection of potential translation problems.
- Detection of residual errors overlooked by test developers.
- Refining of list adaptations to be recommended.
- Addition of translation notes for better guidance of national translators.

22. In this respect, the development of the French source version is like a “pilot translation”, and contributes to provide National Centres with source material that is somewhat easier to translate or contains fewer potential translation traps than it would have had if only one source had been developed. In order to make the French version as equivalent as possible to the English version and therefore to have the “source” status, additional steps are needed: the French version is independently reviewed by a French domain expert, a professional French proof-reader and a specialist who is experienced in equivalence issues. After each review, the translation referee discusses the issues raised with the item developers and, most often, both English and French versions are further amended.

23. Between the Field Trial and the Main Survey, further analyses on the English and French versions are made in order to help with the review of both source versions. The DIFs (Differential Item Functioning) that are identified helps the French and English source development teams to locate potential lack of equivalence between the French and the English source versions.

24. It is important to note that the development of a second source version has a non-negligible impact on the development costs. Additionally, the complex process requires the involvement of experts in translation, education, linguistic, equivalence issues and particular subjects and may render the timeline a bit longer. However when considering the benefits of a second source version, the impact on the timeline may seem pretty trivial.

25. The translation notes that are added at the multiple steps of the process are of very high importance. They give specific guidance on how to accurately translate stems, stimuli or expressions so that they stay equivalent to the source versions. They can also draw the attention of translators on possible translations traps, or on how to adapt specific parts of the text.

26. The few examples given below show different kinds of translation notes that may need to be added:

- Focusing on translation and adaptation guidelines: “Part B of this text is likely to be available in published translation version in many languages. However, the source version has been heavily adapted. Please use a published version to support your translation where possible (however, observing PISA “rules” about translation of literary texts outlined in Translation and Adaptation Guidelines PISA 2009)”.
- Focusing on equivalence:
 - “Translate “Panic-stricken and frantic” to connote the ideas of being struck by uncontrollable fear, and being hurried/chaotic.”
 - “Translate “Excited and impatient” to connote the idea of being keyed up, and restlessly waiting for something to happen.”
 - “Translate “Angry and frustrated” to connote being upset by one’s inability to achieve something.”
 - “Translate “anxious and oppressed” to connote the ideas of being worried and weighed down. Do not use the same word as “apprehension” (last paragraph of stimulus).
 - “Translate “Relaxed and confident” to connote feeling comfortable and sure of oneself.”
- Focusing on needed adaptations:
 - “Adapt to the local currency”.
 - “Adapt “pounds” to kilograms if necessary.”
 - “Adapt “metres” if necessary.”
 - “Adapt the phone number to a fictitious, but plausible toll free phone number for your country.”
 - “Change the website URL to your country’s country code.”

- Focusing on vocabulary issues:
““Casual” means a job that is temporary or irregular.”
“The scientific name of “wild oat grass” is *Avena fatua*.”
“All scientific names should be translated in using the terminology used in the day to day life context”
“Please make sure to use the strict scientific terminology as used in the medical context.”
- Focusing on layout/graphic issues: “Check that the graphic has not been resized. Map’s dimensions on the printed copy: 11 cm by 15 cm. Scale line: 5 cm.”

27. In PISA 2009, the development of the new reading material lasted 17 months, starting in July 2006 and “ending” with the release to the National Project Managers in Mid and End-November 2007. The preparation of the French source version began when the first bundle of units was submitted for Country Review in March 2007 and it was released to countries at the same time as the English source version in November 2007; which sums up the French source version preparation to 8 months.

2.2 OECD/PISA 2009/ERA Option

28. The item development teams for ERA were the same than for the Main PISA 2009 development. National submissions were invited but only one unit came from a national submission.

29. As explained in the introduction, ERA is a computer based assessment that uses simulation of Website. The first step was thus to find an appropriate delivery platform. The two following paragraphs are giving specific technical details in order to give an idea of the complexity and extent of the preparation work.

30. The TAO¹ computer-based assessment platform, developed by the *Centre de Recherche Public (CRP) Henri Tudor* and the *Université du Luxembourg*, was used to deliver the electronic reading assessment. TAO is a modular and versatile platform for collaborative and distributed computer-based assessment delivery and management. It is an “open-source” development. The existing TAO was adapted for use in ERA. The time needed for adapting the platform and to make it operational was of about 18 months

31. The test was delivered by booting each computer from a CD or an USB stick. The test was bundled together with a Linux-based operating system, a beta release of the TAO assessment platform and a Mozilla Firefox web browser for viewing the TAO interface. A licensed version of Flash media player was also bundled with the other software to enable the viewing of the stimulus material. Item template incorporating simulated browser was designed and software was developed to allow items to be reviewed by the participating countries and feedback to be collected online.

32. Because of time constraints and mostly of cost implications, it was decided to only provide an English source version of the ERA material. As the item development teams were identical as for the Main PISA, they had been trained on potential translation problems and used the same translation and adaptation guidelines as references. For the translation notes, it was not desirable (nor possible) to include them in the material itself because of the IT design. They were thus included on the specific coding guide provided in Word format for each unit. Indeed part of the student answers were automatically scored for closed constructed responses (like QCM) while other needed manual coding.

¹ TAO is the French acronym for “Testing Assisté par Ordinateur”.

33. The item preparation process was complex, as the software was being developed while the items were being written, which resulted in a process of continuous negotiation between the test developers, the layout designers and the software designers. These negotiations focused not only on how the stimuli could be presented, but also on which item types and response formats were feasible. The process was further complicated by the fact that 2 systems were involved - one was the authoring software, where stimulus material was presented (and where test takers responded to a small proportion of the items); the other was the test delivery and data capture system, including such features as the area where the task (question) was presented (and where test takers responded to the majority of the items) and the area of the screen showing progress through the test. Making these two systems compatible with each other was complex and time consuming, involving IT staff in 3 countries. Item development began in July 2006 and the item released occurred in November and December 2007, so the development process took slightly longer than for the paper-based assessment of PISA but for a very much smaller item pool. Overall it is estimated that the development of each electronic reading item took about three times as long as each print item.

2.3 OECD/PIAAC

34. One of the particularities of PIAAC is that it is implemented in both paper and computer format. Not all instruments are developed in both formats, for example the problem-solving test is only developed for the computer-based assessment and the reading test for the paper-based assessment only.

35. The source version is US English. The items developers receive training on potential translation problems to be avoided and on specific item notes to provide. As some units and items are indeed common for both paper-based and computer-based assessments, detailed notes are provided for each item and stimulus.

36. Variations in wording between the computer-based and the paper-based tests are necessary because of the particular settings and the way answers will be made. For example where a question would ask in the paper-based version to retrieve and underline specific information in the stimulus, the computer-based version will ask to highlight it; or when a paper-based would ask to “refer to”, the computer version would ask to “look at”. Translation notes (or item notes) that are common to both will be mostly similar as for PISA but systematically provided for each stimulus and item.

37. The paper-based item development is, as in PISA, developed in Word format while the computer-based assessment needs (as for ERA) a specific platform. PIAAC uses also the TAO platform, as in PISA/ERA. The timeline devoted for adapting the TAO to PIAAC and doing the integration and packaging work necessary for the field trial is currently of 21 months: 12 months for building the unique functionality for PIAAC (to support the item types and response modes required by PIAAC) and the Item Management Portal required to support the translation and adaptation of the instruments; 3 months for developing the virtual machine for integrating TAO with the countries’ laptops and producing country specific versions, 6 months for working with countries to test the software and integrate it into their systems. PIAAC is using virtual machine technology to deliver the tests and also uses a freely available software package (VMWareServer).

3. Translation and Adaptation Procedures

38. A *back translation* design has been for a long time the most frequently used procedure to ensure linguistic equivalence of test instruments in international surveys. It requires translating the source version of the test (generally English language) into the national languages, then translating them back to English and comparing them with the source language to identify possible discrepancies.

39. A *double translation* design (*i.e.* two independent translations from the source language(s), and reconciliation by a third person) offers two significant advantages in comparison with the back translation design:

- Equivalence of the source and target versions is obtained by using three different people (two translators and a reconciler) who all work on both the source and the target versions. In a back translation design, by contrast, the first translator is the only one to simultaneously use the source and target versions.
- Discrepancies are recorded directly in the target language instead of in the source language, as would be the case in a back translation design.

40. Reconciliation consists of “merging” the two independent translations in such a way that the resulting national version is as equivalent as possible to the source versions, that all possible translation errors have been corrected; and that the wording is as fluent as possible – without the “translinese flavour” that is so often characteristic to translated material. The aim is to get the right balance: the reconciled version must not be literal to the point that it sounds awkward, but neither should it deviate too freely from the source version, which would be likely to change the functioning of the assessment items in unexpected ways. The person in charge on the reconciliation, so called the reconciler, will also need to document all adaptations made in the material.

41. The boundary between “translation” and “adaptation” is very thin. Ronald K. Hambleton explains that the term “test adaptation” should be preferred to “test translation”: *“Test adaptation includes all the activities from deciding whether or not a test could measure the same construct in a different language or culture, to selecting translators, to deciding on appropriate accommodations to be made in preparing a test for use in a second language, to adapting the test and checking its equivalence in the adapted form. Test translation is only one of the steps in the process of test adaptation and even at this step, adaptation is often a more suitable term than translation to describe the actual process that takes place. This is because translators are trying to find concepts, words and expressions that are culturally, psychologically, and linguistically equivalent in a second language and culture, and so clearly the task goes well beyond simply preparing a literal translation of the test content.”*

42. For the purpose of this document, the term “adaptations” will however refer to

- intentional deviation from the source version that the national centres deem necessary to make the item, sentence, question suitable in their national context or culture;
- sections that need to be adapted to the national context and that are mentioned or required by the item developers.

43. In each of the three OECD surveys, the adaptations that are made by the translators/reconciler need to be recorded in specific adaptation forms (Excel files). The double translation and reconciliation procedure is also used in PISA, ERA option of PISA and PIAAC to different extents. The similarities and differences between these three surveys are summarised in the following paragraphs. Note that the field operation manuals do not need to be double-translated.

3.1 OECD/PISA

44. PISA uses double translation *from two different source languages* because both back translation and double translation designs fall short in that the equivalence of the various national versions depends exclusively on their consistency with a single source version (in general English). Using a single reference

language is likely to give undue importance to the *formal* characteristics of that language. If a single source language is used, its lexical and syntactic features, stylistic conventions and the typical patterns it uses to organise ideas within the sentence will have a greater impact on the target language versions than desirable (Grisay, 2003). The second source version thus often provides to translators alternative ways of translating while staying equivalent to the first source version. In some aspects, it also gives the extent to which the national version can be adapted.

45. In PISA, it is *required* that the national versions of all test and questionnaire instruments used in the assessment be developed through a *double translation and reconciliation* procedure. It is *recommended* that the countries use the English source version for one of the translations into the national language and the French source version for the other one. An alternative (seemingly efficient) method is to perform double translation and reconciliation from one of the source languages, followed by extensive cross-checks against the second source language. When countries can't find competent translators for one of the source languages, they will perform a double translation from only one source version.

46. Countries sharing a testing language are strongly encouraged to develop a common version in which national adaptations would be inserted or, in the case of minority languages, to borrow an existing verified version.

47. Countries are asked to fill in and submit a translation plan documenting their translation procedures. These plans are reviewed and agreed upon by the consortium. The following table summarises the translation procedures as described in the country translation plans for the PISA 2009 Field Trial and presents the number of countries using a certain procedure for the unit translation and for the questionnaire translation.

Translation procedures PISA 2009

	Unit translation	Questionnaire translation
Double translation from English and French source versions	29	24
Double translation from English source version with cross-checks against the FRE source version	8	11
Double translation from French source version with cross-checks against the ENG source version	1	1
Double translation from English source version only	15	17
Adaptations in one of the source versions	17	17
Adaptation made in a borrowed verified version	12	12
Alternative procedure	3	3

48. It is important to note that, from one PISA cycle to another, the number of countries that use both source versions is increasing. Note also that the row referring to adaptations in one of the source versions refers to countries that are testing in English or French and that only need to adapt the source version to their national context.

49. Before the dispatch of the source versions and the beginning of the national translation process, National Project Managers receive detailed translation and adaptation guidelines in both French and English (see Appendix 2).

50. The translation guidelines include:

- *PISA requirements in terms of necessary national version(s)*. According to one of the PISA Standards, students should be tested in the language of instruction used in their school. Therefore, the NPMs of multilingual countries are requested to develop as many versions of the test instruments as there are languages of instruction used in the schools included in their national sample.
- *Information on which parts of the materials had to be double translated, or could be single translated*. Double-translation is required for the tests, questionnaires and for the optional Questionnaires, but not for the manuals and other logistic material (coding workshop documents, item submission guidelines, etc.).
- *Instructions related to the recruitment and training of translators and reconcilers* and the scientific and technical support to be provided to the translation team. It was suggested, in particular, that translated material and national adaptations deemed necessary be submitted for review and approval to a national expert panel composed of domain specialists.
- Description of the PISA translation procedures. Translation and adaptation of Questionnaires.

51. Other sections of the *PISA Translation and Adaptations Guidelines* are intended for use by the national translators and reconciler(s):

- recommendations to avoid common translation traps—a rather extensive section giving detailed examples on problems frequently encountered when translating assessment materials, and advice on how to avoid them;
- instructions on how to adapt the test material to the national context, listing a variety of rules identifying acceptable/unacceptable national adaptations and including specific notes on translating mathematics and science material;
- instructions on how to translate and adapt the questionnaires and manuals to the national context; and
- the check list used for the verification of PISA material.

52. These guidelines are of crucial importance for National Centres as they cover all aspects and steps to be taken in the translation and adaptation of the instruments. (Please refer to Appendix 2 for detailed explanations on all aspects mentioned above.)

53. NPMs also receive sample materials to use when recruiting national translators and training them at the national level. Prior to the PISA 2009 Field Trial, countries were invited to participate to a one-day training on translation and adaptation during one of the NPM meetings.

54. All information released by the PISA consortium is placed on a secure website (MyPISA). This website allows downloads and online submission of various forms and documents.

3.2. OECD/PISA/ERA

55. For the PISA ERA option, as only one source version is developed, the required procedure is a double translation and reconciliation from the English source version. Because of the electronic format, the translation process is considerably different from that in “main” PISA. The translation, reconciliation and verification of ERA units (including coding guides) are managed via an online secure Translation Management System (TMS). From the TMS, translators are able to download the source version of stimuli and items as text files in XML Localization Interchange File Format v 1.1, (XLIFF 1.1). Coding guides are downloaded as Word documents at each stage of the process. The XLIFF text files can be translated into the target language using any commercial or open source translation editor that handles XLIFF 1.1. Each step of the process needs to be included in the Translation Management System (TMS): first and second independent translations, reconciliation, adaptation review, verification, format and layout checks. Note that the design and development of the TMS system was a complex process that took about 6 months.

56. As an international option of PISA, the translation and adaptations guidelines to which National Centres should refer are the same as in the main PISA. Additionally countries receive a specific document that details the adaptation, translation and verification process using the TMS, all interfaces inside this system, system requirements, information about translation editor and the structure of the ERA XLIFF files.

57. It was therefore recommended to the countries to find competent translators that are used to work with such translator editors or to train their PISA translators with the information provided.

3.3. OECD/PIAAC

58. PIAAC differs slightly from PISA because the double-translation and reconciliation process is a recommendation and not a requirement.

59. However, as in PISA, PIAAC provides precise translation and adaptation guidelines (primarily based on PISA guidelines) with a few additional specifications linked to the different testing domains and the two distinct formats of assessment. An online secure system (Item Management Portal) will be used for both paper and computer instruments. It will serve as the primary mean for the distribution of documents as well as the tool for capturing the translated forms for verification.

4. Verification procedures

60. One of the most important quality control procedures implemented to ensure high quality standards in the translated assessment materials consists in having an independent team of expert verifiers, appointed and trained by the Consortium, that verify each national version against the source version(s).

61. In PISA, ERA and PIAAC, international verification is organised similarly even though there are differences in the processes and/or ways of revising the adaptation forms and annotating them. However, the final aim surely stays the same in all of the studies: assuring that the instruments are strictly equivalent to the source version(s).

62. Verifiers are recruited on the following criteria:

- native command of the target language;
- professional experience as translators from the source version language (or one of the two in PISA) into their target language;

- in PISA: sufficient command of the second source language to be able to use it for cross-checks;
- familiarity with the domain of assessment;
- good IT skills; and
- as far as possible, teaching experience or higher educational degrees in psychology, sociology or education

63. Verifiers are trained by the Consortium before the field trial and again before the main study. During these trainings they receive specific information on the study aims, examples of issues overlooked or inadequately addressed in previous verification rounds and a tutorial on comment writing. The team of verification coordinators presents the materials to be verified with an extensive review of the different procedures and of the material key features. Verifiers are also trained with practical exercises in which a number of different kinds of errors are included intentionally. They then need to try and identify what these issues are and to record their finding in the appropriate described way that is specific to each of the studies.

64. As explained previously, countries need to enter their national adaptations in an adaptation spreadsheet. In PISA, ERA and PIAAC, these files are of most importance. They will contain all information collected during the different steps of the verification process.

65. In PISA, for example, these steps are:

- Review of the material by the verifier in the Word document using track changes mode for suggesting the corrections. All corrections that might have an effect on equivalence are recorded by the verifier in the adaptation spreadsheet and are back-translated into English for review by the translation/verification referee. Each verifier has to choose an intervention category corresponding to the described correction.
- Review of the adaptation spreadsheet by the consortium referee. When the national adaptation asked for by the National Centre is controversial or when the translation referee deems that it is necessary for the National Centre to follow the verifier advice, the issue is labelled as key correction.
- In case of disagreement with one or more of the key corrections, the National Centre contacts the Consortium referee and discusses the issue until an agreement is reached.
- Final check of booklets, coding guides and questionnaires. This final check includes two different steps: (i) an “optical” check where the reviewers (separately trained) focus on layout issues such as graphic rendering, question allocations, footers, item numbering, etc. and (ii) the systematic verification of the implementation of the key corrections.

66. Note that in a computer-based assessment, the final check that deals with “on screen” issues, the necessity to solve potential technical issues, the preparation of the national CD and USB stick will need to be taken into account in the verification timeline.

67. In PISA, the adaptations made in the cognitive instruments are first reviewed by the verifiers, then by the consortium translation referee. For the questionnaires and the field operation manuals, though, the National Centres first need to get their adaptations approved by the development team including experts in educational systems. Adaptations to this material are indeed very specific and National Centres need, for example, to submit their adaptations about their school learning programmes, their national language(s), their exclusion type definitions, etc.

5. Summary and points for discussion in the AHELO context

68. In the intent to help with the discussion in the AHELO context, the following table summarises the information given in this document for the three OECD studies.

Brief summary of procedures in PISA, ERA and PIAAC for the cognitive instruments

	OECD/PISA	OECD/PISA/ERA	OECD/PIAAC
Item development	<p>Consortium teams from different languages and cultures</p> <p>National submissions</p> <p>Training of test developers</p> <p>Translation notes inserted in units</p> <p>Material developed in MSWord</p> <p>Two source versions in UK English and French</p>	<p>Consortium teams from different languages and cultures</p> <p>National submissions</p> <p>Training of test developers</p> <p>Translation notes at the beginning of each coding guide (by units)</p> <p>Electronic assessment requiring a specific platform</p> <p>One source version in UK English</p>	<p>Consortium teams in ENG only</p> <p>National submissions</p> <p>Training of test developers</p> <p>Specific detailed item by item notes</p> <p>Both MSWord and specific Platform</p> <p>One source version in US English</p>
Translation Procedures	<p>Double-translation/reconciliation is required.</p> <p>Translation/Adaptation/Verification guidelines and translator kit for trainings</p>	<p>Double-translation/reconciliation is required</p> <p>Translation/ Adaptation/ Verification guidelines</p> <p>Electronic assessment requires an online translation management system</p>	<p>Double-translation/reconciliation is recommended</p> <p>Translation/Adaptation/Verification guidelines and translator module for training</p> <p>Computer-based part of assessment requires an online management system</p>
Verification procedures	<p>International verification</p>	<p>International verification with a process that is specific to the IT environment (using the Translation Management System)</p>	<p>International verification. The process is adapted depending on the assessment type (paper-based or computer-based assessment).</p>
Particular Timing	<p><i>Item development:</i> From the beginning of the development till the dispatch to the NPMs: 17 months</p> <p><i>Translation and verification:</i> Dispatch to countries at the mid and end of November with early countries beginning testing on the 1st of March → a minimum of 3 months. Note that this might vary according to the volume.</p>	<p><i>Item development including set up of TAO:</i> Item development itself: 19 months TAO adaptation: 18 months</p> <p><i>Translation and Verification:</i> Dispatch to countries at the end of November with early countries beginning testing on the 1st of March Development of the TMS: 6 months</p>	<p><i>Item development including set up of TAO:</i> Item development not yet completed. Estimated time for TAO: 21 months</p> <p><i>Translation and Verification:</i> Process not yet completed. Development of the online system: no final estimation yet</p>

5.1 AHELO

69. When discussing procedures for the AHELO study, it is likely that the timeline will be the crucial issue.

70. At the time of this document preparation, the expected timeline for AHELO Field preparation was:

- Call for tender: first quarter of 2009
- Adjudication of contractors: third quarter of 2009
- Translations and cultural adaptations: fourth quarter of 2009
- Final instruments and preparation of the computer platform first quarter of 2010
- Field implementation – Northern hemisphere countries: second quarter of 2010 – April May and Southern hemisphere countries: third quarter of 2010

71. In the following paragraphs, potential aspects to be discussed appear in italics.

72. *In northern hemisphere countries, is it indeed feasible to organise the testing in April/ May at a time when most students are fully busy studying and preparing for their end-of- academic-year exams?*

73. *From the experience in PISA, PISA/ERA and PIAAC, translation and cultural adaptations should come after the final instrument release.*

Item development: (including cognitive and questionnaire instruments)

What would be the most appropriate: computer-based or paper-based assessment?

74. Note that it may be envisaged, for example, to implement a computer based assessment for the cognitive instruments and a paper-based assessment for the questionnaires; or according to the domain of the assessment.

75. The number of students that needs to be tested within the same university, school may render a computer-based assessment difficult to implement. If, for example, about 200 or 300 students had to be tested within a particular university, then a computer-based assessment could become very problematic.

If it is preferred to have a computer-based assessment, how will this fit with the initially devoted timeline?

76. Indeed, the last row of the table shows that the time needed for the preparation of the instruments (including platform preparation) varies from 17 up to 21 months.

77. ERA preparation took 18 months for the adaptation and testing of the TAO. Even though TAO became an Open System, the adaptation for use in PIAAC will take about 21 months. Opting for a computer-based assessment will at least require starting the preparation of the computer platform as early as possible (not waiting for the first quarter of 2010). Costs involved in the development and/or adaptation of such platform are without any doubts (much) higher than if the items were developed for a paper-based assessment. Note also that, for PISA, if the development of the new items for paper-based assessment took as long as the development of the items for ERA, the volume developed was at least three times more. However, as ERA was the first to use and adapt the TAO platform it is likely that the experience gained at the time of the item development may be beneficial for other studies that would use it.

Use of existing materials or development of new material? How to review the existing material for use at the international level?

78. Even though no new material would be developed, it would be needed to: (i) select it carefully with the input of national expert panels, international expert groups, etc.; (ii) review, adapt the selected material to make it suitable for translation into various languages.

79. Such selection and adaptation may not require as long as developing new material but the timeline should be carefully set to allow enough time for this. (at least 6 to 9 months).

Source version: in English only? Would it be possible to consider the development of a second source version or a reference version?

80. The development of a second source version takes time but is feasible within the timeline of item development if the process begins early. Alternatively, a “reference version” can be prepared. A reference version is a version that is developed through double-translation and reconciliation but that does not go through all additional steps required for a source version. Therefore it cannot be used as a basis for translation into other languages but it will help identifying translation problems during item development. Both source and reference versions do have implications on the overall costs.

Translation, adaptation procedures

81. All current international surveys in education are using the double-translation and reconciliation procedures to some various extents. However, as AHELO is a feasibility study, it would be important to first define what the prime priorities are. If the aim of AHELO is to prove that higher education students can be tested, but that it does explore the possibility of building unidimensional scales, the double-translation and reconciliation procedure might not be as important at that stage. But if the feasibility study wants to examine whether university outcomes can be compared in meaningful ways across countries then double-translation and reconciliation should definitely be used.

For AHELO, what procedures would be recommended, or required?

82. Note that double-translation procedures could be a requirement for units, items and questionnaires and a single translation only could be allowed for coding guides.

Consider the impact on timing if the preference goes to a computer-based assessment as this would need special online translation management system

83. At the national level, it may be possible that the translation will be more costly for computer-based assessment than for paper-based assessment. Translators may indeed ask for higher translation fees if they have to manage more technical files and translation software.

Verification

84. International verification is a necessary quality assurance step that should be organised.

If the preference goes to a computer-based assessment the impact on the timeline will need to be considered as verification would also need to be part of translation management system.

85. Note also that the PISA/ ERA verification costs (based on number of characters to be verified) is higher than verification costs for the paper-based assessment of PISA because of the technical aspects involved, such as the IT skills needed and the management of more sophisticated files and software.

REFERENCES

- Grisay, A. (2002). Translation and cultural appropriateness of the test and survey material. In Ray Adams and Margaret Wu (Eds.), *PISA 2000 Technical Report* (pp. 57-70). Paris: Organization for Economic Co-operation and Development.
- Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment, *Language Testing*, 20 (2), 225-240.
- Grisay, A., de Jong, J.H.A.L., Gebhardt, E., Berezner, A., & Halleux-Monseur, B. (2007) Translation Equivalence across PISA Countries. *Journal of Applied Measurement* 8(3) 2007, 249-266.
- Hambleton, R.K., Merenda, P.F., Spielberger, C.D. (2005), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*, IEA Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.
- Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229-224.
- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45, 153-171.
- Hambleton, R.K. (2002). Adapting achievement tests into multiple languages for international assessments. In Porter, A.C. and Gamoran, A., editors, *Methodological advances in cross national surveys of educational achievement*. Washington DC: National Academy Press.
- Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, 1, 1-30.
- Harkness, J.A., Van de Vijver, F.J.R., Mohler, P.Ph (2003), *Cross-Cultural Survey Methods*, Wiley-Interscience, John Wiley & Sons, Inc., Hoboken, New Jersey.
- OECD PISA 2006 Technical Report, *Chapter 5. Translation and Cultural Appropriateness of the Test and Survey Material*
- PIAAC_NPM(2008_10_12):Translation_Tool: PIAAC Translation Tools and Procedures
- PIAAC_NPM(2008_10_13):Translation_guidelines_DA: PIAAC Translation Guidelines for the Direct Assessments
- PISA_TAG(0505)10_1: Linguistic and cultural appropriateness in international assessments: a review
- PISA_NPM(0702)12_1: Electronic Reading Assessment
- PISA_NPM(0709)2_3: Electronic Reading Assessment: Update and Field Trial Arrangements

EDU/IMHE/AHELO/GNE(2008)13/FINAL

PISA_NPM(0709)7b: PISA 2009 Reading Literacy Framework (draft)

PISA_ERA_TMS_FT09_2: Translation Management System version 2:ERA_TMS_FT09_2

PISA_NPM (0809): Translation and Verification Report Field Trail PISA 2009

PISA_NPM(0709)1_ENG: Translation, Adaptation and Verification Guidelines

**APPENDIX 1:
ANTICIPATING TRANSLATION PROBLEMS WHEN WRITING ITEMS. MEMO FOR TEST
DEVELOPERS**

ANTICIPATING POTENTIAL TRANSLATION PROBLEMS WHEN WRITING ITEMS

Memo for test developers

Béatrice Halleux

Andrea Ferrari

Aletta Grisay



Project Consortium:

Australian Council *for* Educational Research (ACER)

cApStAn Linguistic Quality Control

Service de Pédagogie théorique et expérimentale (SPe)

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

National Institute for Educational Policy Research (NIER, Japan)

Westat

Anticipating potential translation problems when writing items

One of the important responsibilities of the PISA consortium is to ensure that the instruments used in all participating countries to assess their students' literacy provide reliable and fully comparable information. Translation and adaptation of test instruments are of high importance in cross-national studies. *“Unless the translations (or, more correctly, test adaptations) are carried out well (...), the results of an international comparative study such as TIMSS or OECD/PISA may be confounded by the consequences of poorly translated assessment materials. To the extent that a test adaptation changes the psychological meaning and/or test difficulty in the target languages or cultures, comparisons of student performances across languages and cultural groups may have limited validity”* (Hambleton, 2002)

The translation procedure used in the first three PISA cycles included the following components:

- Translation guidelines
- Two equivalent source versions of all assessment instruments (in English and French)
- Double translation and reconciliation
- National adaptation forms
- Verification procedures

As we have progressed from PISA 2000 to PISA 2006, and now PISA 2009, procedures and guidelines have been refined and potential issues better determined. Items statistics also pointed out some translation flaws.

This paper, based on the experience gained over the first three PISA cycles, lists a number of recommendations that may help test developers in any PISA domain to anticipate potential translation problems when writing items.

A. Syntactic issues

Multiple-choice items in the incomplete stem format (truncated sentence) are difficult to translate.

“Such items are difficult to translate because the organisation of subject, verb and object in sentences is not consistent across languages. In countries such as Turkey, use of the incomplete stem format meant placing the blanks at the beginning of sentences rather than the end, and revising answer choices to match the format changes, and these changes could have influenced the difficulty of test items.” (Hambleton, 2002).

Example: Note that in the MS PISA 2000 reading item pool, 21 MC items were in the incomplete stem format.

R081 Graffiti

The purpose of each of these letters is to

- A explain what graffiti is.*
- B present an opinion about graffiti.*
- C demonstrate the popularity of graffiti.*
- D tell people how much is spent removing graffiti.*

It would have been preferable to reformulate the question as, for example:

What is the purpose of each of these letters?

- A To explain what graffiti is.*
- B To present an opinion about graffiti.*
- C To demonstrate the popularity of graffiti.*
- D To tell people how much is spent removing graffiti.*

Sometimes questions that have been reworded to avoid truncated sentences may actually behave as truncated items.

The following example comes from the PISA 2003 school questionnaire.

The original question was:

“Is your school’s capacity to provide instruction affected by a shortage or inadequacy of:”

The reworded question was:

Q8 Is your school’s capacity to provide instruction affected by a shortage or inadequacy of any of the following?

(Please <tick> one box on each row.)

		<i>Not at all</i>	<i>Very little</i>	<i>To some extent</i>	<i>A lot</i>
a)	Availability of suitably qualified Mathematics teachers.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
b)	Availability of suitably qualified Science teachers.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
c)	Availability of suitably qualified <test language> teachers.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
d)	Availability of suitably qualified <other national language> teachers.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
e)	Availability of suitably qualified foreign language teachers.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
f)	Availability of experienced teachers.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
g)	Availability of <emergency/replacement> teachers.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
h)	Availability of support personnel.	<input type="checkbox"/> ₁	<input type="checkbox"/> ₂	<input type="checkbox"/> ₃	<input type="checkbox"/> ₄
	etc.....				

In languages that have no passive voice and/or where the verbs are placed **at the end** of the sentence, the “normal” or “fluent” order of words would be something like: “*Does shortage or inadequacy of a) availability of Maths teachers..., b) availability of science teachers, (and so on) the capacity of your school to instruction provide affect?*”

Passive voice doesn’t exist in some languages.

In passive voice, the verbal or sentential form displays the grammatical subject as the recipient or "logical object" of the action expressed by the verb. Using this construction in passages can give rise to item adaptation problems because passive does not exist in certain languages. Such changes influence the structure of the language in the adapted tests and can affect the difficulty of the test items.

Examples:

1. Original sentence in passive voice:
Twenty bus passengers were injured by a bomb blast.
Sentence transformed in active voice:
A bomb blast injured twenty bus passengers

In this example when transforming from the passive to the active voice, the “animated” subject becomes an object and the “unanimated” object becomes a subject. This makes the sentence a bit more difficult to understand

2. Original sentence in passive voice:
The explosion was caused by a terrorist.
Sentence transformed in active voice:
A terrorist caused the explosion.

In this example, transforming the passive voice into the active voice makes the sentence a bit easier to understand as the abstract subject becomes an object and the object becomes an animated subject.

For translators that work in languages that only have the active voice, it becomes **very** difficult when several different passives are in the same sentence:

3. *Twenty of the bus passengers **injured by the blast of a bomb placed by a terrorist were hospitalised in critical condition.***

Try to avoid very long and/or complex sentences.

Examples:

M441 Frequent flyers (unit dropped before the FT)
What is the minimum number of return flights from Extersten to Zetersten required under the new system so that together with one return flight to Paris, Derek would earn more points than for the same number of flights under the old system?

The stems are long and complex sentences in English. In French (and in many other languages), they will become even longer and disproportionately complex. It is better to break such stems into a sequence of simpler sentences.

In PISA 2000, the question stems of a number of items were similarly worded:

“Circle “Yes” or “No” to show whether these activities, according to the material.”

In almost all occurrences, FRE source broke the sentence in two in order to keep it easily understandable. In this particular case, it was:

« Indiquez quelles sont, d’après le texte, les activités Répondez en entourant « Oui » ou « Non ».”

[According to the text, indicate what are the activities Answer by circling “Yes” or “No”.]

Remember that most other languages have much more limited “wh-phrases”¹ than English.

English is probably one of the most flexible languages in the world for the construction of interrogative sentences. Any English adjective or adverb can be used as a basis for a wh-phrase:

Examples:

How often...?
How likely...?
How strongly...?
How large...?
How big...?
How deep...?
How valuable...?
How confident...?

Unfortunately, most other languages have interrogation systems that are much less user-friendly for people who, like test developers, need to ask plenty of questions. The expressions above will require complex periphrases in languages such as French (“*À quelle fréquence...?*”, “*Dans quelle mesure est-il probable que...?*”), which very often will not fit too well with the remainder of the sentence, leading to awkward formulations. Many translators will not have any choice but to change the whole sentence – with some risk of loss of equivalence against the English item.

In many cases it will help the translators if the stem of the question only contains the “core” meaning, and the “to what extent?” part is only conveyed by the answer categories. For example, rather than asking “*How strongly do you agree or disagree with the following statements?*” (with answer categories “*Strongly agree*”, “*agree*”, “*disagree*”, “*strongly disagree*”), it would be easier for the translators if the question just read: “*What is your opinion on the following statements?*”, with the same answer categories “*Strongly agree*”, “*agree*”, “*disagree*”, “*strongly disagree*”.

When these expressions are embedded in sentences without being an interrogation, the issue remains.

For example:

In R219Q03 Employment the ENG version was:

The store could use the information about how you heard about the job to

A work out how popular the store is.

B work out how enthusiastic you are about the job.

C check on the way they find new staff.

D check on the private lives of their employees.

For options A and B that used “how popular” and “how enthusiastic”, the FRE version translated as “*Se faire une idée sur la popularité du restaurant* (get an idea on the popularity of the restaurant)” and *se faire une idée sur votre enthousiasme pour cet emploi* “(get an idea

¹ Wh-phrases in English are those that begin with a wh-adverb (what, why, where, how, etc).

on your enthusiasm for this job). Adapting the options that way resulted in a lack of equivalence between the versions.

In R111Q04, the Eng version asks:

“According to the article, how much skill in the language of the country which they are visiting do exchange students need?”

This was translated in FRE as:

“D’après l’article, quelles compétences dans la langue du pays d’accueil les étudiants doivent-ils avoir?” [According to the article, what skills in the language of the host country do exchange students need ?]

Note that there was, in this example, an additional difficulty linked to the embedded “*which*” in “*the country which they are visiting*” making thus the sentence complex. In translating this expression as “*pays d’accueil*” (host country), the FRE version avoided the problem but lost equivalence with the ENG source.

Do not use “wh-phrases” embedded in other “wh-phrases”.

Simple English interrogative sentences are hard enough to translate in other languages. Do not make the translator’s task impossible by embedding interrogations within other interrogations!

Note that embedding the wh-phrase, as part of one of the nominal groups in the sentence is even worse:

Add translation notes to alert translators to the possible ambiguity of “Which of the following” items (or write the stem otherwise).

Examples:

S253 Ozone

“Which of the questions below can be answered by scientific research?”

S213 Clothes (early version in PISA 2000)

“Which of these claims made in the article can be tested through scientific investigation in the laboratory?”

S209 Tidal Power

“Which of the following statements support(s) the choice of St Malo as a place for building a tidal power station?”

The PISA verifiers became well aware that they have to pay particular attention to complex multiple choice items where the stem includes the expression “*Which of the following...?*”. They now know that each of these will be incorrectly translated in at least five or six of the PISA national versions, even when – like in the third and fourth examples above – the English version contains some cue pointing at the fact that the item is ambiguous (either only one, or more than one statement may be selected).

In many languages, the *wh-* word itself needs to be either singular or plural, which is a trap for inaccurate translators: they will tend to translate it either into a singular *wh-*word (thus suggesting that only ONE of the statements is correct) or into a plural *wh-*word (thus suggesting that there MUST be more than one correct statement). To prevent this problem and keep the ambiguity, the translator should use a more complex wording of the stem (e.g. in French: “*Laquelle ou lesquelles des affirmations suivantes justifie(nt) le choix...*”).

It would be much safer to add a translation note to all such items, in order to explicitly draw the translator’s attention to the issue:

“*Make sure that in your national version the wording of this question calls for only one answer*” or “*Make sure that the wording calls for more than one answer*”, or “*Make sure that the wording calls for EITHER one OR more than one answer*”.

OR to reword it, as it a Yes/NO complex multiple choices. Note that in that case, the question should be formulated in such a way that in solving one problem, another problem is not introduced.

Try to avoid “inversion” of predicative complements (attributes).

Example:

S304 Water

A desirable characteristic of a drinking water installation is that the energy costs for making the drinking water should be low.

R101Q07 Rhino

The passage suggests that the main importance of the male rhinoceros’s visit (lines 87–103) was that it gave Anna Merz the chance to

At the **semantic** level, “*a desirable characteristic of a drinking water installation*” is a predicative complement (attribute) of “*the energy costs for making the drinking water should be low*”. In the English syntax, such complements are very often shifted into the status of **grammatical** subjects, while the “actual” subject becomes a predicative complement (attribute), like in the two examples above.

In many languages this kind of “inversion” is not possible, or results in extremely awkward sentences. On the other hand, if the translator just changes the grammatical status of the two parts of the sentence to avoid this problem, the emphasis is lost (“*Low energy costs for making drinking water is a desirable characteristic of a drinking water installation*”).

It would be preferable to reformulate the stem, e.g.:

“*In a drinking water installation, it is desirable that ...*”

Similarly, in the second example, the FRE version needed to adapt the sentence as: “*The passage suggests that the male rhinoceros’s visit (lines 87-103) was particularly important because it gave Anna Merz the chance to...*”

In this specific case, however, because the item is on an incomplete stem format, the whole sentence would need to be revised taking into account what is explained above.

Avoid sentences with “hidden” double negations.

Test developers systematically avoid using overt double negations, because they know that they almost always result in too difficult sentences in the stimulus, or (if used in the item stem), in flawed items. However, they may sometimes pay less attention to hidden double negations:

Example:

This does not mean that the violence in cartoons is harmless.

This sentence contains only one grammatical negation (This does not mean), but the word harmless is, in fact, a second negative expression, which results in a semantic double-negation.

In languages that do not have a lexical equivalent of the word “harmless”, the translator will have to translate into “*This does not mean that the violence in cartoons is not harmful*”, thus using a formulation that explicitly contains a double negation, and that will be more difficult than the English source.

Pay attention to articles in generic statements.

Examples:

S307 Tidal Power

“A tidal power station uses a different method to generate electricity from many other power stations, which burn coal to produce electricity.”

S326 Milk

“... the milk of three mammals, a cow, a wolf and a human.”

In languages such as French or Italian, these sentences often produce awkward translations. This is because, contrary to English, the equivalent of the ‘indefinite’ article (English “a”; French “un, une”; Italian “uno, una”) can’t really be used to convey a generic meaning; they are normally used to refer to a *specific* tidal power station, an *individual* cow or wolf, etc. In French, to avoid the problem, one has to turn the sentence into plural (“*Tidal power stations use a different method from other power stations*”), or use the ‘definite’ article (“*the cow*”, “*the wolf*”), or even, in the case of the second example, change the whole sentence : “*le lait de trois espèces de mammifères : le lait de vache, le lait de louve et le lait humain.*” (Back translated in English it would be: “*the milk of three species of mammals: the cow milk, the wolf milk and the human milk*”)

When reconciling translations of Science materials (where generic statements are legion), dozens of similar occurrences had to be corrected. It will be useful to pay attention to this issue when using Science-related material in the Reading units.

B. Vocabulary issues

Add translation notes with Latin names of unfamiliar animals or plants.

Names of living beings in PISA materials may raise two translation problems:

- (i) Since most of the materials refer to “real life” (not school-related) situations, the name used in the items is often the English “common-language” name, not a “scientific” name. The target language may (or may NOT) have a corresponding “common-language” name for the same animal or plant;
- (ii) Even when common-language names exist, the translator must check whether they are correct, since common-language names in natural languages sometimes apply to a larger range of species than in scientific language.

Examples:

All names of animals and plants in S126 Biodiversity were particularly hard to translate, e.g.:
 LEAF HOPPER: *sextius virescens* (also called “acacia tree hopper”) or *sertorius australis*.
 BUTCHER BIRD: generic name for all *Laniidae*, i.e. all birds in the world that have ‘*lanius*’ as their first Latin name.
 TEA TREE: was not the well-known tea plant, but the Australian name of *melaleuca alternifolia*.
 HONEYEATER: generic name of all birds that have Latin names starting with ‘*meliphaga...*’, ‘*certhionyx...*’, ‘*melithreptus...*’ and several other species.
 NATIVE CAT: common name for all Australian marsupials that have Latin names starting with *dasyrus* (thus, nothing to do with what Europeans know as a *wildcat*).

It would therefore greatly help the translators if translation notes are added in the units containing such names, pointing out the correct Latin name, so that the translators can check in their national dictionaries.

This does not apply to Science materials only. For example, in R119 Gift (a story situated in North America, probably in Mississippi) the author used a common language name “*panther*” to mean a *cougar* (Latin name: *puma concolor*), which was confusing for many translators around the world. It was often translated into “*leopard*” (Latin name: *panthera pardus*), an animal which does not exist in that geographical area.

In another Reading unit, R077 (about Flu immunisation) the terms *flu* and *influenza* appeared in the stimulus. If these words seem quite common in the everyday language, the translation guidelines call for the use of two different terms in the translation. However in many languages there is only one term to translate flu-influenza. (for example “grippe” in FRE). This difficulty was overcome in a science unit in PISA 2006 (S477Q04 Mary Montagu) where the term *influenza* followed by *flu* within parenthesis was used, allowing countries that have two terms to use them and countries that have only one term to delete the parenthesis.

Just for fun, here is a mail exchange about ONE of the S126 Biodiversity names:

 From: Aletta Grisay
 Sent: 06 January 1999 12:23
 To: Gayl O'Connor, Ray Adams

Dear Gayl,

In S126 Biodiversity, there is a problem with one of the birds in the graphic, the *honeyeater*. The latin name seems to correspond to '*bee-eater*', which would translate into '*guêpier*' in French, but was translated into '*colibri*' by my translators.

Colibri is clearly wrong, since this bird does not eat larvae nor *cicadellidae* (*leaf hoppers*). But, first, Steve checked in the Internet, and found that the bee-eater does not resemble at all the bird in the graphic. And, second, I can't possibly accept to correct the French version into '*guêpier*', since this word means '*wasp-eater*' – which would be extremely confusing: there is a wasp (French: *guêpe*) in the graphic, but the wasp is NOT in the bee-eater food chain!

Steve suggested to change the name into another bird in the same family, which resembles strongly the bird in the graphic and who eats larvae and *cicadellidae*; the Latin name is '*certhidae*', and the French name ('*grimpereau*') would raise no problems. Another acceptable bird would be '*coracidae*' (French '*rollier*') who also resembles the bird in the graphic, has appropriate food habits, and no confusing name. Let me know if any of these two birds could be found in the Australian chain food represented in the graphic (I would not want to be fined by the Australian authorities for importing extraneous European birds into your ecologic systems!).

Aletta

From: OConnor Gayl
Sent: 07 January 1999 09:22
To: Adams Ray

Ray,

Honeyeaters eat insects as well as nectar. I do not know what the term '*colibri*' means. Aletta does not say, but it could well be that it is ok, based on the fact that the honeyeater DOES eat leaf hoppers and larvae. I can't make a decision on the suggestion to use the name '*certhidae*' because it is not clear to me to what this is referring...it would be helpful if I knew the common name (in English) of the bird they mean.

Gayl

From: Aletta Grisay[SMTP:agrisay@ibm.net]
Sent: 07 January 1999 12:27
To: Adams Ray, Cc: Gayl O'Connor

Ray & Gayl,

I wasn't able to reach Steve, who has the appropriate dictionaries.

a. According my own dictionary, '*Colibri*' is '*hummingbird*'. To my knowledge, these birds only eat nectar, and their latin name is different, therefore out goes the colibri.

b. The CORRECT translation of honeyeater in French is '*guêpier*', according what Steve wrote yesterday. The Latin name and food habits are in accordance with the stimulus. BUT on one hand, I do not want to have '*guêpier*' in the French version, for psychometric reasons (the name in French would be highly confusing). On the other hand, Steve suspected that the drawing of the bird might be wrong, since the bird in the graphic did not resemble the picture of the honeyeater that he saw in his dictionary (I can't confirm: my dictionary does not have pictures!).

c. '*certhidae*' ('*grimpereau*' in French) is, according to my dictionary, a '*tree creeper*'; Steve said that it reasonably resembles the bird pictured in the graphic and eats the same disgusting things. Therefore, for the sake of the item credibility, I would by far prefer to have in

the graphic a '*tree creeper*'/'*grimpereau*' instead of a '*HONEY eater*' who eats LARVAE in English and a '*guêpier*' who is supposed NOT to eat wasps ('*guêpes*') in French. But, of course, if the tree creeper does not exist in the kind of sites described in the item, I would not insist in introducing it!

Aletta

Date: Sat, 09 Jan 1999 11:17:33 +1100

De: Adams Ray <adams@acer.edu.au>

A: agrisay@ibm.net, <oconnor@acer.edu.au>

Aletta,

Because of the translation problems we suggest that you use "*tree creeper*". Gayl says that the graphic is not really like a tree-creeper "*tree creepers are actually quite squat birds, and would not have such a pronounced pointy beak*" but it is better to leave the picture as it is.

Ray

Remember that English adjectives with some “subjective” connotation will NEVER have appropriate equivalents in ALL other languages.

Example:

Item 8 in R119 Gift presents a dialog where one of the speakers says that "*the woman in the story is heartless and cruel*", while the other speaker thinks she is "*a very compassionate person*". The item asks the student to give evidence from the story to show how each of these speakers could justify their point of view.

R119Q08 is an item that heavily relies on the correct interpretation of words such as "*compassionate*" and "*cruel*". The closeness with the English meaning of the corresponding word used in the national versions varied significantly across the PISA languages, which resulted in several item/country interactions.

Items that rely on synonymy or antonymy between particular words are always very problematic in a multilingual study, due to the fact that equivalent pairs of words that have exactly the same connotations as the English pair are almost impossible to find in all languages.

In the IEA Six-subject study a vocabulary test was used as a surrogate measure of students' verbal fluency. "*The item format that was selected was that of word pairs where the words of a pair must be judged to be approximately either synonyms or antonyms, that is, either the same in meaning or nearly the opposite in meaning*" (Thorndike, 1973). About 300 word pairs (distributed in ten classes of anticipated difficulty) were circulated to the countries' experts, "*...who were asked to indicate which of the pairs it would be either difficult or impossible to translate. (...) All word pairs were dropped from further study in which more than one country indicated difficulty in the matter of translation*". About 40 out of the 300 pairs survived both this review and the Field Test, and were used in the Main Study. However, "*as the data were received and examined, it seemed rather doubtful whether either of the original expectations, i.e. of equal average difficulty for the complete set or of high correlation of difficulties from one language to another, was borne out. (...) It appeared that the items tended to get systematically easier in translation (...)*", and the dispersion of difficulties varied significantly across countries. "The average correlation of item difficulty

from language to language (...) was only about 0.4 or 0.5". The author concluded that, while the measure obtained could be used for within-country analyses, "*it would only become suspect if it were used for cross-language comparison*".

If the example given above (R119 Gift) doesn't have the same question format (word pairs), it presents the same characteristics.

Similar problems occurred in an item where the student was asked whether a given decision (or interpretation of evidence) could be considered as "*wise*", or "*reasonable*". For example: "Which one of the following is likely to be the most reasonable estimate of...". Many languages do not have a proper word for it, and the verifiers were busy correcting translations like "most intelligent", "most wise" and so on.

A last example from PISA 2000, is coming from the unit R101Q08 where the students were asked to interpret the text and determine the feelings of a person in front of a rhinoceros.

A translation note with precise indications would have been very useful in order to try to reach the best possible equivalence.

Affective Questionnaire items are particularly sensitive to this specific kind of vocabulary issue. For example, in PISA 2003 there were problems with the translation of the stem of the Maths self-efficacy scale, where the student was asked to what extent he felt "*confident*" in solving a number of Maths problems. The word used to translate "*confident*" in many other languages is likely to convey either a higher or a lower degree of certainty/uncertainty than the English word, and this may have some impact on the distribution of responses.

Avoid when possible the use of metaphors or add translation notes.

A well-known example of literal translation of an English metaphor is "*Out of sight, out of mind*" translated as "*invisible, insane*".

In R081 Graffiti, question 2, the ENG source version used the word "*costs*" with both its "normal" meaning ("*financial costs*") and a more "metaphorical" meaning ("*damages caused to public buildings and to the reputation of young people*"). No proper corresponding word could be found in many national versions. A number of DIFs were observed for this item. Most of the FRE versions used the word with italics and double commas to help preventing a too literal understanding. Apparently this was not sufficient to avoid a differential functioning of the item.

In France, the word "*coût*" was consistently adapted into a less ambiguous word "*prejudice*" [damage, harm], also used with double commas – and the item appeared to be easier.

In R119Q04 Gift, the question was:

When the woman says, "and then I'll see to you" (line 92) she means that she is

- A *sure that the cat won't hurt her.*
- B *trying to frighten the cat.*
- C *intending to shoot the cat.*
- D *planning to feed the cat.*

For the FT this item had poor statistics, in the MS version a translation note was thus added to clarify the meaning of “and then I’ll see to you”: “Please ensure that the phrase, “and then I’ll see to you” allows BOTH of the following interpretations: “and then I’ll feed you” AND “and then I’ll shoot you”. Indeed translated literally one or even both interpretations would be impossible in many languages.

The multiple choice question 3 in the unit R228Q03 asked the students what the message “Remember to smile on cloudy days” meant that they should do. This item had poor statistics and was not retained for the MS. Providing a translation note would have most probably helped the translators to try and find a similar metaphor.

In R232 Unicef, the stimulus used the following two expressions: “The solution is in your pocket” and “the tip of the iceberg”. It is likely that in a number of languages one or both of these expressions could not be translated literally. For example, in FRE the first one read: “the solution is in your purse” in order to render the correct meaning. Again in that case, a translation note clarifying the meaning of these expressions would have been necessary.

Definitions of quantification limits can vary from one language to another.

Examples

X601 Cinema (early version)

Suitable only for persons over 18 years old.

In some languages the notion “over” doesn’t include the 18 years old, in other it does. It was necessary to make clear that the 18 years old were included. The sentence was reworded: “Suitable only for persons of 18 years and over”.

R088Q03 Labour (scoring)

Code 21: Allow approximations between 949,000 and 950,000 in figures or words

In languages like German and French the word “between” excludes each boundary (in this example 949,000 and 950,000). However the correct answer should include these limits. Writing “from 949,000 to 950,000” would avoid this problem. In coding guides (for the attention of coders, not students) the limits can be made even clearer by using mathematical symbols for “less than”($<$), “less than or equal”(\leq), etc.

The same kind of problem occurred in items referring to buildings with storeys. In Australia a two-storey house is a house that comprises the ground floor and a first floor. In French-speaking countries a two-storey house is a house with the ground floor, a first and a second floor. In M535 Twisted building, the notion was made clear in writing: “**Above the ground floor** there are 20 storeys containing apartments” and “*The top floor (the 20th floor above the ground floor) is at right angles to the ground floor.*” and also “*Draw in this diagram the plan of the 10th floor above the ground floor, showing how this floor is situated compared to the ground floor.*”

Avoid asking for a determinate number of words to be chosen from a list or sentence.

Example:

S444Q06 White storks

The scientists wanted to find out more about the death toll of migrating birds in other parts of the world.

You can use only three of the words underlined above as keywords in an internet search for information. List the three words that would produce the most useful information.

In a number of languages (as German and Dutch) the two underlined words “*migrating birds*” are translated as a single word, which is a big issue as 2 of the keywords were “migrating” and “birds”. In some cases the same issue arose with the two words “*death toll*” where “*death*” was one of the keywords. This item required quite a number of heavy adaptations resulting in a biased item.

Moreover in some other languages (as Greek), National centres or verifiers noted for example: “*for the sentence to work grammatically, the 3 key words must be in genitive/plural. But in an internet search, students would more naturally choose the words in the nominative case/singular.*”

Thus, the two items of this kind used in the FT 2006 were dropped for the MS.

Plural form of nouns after numerals can vary according to the quantity.

Examples

Answer:.....zeds

Answer:.....cans

In some languages, having to put a noun after a numeral in the answer space poses a problem, the ending of the noun being variable depending on the numeral it follows, i.e. it can prompt certain answers and exclude other ones.

In Slavonic languages, for example, one needs to use a different ending for the word zeds depending on whether there is 2-3-4 zeds or 5 zeds or more.

This kind of prompt could be avoided and be replaced in these examples by: “Cost in zeds:.....” and “Number of cans:.....”

Avoid when possible the (unnecessary?) use of too technical words.

Example:

“nylon thread, metallic silver thread, wax linen thread, polyester thread, cotton embroidery thread, and yarn.”

Dozens of translators around the world got headaches in trying to find out what some of these threads were, and 25 verifiers got headaches in trying to fix their errors.

Some common English words are hard to translate.

- ***Evidence*** as in “Give ***evidence*** from the table that supports the idea that this story might be true” and in “Support your answer with ***evidence*** from the text”. (R11Q02B)

A number of languages do not have the exact equivalent of the word “*evidence*” (to my knowledge, for example, none of the Romance languages has it). Then the translator is in trouble finding an appropriate word:

“Give an element from the table...”, “Quote some data from the table...”; “Find a proof in the table...”.

Or as in “*Explain whether you think there is any evidence for concern*”. It is also quite difficult to maintain the conciseness and the vagueness of the English expression “*evidence for concern*”. The translation would need to be something like: “explain whether you think there is any element that might support these concerns” which may be less ambiguous (thus a bit easier) than the English version.

- **Idea**, when used to mean “*hypothesis*”, or “*conjecture*” or “*theory*”, as in “*Give evidence from the table that supports the **idea** that this story might be true*” or in “*By observing the behaviour of the flies, what was the main **idea** Hoy was testing?*”. I suspect that in a number of languages, when using the word « *idea* » instead of « *hypothesis* », the meaning is not as neutral as in English. It rather has some slightly negative connotation (an “*idea*” in this case tends to be a “*poor*”, or a “*very uncertain*” or an “*odd*” hypothesis). I’d suggest to add a translation note to indicate the cases when using the more ‘formal’ word “*hypothesis*” is acceptable, and those when the translator must avoid it and find some alternative wording.

Other difficult words are:

- **Insight**
- **To value**
- **A range of**
- **Pattern**

Translators are also in trouble with some English expressions:

- **More likely than not:** Example: *More likely than not, there will be an earthquake in Zed City at some time during the next 20 years.* This is a perfect example of translation trap: such a simple and efficient expression in English, but in almost all other languages one needs two lines to express the same idea.
- **In which x of the y situations:** Example: “*In which two of the four situations...*”. Such expressions are not translatable in languages like French and other Roman languages without totally changing the sentence and using more than 10 words.
- **For not longer than:** Example: R240Q01 Pain Reliever. The question was: “*What is the maximum number of days you should take this medicine?* The answer can be found in the stimulus as “*for not longer than 7 days*”. In many languages (as French) it is impossible to translate it with the negation (not longer) which emphasises the “maximal” number of days. The FRE read as “*pendant 7 jours au plus*” which is less straightforward. To have a better equivalence the ENG source could have been changed, for example, as “for 7 days at the most”.

Use of scientifically appropriate terminology versus everyday life terminology.

Natural languages often use less than precise words to describe chemical, physical or biological phenomena – or even words that may convey common misconceptions, such as

‘weight’ used to mean ‘mass’, ‘heat’ used to mean ‘temperature’, ‘circle’ used to mean ‘disk’, etc.

For the Science materials to be developed but also in the scientific texts used in Reading, it will be particularly important to add translation notes to help translators distinguish the cases when they should use “common language” terms that may occur in real life situations involving some scientific aspect, from those where the context is purely “scientific” and use of strictly appropriate terminology is needed.

It may also be important to address possible conflicts between the “authentic” nature of the materials used in PISA (which implies the use of stimuli where scientific phenomena may be described in everyday language) and the requirement, in a number of school systems, that when “doing science” students should only be instructed using “rigorous” terminology. Such conflicts should be discussed with the NPMs and with the representatives of their Science National Committee who attend the Science Forum – so that a consensus can be reached on the common instructions that must be conveyed to all translation teams when they have to deal with such issues.

In addition, it would be essential that strong emphasis is put on the need for the NPMs to systematically identify in their *Test Adaptation Spreadsheets* all cases when these kinds of terminology problems were encountered, and what was their proposed solution, in order to ensure that all possible doubtful adaptations are discussed with the consortium.

Examples:

R240Q01 Pain reliever

“INDICATIONS: Headaches, muscle pains, rheumatic pains, toothaches, earaches. RELIEVES COMMON COLD SYMPTOMS.

DOSAGE: ORAL. 1 or 2 tablets every 6 hours, preferably accompanied by food, for not longer than 7 days. Store in a cool, dry place.

CAUTION: Do not use for gastritis or peptic ulcer. Do not use if taking anticoagulant drugs. Do not use for serious liver illness or bronchial asthma. If taken in large doses and for an extended period, may cause harm to kidneys. Before using this medication for chicken pox or influenza in children, consult with a doctor about Reyes Syndrome, a rare but serious illness. During lactation and pregnancy, consult with a doctor before using this product, especially in the last trimester of pregnancy. If symptoms persist, or in case of an accidental overdose, consult a doctor. Keep out of reach of children.”

In PISA 2006, some countries had a tendency to choose the most familiar term rather than the exact scientific medical term. In this case a translation note would be needed, specifying whether the medical terms have to be translated with the exact medical terms or with common familiar terms. Using more familiar terms whereas we expect them to use the exact medical terminology would make the stimulus easier to read.

Weight/Mass in M836 Postal Charges (PISA 2003 FT):

The postal charges in Zedland are based on the weight of the items (to the nearest gram), as shown in the table below:

Weight (to nearest gram)	Charge
Up to 20 g	0.46 zeds
21 g – 50 g	0.69 zeds
51 g – 100 g	1.02 zeds
101 g – 200 g	1.75 zeds
201 g – 350 g	2.13 zeds
351 g – 500 g	2.44 zeds
501 g – 1000 g	3.20 zeds
1001 g – 2000 g	4.27 zeds
2001 g – 3000 g	5.03 zeds

In daily life contexts people use “weight”, not “mass” and in this unit it didn’t have a scientific impact on the item, so the countries were asked not to adapt “weight” as “mass”. However many countries sent remarks and comments from their experts. Here is one of the comments we received:

“In both Switzerland and Italy the term Weight, followed by the unit Kg is not anymore accepted in Maths and Science. Therefore we should translate Weight into Mass if we want to use the correct terminology, although this could impair our students.

What is the rationale, internationally, for maintaining the term “weight” (which is what is used in both the French and English version) given that is used in a Maths test and that the term is wrong from both a mathematical and scientific perspective?”

We received similar comments from some other countries.

The same type of problem occurred in Mathematics with words such as *line/segment* or *circle/disk*.

Line/segment: in M161Q1 where the English stimulus read: “Triangle PQR is a right triangle with right angle at R. The line RQ is less than the line PR. M is the midpoint of the line PQ and N is the midpoint of the line QR. S is a point inside the triangle. The line MN is greater than the line MS.”

In the French version the word “line” was translated as “segment” to be mathematically correct as requested by national experts.

Insert translation notes when adaptations are needed or allowed.

Examples

R092Q03 Movie

Imagine that you are planning to use the voucher on a Saturday.

You can use it for the session starting at

2.00 p.m.

10.30 p.m.

8.00 p.m.

5.30 p.m.

R242 Marathon

“University of Maryland senior Stacy Chanin on Wednesday became the first person to swim three 28-mile laps around Manhattan”.

In both cases a translation note asking or allowing the country to adapt the hours and the miles to what is commonly used in their country or to what is in used in the local cinemas (for the first example) would have been useful. Even if this is not a major translation issue, the country that translates the source version as such (without adaptation) might make the unit more difficult to understand for their students.

C Graphics and stimulus material

Generally, graphic artists will redraw graphics attached to items selected for inclusion in the field trial. However, redrawing will not take place until after items have been distributed to NPMs for feedback, so care should be taken to design graphics exactly as you wish them to appear.

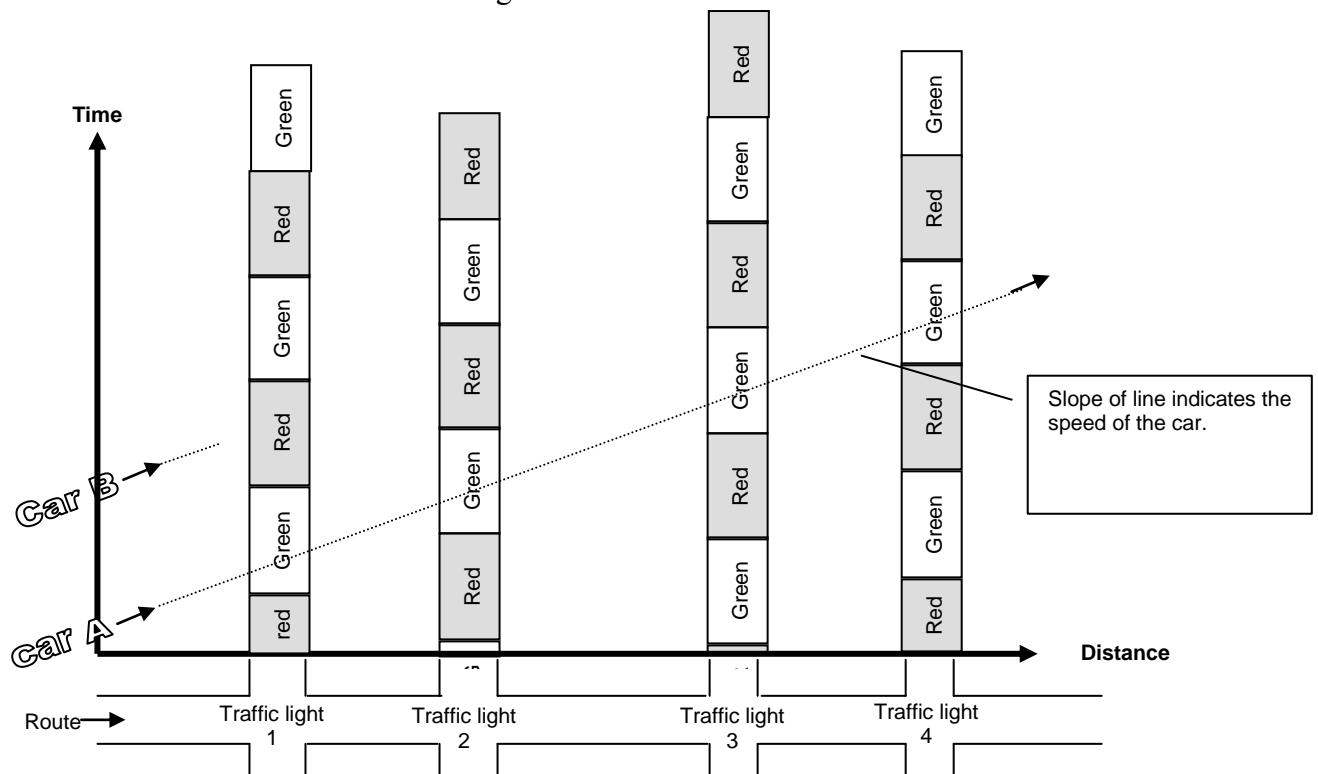
There are several things that must be kept in mind when designing the final version of a graphic in order to avoid translation problems. It is wise to keep these things in mind from the outset, so that it will be possible to reproduce the same appearance in the final graphic in all languages.

Text elements.

Text elements in graphics should always be be in easily accessible boxes, and never require complex manoeuvres like ungrouping/grouping. It appears that many countries/verifiers will not be able to perform such actions and will resort to other less desirable fixes or give up altogether.

Text boxes should be calibrated taking into account that many languages (German for example) are not as concise as English.

The following example (from PISA 2003 field trial) had both problems. The translators needed to ungroup the graph to have access to the word in “Word art”(which also means that they have to copy it in a separate document and reinsert it afterwards) and had problems with the size of the text boxes for “route” and “green/red”.



It is also important to keep in mind that for some languages (such as Arabic) the two axes would need to be inverted as the reading is done from right to left.

Text and graphic sizes

When it is intended to have the stimulus material to fit on only one page or specific frame, please provide enough space for translation in any language. For example, French and even more German need much more space than ENG. Generally it is recommended to the countries to modify slightly the line spacing or the font size, but this becomes impossible when the ENG already uses a very small font size.

In the unit R103 Sphinx, the stimulus in ENG is in font size 10 but the FRE do not have any choice but to use the font size 8. Please see below what difference it makes from one version to the other:

“This time, the government did it right. Its 10-year fix-up, unveiled this week, used the expertise of skilled conservationists from around the world”

“Cette fois, le gouvernement a bien fait les choses : pour les travaux de réparation, qui ont duré 10 ans et dont on célébrait cette semaine l’achèvement, il a eu recours aux compétences de restaurateurs chevronnés venant du monde entier. ”

The FRE text appeared much less attractive, denser and quite difficult to read.

Two other examples are given in the next two pages:

R122 Just Art, where the ENG text needs the whole page and where less concise languages might have experienced difficulties.

R241 Warranty

In this case if the ENG frame is smaller, the text inside takes all the available room.

Unfortunately, enlarging the frame needs ungrouping several times the picture, then playing with the columns to keep a similar layout as the source, then regrouping everything.

D Cultural issues

The Consortium has in place a variety of mechanisms to address potential cultural issues, including the following:

- item development teams in different countries/cultural settings
- inclusion of items submitted by participating countries
- international PISA Reading Forum Reviews
- National Reviews
- Reading, Math and Science Expert Group Review

However, some attention to cultural references is still needed. In this era of “globalism” western cultural references are not necessarily universal, and some could be unfamiliar, misunderstood or even negatively connotated in some participating countries.

In some cases, it is relatively easy to purge an item of an inappropriate cultural reference. E.g: The Pisa 2003 Math item “Fence” included a reference to “enclosures for pigs”. This was adapted to “enclosures for sheep” in the version for Indonesia, a Moslem country, with no effects on the difficulty of the item.

In other cases, a whole item can revolve around a concept familiar to westerners but unfamiliar in other cultures. For example, in the Problem Solving unit “Children’s Camp” (FT version), the problem was Anglo-Saxon cultural references, not easily transposed even to other western cultures. The issue was meal patterns, i.e. the distribution of light/heavy meals during the day and the name for these meals.

Other examples: S321 Heat Loss from FT 2003, Heating and Insulation from PISA 2000.

R081Q01 Graffiti:

The purpose of each of these letters is to

- A explain what graffiti is.*
- B present an opinion about graffiti.*
- C demonstrate the popularity of graffiti.*
- D tell people how much is spent removing graffiti.*

This item was retained for the MS without any changes except that the concept of graffiti was now explained before the items, trying to palliate potential cultural problems linked to knowledge about graffiti (or lack thereof).

R118 Bullying

The theme of this unit is very well known and treated at school with the students and parents in some countries (ENG speaking for example). This is not the case in other cultures; it is even difficult to find a single perfect translation for “bullying”. In FRE for example, you will find at least three different ways to translate it. It is likely that this unit was somehow culturally biased even if the theme fitted very well in the interests of 15-year-olds.

It is true that such units or items were dropped or modified before the main study but knowing in advance that such topics engender cultural issues could avoid developing such items.

Final comment

As a conclusion, when we consider all of the activity that goes into developing the items that are finally included in the main study instruments in each PISA country, it is clear that the cost of developing each PISA item is extremely high. Translation-related issues can render an item immediately unusable, and often such issues do not come to light until the item development process has advanced quite a long way. Any success we can have in preventing translation problems at the earliest possible time will result in better items, and a more effective and cost-efficient item development process.

Finally, here is an encouraging quotation to remember when working hard on the item development: “*One of many reasons of the success of recent TIMSS and OECD/PISA studies is the substantial effort that has gone into the source-language test development with clearly defined constructs and test specifications, careful item development and field-testing, and other activities associated with proper test development.* (Hambleton, 2005)

References

Hambleton, R.K. (2002), Adapting achievement tests into multiple languages for International Assessments, in A.C. Porter and A. Gamoran, Eds, *Methodological Advances in Cross-National Surveys of Educational Achievement*, Board on Testing and Assessment (BOTA). Washington, DC: National Academy Press, 58-70.

Hambleton, R.K., Merenda, P.F., Spielberger, C.D. (2005), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*, IEA, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.

Thorndike, R. L. (1973), *Reading Comprehension in Fifteen Countries*, IEA International Studies in Evaluation III. Stockholm: Almqvist and Wiksell.

APPENDIX 2: PISA 2009 TRANSLATION AND ADAPTATIONS GUIDELINES

PISA 2009 TRANSLATION AND ADAPTATION GUIDELINES

Doc: NPM(0709)1

National Project Managers' Meeting
Dubrovnik, Croatia

24-28 September 2007



Core A Consortium:

Australian Council *for* Educational Research
(ACER)

Unité d'analyse des systèmes et des
pratiques d'enseignement (aSPe)

cApStAn Linguistic Quality Control

Deutsches Institut für Internationale
Pädagogische Forschung (DIPF)

National Institute for Educational Policy
Research (NIER, Japan)

Westat

INTRODUCTION.....	4
OVERVIEW OF THE PISA TRANSLATION, ADAPTATION AND VERIFICATION PROCEDURES.....	4
<i>Double translation and reconciliation.....</i>	4
<i>Translation guidelines.....</i>	5
<i>Translation notes.....</i>	5
<i>International verification.....</i>	5
<i>Adaptations approval procedures.....</i>	5
<i>“Dodgy” items analyses.....</i>	5
1. GENERAL INSTRUCTIONS.....	7
<i>Materials.....</i>	7
<i>Security requirements.....</i>	8
<i>Translators.....</i>	8
<i>Reconcilers.....</i>	8
<i>Use of electronic files.....</i>	9
<i>Harmonisation.....</i>	9
<i>Consortium’s errata.....</i>	10
2. RECOMMENDATIONS ABOUT TRANSLATION TRAPS IN TEST MATERIALS.....	11
LAYOUT AND GRAPHICS.....	11
<i>Respect the layout and presentation shown in the source document.....</i>	11
<i>Check the rendering of illustrations and graphic elements.....</i>	11
LINGUISTIC DIFFICULTY.....	12
<i>Length of sentences.....</i>	13
<i>“Common” vocabulary vs “scientific” terminology.....</i>	13
<i>Affective vocabulary.....</i>	14
<i>Idioms and metaphorical expressions.....</i>	15
<i>Abstract words or turns of sentences.....</i>	15
<i>Particularly difficult-to-translate English words.....</i>	15
<i>Active vs passive turns of phrase.....</i>	16
<i>Negations.....</i>	16
<i>Reference chains.....</i>	17
<i>Generic articles.....</i>	17
COMMON PSYCHOMETRIC TRAPS.....	17
<i>Do not modify the length of key and distractors in multiple-choice items.....</i>	18
<i>Be careful with truncated items.....</i>	18
<i>Respect literal matches.....</i>	18
<i>Provide all the information contained in the item.....</i>	19
<i>Check all of the “Which of the following...?” questions.....</i>	19
<i>Try to respect the order of information in the question stems.....</i>	19
<i>Small details are sometimes important.....</i>	20
OTHER.....	20
<i>Pay attention to scoring instructions.....</i>	20
<i>Respect all translation notes.....</i>	20
3. NATIONAL ADAPTATIONS OF TEST MATERIALS.....	22
WHAT SHOULD BE ADAPTED, WHAT SHOULD NOT BE ADAPTED?.....	22
<i>Never adapt the format of an item.....</i>	22
<i>Do not include explanatory notes nor additional instructions.....</i>	22
<i>Adapt mathematical symbols, abbreviations and formulas.....</i>	23
<i>Adapt mathematical and scientific vocabulary.....</i>	23
<i>Check the use of scientific and mathematical terminology vs “common language”.....</i>	23
<i>Check “common language” quantitative expressions.....</i>	24
<i>Do not adapt the “zed” currency.....</i>	24
<i>Names of persons and locations can usually be adapted.....</i>	24
<i>Some names should NOT be adapted.....</i>	25
<i>Check that adaptations are consistently applied.....</i>	25
<i>Think of possible other adaptations that may be needed in your country.....</i>	25
<i>Document all national adaptations made in the test materials.....</i>	26

<i>What about the PISA test material that was submitted by your country?</i>	26
4. TRANSLATION AND ADAPTATION OF QUESTIONNAIRES	27
<i>Keep in mind that some respondents will misunderstand anything that can be misunderstood</i>	27
<i>Avoid translating complex English questions into “translanese”</i>	27
<i>Do not change the question layout</i>	27
<i>Pay attention to omissions that may change the item format or the item scope</i>	28
<i>Check the level of generality of the words used</i>	28
<i>Pay special attention to the wording of answer categories</i>	29
<i>Never change the order of answer categories</i>	29
<i>Do not introduce mismatches between answer categories and the wording of items</i>	29
<i>In translating attitude items, pay attention to words that convey affective nuances</i>	30
<i>Avoid increasing the social desirability of self-reported responses</i>	30
<i>Correctly adapt all <bracketed> terms and expressions that relate to school organisation</i>	31
<i>Document all Questionnaire adaptations</i>	32
5. TRANSLATION AND ADAPTATION OF MANUALS	32
<i>Notes to NPMS</i>	32
<i>Procedures that should NOT be changed in the Manuals</i>	32
<i>Procedures that may be changed or adapted</i>	32
<i>Documenting adaptations for Manuals</i>	33
APPENDIX A	35
CHECKLIST FOR THE VERIFICATION OF PISA TEST MATERIAL	35
<i>Verification of test units</i>	35
<i>Verification of questionnaires</i>	37
<i>Verification of the layout/presentation of the booklets (final optical check)</i>	37
<i>Other Remarks</i>	37

INTRODUCTION

OVERVIEW OF THE PISA TRANSLATION, ADAPTATION AND VERIFICATION PROCEDURES

The PISA studies are carried out in a large number of countries with different languages, different cultures and different school systems. The aim is to assess 15-year-old students' proficiency in three domains: Reading, Mathematical and Scientific literacy, using strictly equivalent **test batteries**, translated and adapted into each of the languages of instruction of the participating countries. The assessment instruments also include **questionnaires**, used to collect information on the students' educational environment, that might help explaining the differences in achievement observed among groups of students within and across countries, and **manuals** designed to minimize any potential source of bias in field operations and test administration procedures.

In order to achieve international comparability of the data collected in the study, the equivalence of all national versions is an essential requirement, which means that the translation of materials must imperatively be of extremely high quality in all national versions used by participating countries. It is particularly important to ensure that the translation process does not introduce biases likely to distort international comparisons:

- By making more difficult (or easier) the comprehension of texts, graphics or tables used as stimulus in the various test units;
- Or by unintentionally modifying the difficulty of the questions asked of the student, through a formulation that changes the type of mental strategy required;
- Or by introducing ambiguities that could impair (because of lack of comparability across countries) some of the variables collected through the questionnaires on the educational context;
- Or by adapting part of the material to the national context in ways that change too extensively the data collected (e.g. when adaptations in the Manuals result in undesired changes in the sampling, test administration or marking procedures).

PISA has therefore implemented a number of procedures to ensure equivalence across all national versions:

Double translation and reconciliation

To comply with PISA translation standards, it is *required* that the national versions of all test and questionnaire instruments used in the assessment are developed through a **double-translation and reconciliation** procedure¹. That is, two independent translators should first translate the source material into the target language; then a third person should reconcile these two translations into a single national version.

The Consortium develops two parallel source versions of the test and questionnaires², one in English and the other in French. It is *recommended* that the countries use the English source version for one of the translations into the national language; and the French source version for the other. In previous PISA studies, **double translation from the two source languages** proved to be most effective in

¹ No double translation is requested for manuals.

² Manuals are developed in English only, except for the script that will be read to students during the test administration, which is prepared in both English and French.

reducing translation errors and producing high quality material. In those countries where finding competent translators from both languages would be a problem, an alternative (also very efficient) method could be double translation and reconciliation from one of the source languages, followed by extensive crosschecks against the second source language (Grisay, 2003).

Translation guidelines

Detailed translation/adaptation guidelines have been developed for use by PISA translators. They are described in this document.

Translation notes

A number of translation notes have been embedded within the materials, to help translators with specific adaptation or translation problems.

International verification

As part of the quality insurance procedures implemented in PISA, all participating countries will be requested to submit their national versions for verification against the English and French source versions by a team of independent translators with native command of the various PISA languages, who have been specially appointed and trained by the Consortium. Appendix A in this document contains the checklist that will be used by the international verifiers to identify possible translation problems and to suggest edits and corrections.

Adaptations approval procedures

The NPMs will be asked to document any national adaptations they may want to include in the materials by describing them in specially designed forms: *Test Adaptation Spreadsheets* (TAS) for cognitive tests, *Booklet Adaptation Spreadsheet* (BAS) for the common *Test booklet parts*, *Questionnaire Adaptation Spreadsheet* (QAS) for Questionnaires, and *Manual Adaptation Spreadsheet* (MAS) for Manuals. The QAS and MAS will need to receive approval from the Consortium before the adaptations are entered and the national materials are finalised.

“Dodgy” items analyses

Item analyses will be conducted on both the Field Trial and Main Study data in order to check whether any test item appeared to have poor psychometric qualities. Participating countries will receive information on the items that behaved poorly in their dataset. They will be requested to review them for translation or other flaws, and either to correct possible bugs (at the Field Trial phase) or to advise on item deletion (at the Main Study phase).

A more detailed description of these procedures can be found in the *PISA NPM Manual* and in the PISA technical reports.

THE PISA TRANSLATION AND ADAPTATION GUIDELINES This document is intended for use by all national staff in PISA participating countries that may be involved in translation and adaptation of PISA assessment materials, that is:

- The NPM, or the person(s) who will co-ordinate translation activities at the National Centre. If different persons or different teams are in charge of the development of national versions in different languages, each co-ordinator or each team should receive copies of the *Guidelines*.
- Translators who will be in charge of producing the national version(s) of the materials. It is recommended that the *Guidelines* be used as training materials in the session where translators will be instructed about the aims of the PISA study and about the special requirements that must be paid attention when translating assessment materials.
- Reconcilers who will be in charge of reconciling two independent translations received from the translators into a final national version as equivalent as possible to the source versions provided by the Consortium.
- Domain experts who will be in charge of reviewing the materials for appropriateness of content and terminology.

The development of the PISA *Guidelines* was based on existing literature on international test adaptation (Hambleton, 1994; Hambleton and Patsula, 1998 and 1999; Jeanrie and Bertrand, 1999, Hambleton, 2002), on similar guidelines developed for previous IEA studies (O'Connor and Malak, 2000; Kelly and Malak, 2001), and on experience acquired through the verification of the PISA materials.

Section 1 of the document contains general instructions. Section 2 includes a number of recommendations concerning the most common translation traps encountered when translating test materials. Section 3 lists adaptations that can or cannot be introduced when translating PISA cognitive tests. Sections 4 and 5 contain specific notes on translation and adaptation of Questionnaires and Manuals. Appendix A contains the Verification Check-list used by PISA international verifiers³. Appendix B (currently in preparation) will be circulated at the NPM meeting in Dubrovnik. It will contain special notes intended for national translation teams who will be in charge of translating the materials into non Indo-European languages.

³ The examples used throughout this document to illustrate possible translation problems were mainly drawn from materials prepared for the PISA 2000, 2003 and 2006 Field Trials. Most of these items were NOT used in the final assessments. However, *some* of the examples were drawn from the actual materials used. Please consider all of the examples as embargoed materials.

1. GENERAL INSTRUCTIONS

Materials

It is the responsibility of the National Project Manager to ensure that s/he has duly received both source versions (in the English and French language) of all the materials to be translated and to organise their distribution to the team of translators whom s/he will have appointed for this work. The materials to be translated include:

- **Mathematics, Reading, and Science cognitive tests**, presented as an item pool composed of separate electronic files, each of which will contain one complete test unit, i.e.: a *stimulus* (text, graphic, diagram, ...), which shall be used as a basis for the work of the student; one or several *questions*; and the *scoring instructions* to be followed when coding the answers. Note however that the translation of the scoring instructions of automatically coded items is not required. Such items can be identified by the item label (in light grey at the right side of the question heading), which is not followed by item codes. For example: Question 1 of unit R409, labelled *R409Q01*, is an automatically scored item, whereas Question 3 of unit R409, labelled *R409Q03-0129*, is an item that will need to be coded by national coders.

Some of the units will be identical to units used in previous PISA studies, and will be used as anchor materials to link the new results to those obtained in previous assessments. Those “link” units will not need to be translated in countries that participated in the PISA studies from which the link materials were drawn. They may, however, need some revision in case any item was dropped from the unit, or minor changes had been introduced in the scoring instructions by the Consortium.

All countries will need to translate or to adapt the materials newly developed for the current PISA assessment. Countries that did not participate in previous studies will have to translate both the “new” and the link materials. A more detailed description of the test material to be translated can be found in “Material Preparation FT09”.

Many units will contain *translation notes* for use by translators. These notes obviously do NOT need translation.

- **A Student Questionnaire, a School Questionnaire and optional questionnaires.** The School and the optional questionnaires will be provided in two formats: with and without explanatory notes. Explanatory notes are included, in particular, to help with translation and adaptations. They do NOT need translation.

The Student Questionnaire will be provided in the two above formats but also in the two following formats: a) a version with ‘TA notes’ only, which contains just one occurrence of every question and which will be used for revision, translation and verification only and b) different forms that will be used in the Field trial.

In the Student questionnaire prepared for the Field Trial, some questions may appear twice (with only small differences in the question format or in the answer categories, for use in different rotated forms of the questionnaire). Care must be taken, when translating those “alternative” forms of a same question, to keep the alternatives identical except for the specific difference that has been deliberately included in order to assess which question format will be the most efficient to elicit student’s answers.

- **Manuals** intended for the *School Co-ordinator* (SC) and for the *Test Administrator* (TA) (in English only). These will be provided both as separate manuals (for countries where the school co-ordinators will be different persons from the test administrators) and as a single

combined manual (for countries where the same persons will be in charge of both school co-ordination and test administration).

Both manuals will include a number of boxes containing *Notes* intended for the NPM. These will draw the NPM's attention on aspects of the manuals that need national adaptations. The NPM notes obviously do not need translation.

In the *Test Administrator manual*, a special section called "Test administration script" will contain the instructions that the Test Administrator will have to read word-to-word to the students during the test session. Particularly strict equivalence against the source version is required when translating this section of the manual.

- **The booklet shell** (booklet cover pages, general instructions to be included at the beginning of each test booklet, etc.).

A number of requirements need to be respected in the layout of the cover pages of all PISA assessment materials. These are listed in the Material Preparation FT09 document.

The general instructions included in the students' booklets must be identical to those that are included in the *Test Administrator Manual*. Please make sure that these are translated only once and included in both the *Manual* and the booklets without any change.

Security requirements

Translators, as well as any other person who handles these documents (members of the national research team, of the National Committee of Experts, secretaries etc.) should note that the whole of the PISA material is under embargo, and therefore must be kept **strictly confidential**. A copy of the PISA Confidentiality Agreement is included in the NPM manual.

Translators

PISA translators should have:

- Perfect command of both the source language and the target language;
- Solid command of the domains covered by the test, in particular for the tests of Mathematics and Science. It may therefore be necessary to turn to pairs of different translators for the translation of each of the various tests;

Each translator should receive the following documents:

- A copy of the present *Translation Guidelines*;
- The source version of the materials (or sections of the materials) s/he is requested to translate;
- A copy of the *National Adaptations spreadsheet* related to the materials s/he is requested to translate;
- A timetable for the return of the translated documents;
- The address and telephone number of the NPM (or of the member of his/her staff in charge of the co-ordination of translation activities).

Reconcilers

In the PISA translation process, the role of the reconciler(s) is particularly important.

The main task of the reconciler will be to “merge” the two independent translations in such a way that the resulting national version is as equivalent as possible to the source versions, that all possible translation errors have been corrected; and that the wording is as fluent as possible – without the “translanese flavour” that is so often characteristic to translated material. This means that the reconciler’s role can *never* be limited to just selecting the “best” translation out of the two and briefly proof-reading it! First-hand translations *always* need extremely accurate and in-depth reworking.

In fact, the reconciler will have to read both the translations, to check each sentence against the source version, and to in-depth rework the translated text in order to make it as accurate and fluent as possible. The aim is to get the right balance: the translation must not be literal to the point that it sounds awkward, but neither should it deviate too freely from the source version, which would be likely to change in unexpected ways the functioning of the assessment items. When in doubt, reconcilers who can read both English and French may find very useful to have a look on both source versions, to have a more precise idea of the degree of freedom in translation that would be considered as acceptable.

A second important task will be for the reconciler to document all national adaptations that the translation team deemed necessary in the material. S/he will complete for this purpose the *Adaptation spreadsheets* received together with the materials, where all units included in the material will be listed, and all cases where national adaptations are expected by the test developers will already be highlighted. S/he will describe the adaptations made in each of these cases, and mention all possible additional adaptations next to the units where they occurred.

It is important that the reconciler has in-depth knowledge of the school system of the country (in particular for the school and student questionnaires). The knowledge of the characteristics of the school systems of other countries (particularly those where the source language is used) can be a valuable asset.

The reconciler, together with the NPM, will also be in charge of submitting the translated material and the proposed National Adaptations to the country’s *National Expert Committee* for discussion and possible edits, then to submit it to the Consortium for international verification.

Use of electronic files

It would greatly help that the translators and the reconciler be sufficiently familiar with MS WORD to be able to use a copy of the source version of the material as a basis for their translation (where they will overwrite the English or French text, replacing it with the text in their target language). This would be a very efficient and time-saving way of reproducing the layout, styles, fonts and graphic material from the source version, while reducing the risk of errors or omissions.

It is recommended that in the training session organised by the NPM to familiarise their translation team with the materials and with the PISA Translation Guidelines, some time be spent in instructions about the electronic aspects of the job. In particular, all translators should know (i) how to access the text boxes in the graphics included in the materials, in order to translate labels or captions; (ii) how to use the “track changes” function when revising the materials; (iii) how to use the spelling and grammar tools included in WORD; (iv) how to name or rename the electronic files to avoid confounding “old” and “new” versions of a same translated unit; (v) how to apply styles and formats; (vi) how to use the Excel spreadsheets where they will document any suggested adaptations.

Harmonisation

It may be useful that the translation team lists the headings and other recurrent expressions that occur very frequently in the materials, such as: “*Question*”, “*item*”, “*task*”, “*Explain your reasoning*”; “*Give*

an explanation to support your answer “*Show your work*”; “*Scoring*”; “*Full credit*”, “*partial credit*”, “*no credit*”; etc) and decide from the start how each of them will be translated.

Similarly, it would help that all translators are provided with a copy of their country’s typographic and style conventions, to harmonise their decisions in terms of use of abbreviations, punctuation, use of brackets, use of italics, etc.

Having copies of the most current Mathematics and Science textbooks used in national secondary schools may greatly help the translators in finding the appropriate terminology when dealing with PISA Mathematics and Science materials.

Consortium’s errata

The PISA test developers intend to limit as much as possible the changes or edits done to the source versions after their distribution to the NPMs for translation. However, at least *some* errors will most certainly be identified, requiring that errata and/or new versions of certain test units (with changes highlighted in “track change” mode) be sent to the national Centres during the translation process. **Please make sure that all those corrections are implemented in your national material.**

Many reviews of both the source versions and the national versions will take place throughout the process leading to the final version of the instrument. The risk of the team of translators and/or the reconciler working on a version that is not up to date, or of their forgetting to insert one of the required corrections, is quite real. It is therefore crucial to organise meticulous monitoring of successive versions throughout the process.

2. RECOMMENDATIONS ABOUT TRANSLATION TRAPS IN TEST MATERIALS

Many factors can undesirably bias the answers to items presented to the students in different languages. The list of advice provided below will help you to control some of the most commonly encountered among these. Take good note of the fact that they are to be regarded as *advice*. In practice, too strict an application of one or the other of the following recommendations may lead to an overly cumbersome translation, or it may occur that applying one recommendation would imply violating another. One should then give the priority to one or the other, with a view to choosing the lesser of two evils. So please mainly regard the following as an informative list of what MIGHT raise a problem with respect to equivalence between the source version and your version.

LAYOUT AND GRAPHICS

Respect the layout and presentation shown in the source document.

It often happens that a translated document is longer than the original. This is due to the characteristics of each language, and does not seem likely to have a significant effect on performance (for example, languages less concise than English often have a morphology and syntax which are more redundant, and overall the latter is likely to make up for the former in terms of reading difficulty).

However, sizeable changes must be avoided in the **page set-up** of the test — which would oblige the student, for example, to turn a page to read the questions, whereas in the source version text and questions appear side by side. If necessary, use a slightly smaller or bigger font than in the original, if this enables you to keep a page set-up true to that of the source version.

- In the PISA 2003 FT, three countries had a last minute page-numbering problem when printing some of their test booklets, which resulted in all odd pages becoming even pages and all even pages becoming odd pages. This destroyed the layout of many of their test units – where the original stimulus was supposed to appear on an even page, and the corresponding items in the opposite odd page. Due to the incorrect page numbering, students had to turn the page to find the questions corresponding to most of the stimuli. The difficulty of those items appeared in the analyses to have been significantly affected.

Think about the motivation of students. Presentation that is denser, big blocks of text, which are less spaced out, printing which is more careless than the original may encourage some of the weaker students to turn the page without answering the questions.

Respect the **typographical cues** provided in the original materials:

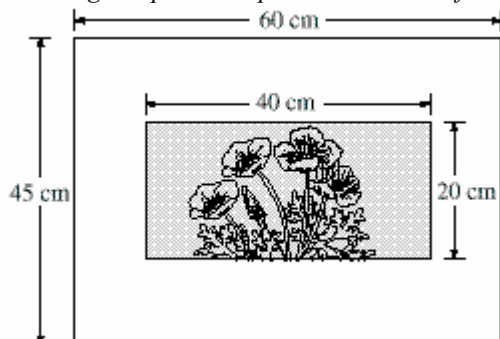
- Make sure that **titles**, **subtitles**, possible **numbering of lines** and paragraphs, and **numbering of items** are the same in your national version as in the source version.
- Check whether the **instructions** given to the students are consistent with your final layout (e.g. “Consider the diagram *below*”, “In the box *above* you are given a formula”, etc.).
- Make sure that any word or expression that is emphasised in the original text by using **bold characters** or **underlining** (or both) has exactly the same emphasis in your target version.

Check the rendering of illustrations and graphic elements.

Check that no alteration occurs while importing these elements in your translated file, or during printing.

- In the following example, taken from the IEA/TIMSS survey, too pale a rendering of the graphic element of this item when printed can make the grey tint which is used as background for the flowered picture disappear, making distractor D much more attractive than in the original version.

A rectangular picture is pasted to a sheet of white paper as shown.



What is the area of the white paper not covered by the picture?

- A. 165 cm^2
- B. 500 cm^2
- C. 1900 cm^2
- D. 2700 cm^2

Do not forget to translate the **text boxes** contained in these graphic elements (in particular the keys, captions, scales, units of measure). Many of the bugs observed in previous PISA studies resulted from small errors in the text boxes included in some of the graphics:

- In R076 Iran Air, one of the items asked what would be the check-in time if you wanted to fly from Tehran to Paris on a Tuesday. But the students in one of the PISA countries could not answer correctly this item, since in the corresponding table the abbreviation for Tuesday had been wrongly replaced with the abbreviation for Thursday.
- In R040 Lake Chad, various national versions had problems with the abbreviations BC and AD. In addition, some translators wrote Before Christ in full in the stem while leaving the abbreviation in the table, which caused a bad functioning of the item related to that table.
- Many problems were observed with the size of the graphic that was used as a stimulus in M305 Map. For the student to be able to correctly estimate a distance using the scale at the bottom of the graphic, the dimensions of the map in their booklet should have been 11cm x 15cm, and the length of the scale should have been 5cm. However, in almost all of the national versions submitted for verification, the dimensions differed substantially from the requested measures.

The NPMs will be requested to submit to the Consortium sample hard copies of some of the pages of their booklets, so that the quality of their printed materials can be checked in advance of the dispatch of materials to the schools.

LINGUISTIC DIFFICULTY

Avoid complicating or simplifying the vocabulary and the syntax. This applies to both the text materials used in the stimulus, and the wording of the items (both in the stem and in proposed responses).

Length of sentences

It is no coincidence that average word length and sentence length almost always appear among the indices of complexity used in readability formulas, whatever the language for which the formulas have been developed. Longer words tend to be less frequent, more technical and/or more abstract than short words. The basic vocabulary of a language (the most frequent and easiest 1500 to 3000 words of a language) is, more often than not, made up of very short words. Long sentences often contain many subordinate clauses and/or embedded clauses; the word order and the syntax in those sentences are usually more complex than when the passage is made up of two or three separate sentences rendering the same content.

- In the following example, the *b* version is more complex than the *a* version, not only because of its vocabulary (term: *relief variation*) but also because of its syntax (subordinate clause):

- It was easy: the route of the marathon consisted of few important differences in height.*
- I found that easy, inasmuch as the route of the marathon consisted of few important relief variations.*

Items where the student is asked to **quote a sentence** in the text that contain some specific information may be seriously affected when, for example, the original sentence was embedded into another sentence, or split into two new sentences.

“Common” vocabulary vs “scientific” terminology

As a general rule, translators should try their best to avoid translating difficult words in the text by using easier words, or paraphrases that use more common terms. Conversely, common language terms used in the source materials should not be translated into more ‘technical’, or ‘scientific’ or ‘literary’ expressions.

- In S434 Radiations, do not translate “*Our skin gets red*” into “*Our epidermis gets red*”. Not only the word “*epidermis*” is much less common, but in this particular case it is also inappropriate, since UV radiations may affect all layers of the skin (i.e. both *epidermis* and *dermis*).

“Common language” vs “scientific terminology” represents a particularly important challenge when translating the PISA Mathematics and Science test materials, due to the very nature of the study, which aims at assessing students’ **literacy** (i.e., how well can the students apply *in life situations* their mathematical and scientific knowledge and skills?).

Natural languages often use less than precise words to describe mathematical concepts, or chemical, physical or biological phenomena – sometimes ‘common’ words may even convey usual misconceptions, such as ‘*weight*’ used to mean ‘*mass*’, ‘*heat*’ used to mean ‘*temperature*’, ‘*circle*’ used to mean ‘*disk*’, etc.

- For example, when discussing about diets, people would say: “*I’m watching my weight*” rather than “*I’m watching my mass*”.

In a number of PISA test units, where every day’s life contexts are used, it is recommended NOT to change any such ‘common language’ expression into the ‘strict’ terminology:

- For example, in M836 Postal Charges, the items are about the cost for dispatching an 80 g parcel. A table indicates that postal charges are 0.46 zeds for “*weights up to 20 g*”; 0.69 zeds for “*weights from 20 to 50 g*”, etc. In this context, it would appear as weird to use “*mass*” rather than “*weight*”.

Some other units have a more formal Mathematics or Science context, where accurate terminology is needed:

- For example, the context in S133 Research is a scientific investigation, thus appropriate scientific terminology was used. A translation note indicates that “*Throughout this unit the words weight, weigh, mass should be used as shown.*” In some occurrences, both ‘common language’ and ‘scientific’ terms are used: “*The balance will be used to weigh (find the masses of) the paper clips*”.

Sometimes, even in formal Mathematics or Science items, the English source version uses ‘common language’ expressions, which may appear as less acceptable in other school systems. When in doubt, please check with your national experts on acceptable or unacceptable use of natural language in the materials, and contact the Consortium for any query related to these terminology issues.

- For example, in M161Q1, the English stimulus read: “*Triangle PQR is a right triangle with right angle at R. The line RQ is less than the line PR. M is the midpoint of the line PQ and N is the midpoint of the line QR. S is a point inside the triangle. The line MN is greater than the line MS.*” However, the word “*line*” was replaced with “*segment*” in the French source version, as requested by national experts from French-speaking countries.

In a number of cases, translation notes have been added in the materials to help the translators with terminology:

- For example, in S410 Sickle Cell Anaemia, a translation note reads: “*Throughout this unit, ‘everyday language’ should be used in preference to medical terminology. For example (in French): “anémie falciforme” rather than “drépanocytose” and “globules rouges” rather than “hématies”.*”

Identifying the exact national name may sometimes be hard when the source version uses common English or French names of animals or plants, since these common names often cover a range of different species. Translation notes have been added to provide translators with the **Latin name used in scientific classifications**. This will help you retrieving the correct equivalent in your national dictionaries or encyclopaedias.

- For example, in S408 Wild Oat Grass, a translation note reads: “*The scientific name of ‘wild oat grass’ is Avena fatua*”.

Affective vocabulary

Test items based on nuances of vocabulary often raise difficult translation problems. It is rarely easy to find terms in one’s own language having exactly the same connotations as those in the source language.

- One of the questions in R111 Exchange read as follows:

Basically, the tone of this article is

English Version:	French Version 1:	French Version 2:
<i>A. stern</i>	<i>A. sévère</i>	<i>A. dur</i>
<i>B. cautious</i>	<i>B. prudent</i>	<i>B. circonspect</i>
<i>C. humorous</i>	<i>C. humoristique</i>	<i>C. comique</i>
<i>D. encouraging*</i>	<i>D. encourageant*</i>	<i>D. encourageant*</i>

We can expect that the pattern of answers to this item will not be the same in the French version 1 as in version 2, because of the slightly less attractive character of distractors A and B in version 2.

Idioms and metaphorical expressions

Too literal translation of idiomatic expressions is a frequent source of awkwardness in translated materials, and probably the most common cause of “translinese” flavour. Failure in finding a fluent equivalent expression in the target language may result, in extreme cases, in unintelligible passages or sentences.

- In S444Q6 White Stork, a too literal translation of the sentence “*I’d drive a little further if it made migrating birds safer*” would be meaningless in French (“*Je conduirais un peu plus loin si cela pouvait rendre plus sûre la migration des oiseaux*”). In French, the correct idiom for “*I’d drive a little further*” would be “*Je roulerais quelques kilomètres de plus*” (i.e., “*I’d drive a few more kilometres*”)

As far as possible, make sure that the metaphoric meaning of English expressions like “*to see to it*”, “*in small steps*”, “*to disappear in the thin air*” is adequately rendered through similar idioms in your language.

- In R081 Graffiti, item 2, the source versions used the word “*costs*” with both a “restricted” meaning (“*financial costs*”) and a more “extended” meaning (“*damages caused to public buildings and to the reputation of young people*”). No proper corresponding word could be found in many national versions, which used instead “*damages*”, “*problems*”, “*burdens*” etc. A number of DIFs (Differential Item Functioning) were observed for this item.

Abstract words or turns of sentences

Do not unnecessarily modify the degree of abstraction of the sentence by using nouns where the author uses verbs, or vice versa:

- In the following example, *a* will be more difficult than *b*:
 - The presence of humour in a violent scene can increase the chances that viewers will imitate or learn aggression.*
 - When humour is present in a violent scene, viewers are likely to imitate or learn aggression.*

Particularly difficult-to-translate English words

Two English words, very frequently used in the PISA Science materials, are particularly difficult to translate:

- **Evidence** as in “*Give evidence from the table that supports the idea that this story might be true*”.
A number of languages do not have the exact equivalent of the word “*evidence*”, and translators are often in trouble finding an appropriate word: “*Give an element from the table...*”, “*Quote some data from the table...*”; “*Find a proof in the table...*”.
In a sentence such as “*Explain whether you think there is any evidence for concern*”, it will probably be quite difficult to maintain the conciseness and the vagueness of the English expression “*evidence for concern*”. The translation would need to be something like: “*Explain whether you think there is any element that might support these concerns*”.

- **Idea**, when used to mean “*hypothesis*”, or “*conjecture*” or “*theory*”, as in “*Give evidence from the table that supports the idea that this story might be true*” or in “*By observing the behaviour of the flies, what was the main idea Hoy was testing?*”.
In a number of languages, when using the word « *idea* » instead of « *hypothesis* », the meaning may be less neutral than in English. It rather will have some slightly negative connotation (an “*idea*” in this case tends to be a “*poor*”, or a “*very uncertain*” or a “*cranky*” hypothesis).

Other difficult words or expressions are:

- **Insight**
- **To value**
- **A range of**
- **Pattern**

Translators are also often in trouble with a number of English expressions, e.g.:

- **More likely than not** (e.g., “*More likely than not, there will be an earthquake in Zed City at some time during the next 20 years*”. This is a simple and efficient expression in English, but in most other languages two or three lines will be needed to express the same idea.
- **In which x of the y situations** (e.g., “*In which two of the four situations below...* ”). Such expressions are often not translatable without totally changing the sentence: Many languages would have to turn a question such as “*In which two locations were most eggs laid?*” (S425 Penguin Island) into the equivalent of “*What are the two locations where most eggs were laid?*”

Active vs passive turns of phrase

If possible, avoid, translating an active turn of phrase in the original by a passive one, or vice versa.

- In the following example, the *b* version increases the difficulty of the sentence, not only by the use of the passive form, but also because this version (as is often the case) uses as the subject an abstract word (*problems*) instead of an animate term (*families*):
 - Many Russian families traditionally present their children with hundreds of problems of this type.*
 - In Russia, traditionally, hundreds of problems of this type are presented to children by their families.*

Negations

Be very vigilant during the translation of passages containing negations, especially when double negations are involved: the latter significantly increase the difficulty of the sentence.

- In the following example, the *b* version is more difficult than *a*, as a result of the replacement of a negative term (*harmless*) by a double negation, syntactic and lexical (*is not harmful*):
 - This does not mean that the violence in cartoons is harmless.*
 - This does not mean that the violence in cartoons is not harmful.*
- A typical example of flaw caused by a negative expression was Question 6 in S268 Algae. When translating the stem of this item (*Which of the following best explains why green algae are most often restricted to the top 100 metres of the Ocean?*) one of the translators turned it

into a negative expression (“...*why, in the ocean, green algae are almost never found below a depth of 100 meters?*”). This caused a DIF in that national version.

Reference chains

Modify as little as possible the reference chain(s) which is/are contained in the passage. A reference chain is the set of occurrences in the text where the same character or the same notion is alluded to, often with the help of various linguistic tools (pronouns, synonyms, etc.).

- The following passage, from an article on violence on TV, is quite complex. It contains three of these chains, one relating to the notion of violence (V), the other to the notion of punishment (P), and the third referring to the young viewer (Y):

“If the punishment (P1) for violence (V1) is delayed until the end of the program, this deterrent (P2) may go unnoticed by a young child (Y1). Punishment (P3) must occur in the same scene for a younger viewer (Y2) to connect it (P4) to the original aggressive behaviour (V2) which gave rise to it (P5).”

In a case of this type do your best to respect the nature of the elements of reference:

- Repeat the word if the author repeated it (*punishment* in P1 and P3).
- Use a synonym if the author used one (*violence* in V1 / *aggressive behaviour* in V2; *punishment* in P1 and *deterrent* in P2).
- Use the combined repetition and synonym when this is the case with the author (*young child* in Y1 and *younger viewer* in Y2).
- Use pronouns where the author uses them (P4, P5).

In many languages, repeating words in a text is less readily accepted than in English, thus translators often tend to use synonyms rather than repeating same words. Note, however, that the text above will become more difficult, for example, if you choose to avoid the repetition in P3 by using a synonym (*sanction* instead of *punishment*), or in Y2 by using a reference by position (*the latter* instead of *younger viewer*).

Generic articles

The English ‘indefinite’ article (“*a cell*”, “*a molecule*”) is often used to convey a generic meaning, such as in the sentence: “*The milk of a cow, a wolf and a human do not contain the same amounts of proteins*”. This differs from other languages, where such articles are normally used to refer to an *individual* cow, or wolf, or cell, or molecule. To prevent misinterpretation, especially when translating Science materials, please turn these expressions in the way that is used in your language to express the same “generic” idea.

- For example, the sentence “*A tidal power station uses a different method to generate electricity from many other power stations*” needs to be turned into plural when translated into French (“*Tidal power stations use a different method from other power stations*”).

COMMON PSYCHOMETRIC TRAPS

When translating the items, avoid involuntarily providing clues that direct the student towards the correct answer or make a wrong answer more attractive.

Do not modify the length of key and distractors in multiple-choice items

In multiple choice items, make every effort not to unnecessarily modify the respective lengths of key answer and any of the distractors. Long answers are more attractive than short ones; therefore, the item might become easier in your version if the correct answer is more elaborate (in relation to other answers), than in the source version. On the other hand, the item might become more difficult if an incorrect answer stands out from the others because of its length more than it does in the source version.

Other factors, like differences in sentence structure or differences in style or tone of the translated text, may also cause one of the proposed answers stand out from the others more than it does in the source version.

Be careful with truncated items.

It may be quite challenging to translate in certain languages those item stems that are “truncated sentences” in the ENG or FRE source versions (i.e. multiple choice items where the stem is an incomplete sentence, and where the student must select the most appropriate answer among four or five proposed “endings”). Make sure that all proposed endings have in your language the same syntactical match with the incomplete stem; otherwise the item will measure grammar skills instead of the intended process. If this is not possible in your language, use complete sentences in both the stem and proposed answers.

Respect literal matches

If the stem of the item (or some of the distractors) literally reproduces expressions contained in the text, take care that the same applies for the translation. If, on the contrary, the author of the item uses a different formulation from that of the text (synonym, indirect allusion) do not simplify the student’s task by using words of the text or derivatives of the same word in the stem or distractors. This aspect deserves special attention, since it is an extremely frequent cause of bad item functioning.

- In M150 Growing Up, the source version contained a literal match between the stimulus text (“*In 1998, the average height of both young males and young females in the Netherlands...*”) and one of the captions in the accompanying graph (“*Average height of young females in 1998*”). A DIF occurred in a national version where the translation of the stimulus text did no longer match the caption in a literal way (“*maids*” instead of “*young females*”).
- Flaws due to the translation of literal matches into synonymous matches also occurred in R088 Labour Force, question 4, when the national version of the stem did not contain exactly the same expression as one of the boxes in the tree diagram (“*Not in Labour Force*”), or in R102 Shirts question 5, where the translator did not use in the stem a word-to-word parallel for one of the characteristics mentioned in the table (“*No iron*”).
- Question 2 of S404 Food safety asked: “*Why are temperatures below 4 °C recommended for safe food storage?*”, and the proposed answers were as follows:

- A Bacteria die at these temperatures.
- B Bacteria stop multiplying at these temperatures.
- C Bacteria become frozen at these temperatures.
- D Bacteria multiply very slowly at these temperatures.

The proper functioning of this item will be highly dependent on whether your national version contains the appropriate matches between the relevant text boxes in the graphic used as stimulus and responses B and D:

“Warming temperatures prevent multiplication of most bacteria; and Chilling temperatures restrict multiplication of bacteria. Some food poisoning bacteria may increase in numbers very slowly.”

Provide all the information contained in the item.

- For example, if the item says: *“On line 20, the author uses the expression...”*, do not forget to check whether this passage is in fact on that line in your version. A bug was observed in S132Q04 Research, in a national version where the line numbers had changed unexpectedly in the final printing of the booklets. This made the item more difficult for the students: *“In lines 13 to 15, Hoy’s experiment was described (...) What was the main idea he was testing?”*
- Do not omit any important word, particularly in Mathematics items. Omission of the expression *“on average”* in M150 Growing Up, question 3 caused a DIF on one of the national versions (the English source version of the item read: *“Explain how the graph shows that on average the growth rate for girls slows down after 12 years of age”*).

Check all of the “Which of the following...?” questions

In multiple-choice items, respect the indication contained in the source version on whether *only one answer* or *more than one answer* are asked of the student. Pay special attention, for example, to all items where the English stem reads as *“Which of the following...”*. Most of them do not contain any information at all on whether the student should circle *only one* or *several* of the proposed statements. This ambiguity MUST be conveyed by the translated stem, which is not an easy issue for languages where the interrogative pronoun *“Which”* carries an intrinsic plural or singular morphologic mark.

- In S133Q04 Research, the English stem reads *“Which of these other pieces of information would you need to know to decide which thread would be the best to use for a necklace?”*. This question calls for either one or more than one answer. In the French source version, the stem had to be turned into a *“singular/plural”* expression in order to convey the same meaning (*“Laquelle ou lesquelles des informations suivantes devez-vous connaître pour décider...”*).

Items like the one in the above example were very often mistranslated in previous PISA studies when the national version was developed using only the English source version. No errors or fewer errors were observed in national versions that used double translation from *both* the English and French source versions.

Try to respect the order of information in the question stems

The order in which the author has presented the various pieces of information contained in the stem of an item is often important. Try to reproduce that order insofar as possible.

- When the stem is long, you may occasionally observe that the author has privileged certain elements of the question by placing them either beyond or right at the end of the phrase. Try to highlight the same elements in your national version.

All conventions with respect to word order significantly differ from one language to the other. However, various stylistic devices often make it possible to enhance this or that segment of the question. Teachers generally prove to be outstanding judges as far as the way of formulating the questions of a test is concerned. If some members of your national panel are teachers, draw their attention on that point and ask them to be particularly aware of it: they will probably assist you in improving those items whose formulation is somewhat awkward or unclear due to the translation.

Small details are sometimes important

Mistranslations sometimes occur in extremely “easy” sentences, just because they were so simple, thus the translator paid less attention!

- In S409Q01 Cleaning Water, the meaning of the sentence “*It is important to have a source of good drinking water*” would be seriously changed if it is translated into: “*It is important to have a good source of drinking water*”.
- In S413 Plastic age, do not translate: “*Should the following items be made out of this material?*” into “*Can the following items be made out of this material?*”. The meaning would not be the same. Many items *can* be made out of a material, not all of which *should* be made.

It may occur in a multiple-choice item that some of the alternatives proposed only differ by one **key detail**. Be particularly vigilant: it does occur that, during the course of the translation procedure, the element that makes the difference between both responses may be toned down, thus impairing the item.

OTHER

Pay attention to scoring instructions.

More often than not, when an item proves easier or more difficult in one language than in another one (for students of equal ability), this is because the formulation has somewhat modified the strategy required to find the correct answer. *Scoring Instructions* can help the translator understand the nature of the item, because they contain information about the strategy that the test authors wish to assess.

Respect all translation notes.

A number of *Translation Notes* have been added into the material by the test developers whenever they thought it necessary to draw your attention to an important aspect. Scrupulously conform to the instructions contained in these notes. Depending on the particular case, their aim is:

- To ask the translator to imitate as closely as possible one or the other of the stylistic characteristics of the source version (for example, a somewhat outdated expression in a literary text, or the familiar tone of a letter).
- To point at the cases when strict scientific or mathematical terminology should be used and cases when every day’s language is preferred.
- To draw the translator’s attention on particularly important cases of literal/synonymous matches between item stem and stimulus.
- To point out the aspects for which translators are explicitly requested to enter a national adaptation. Notes of this type are particularly frequent in questionnaires and Manuals.

- To indicate a particular case where the translated text must remain strictly true to the original and where even a slight deviation could compromise the efficient functioning of one or several test items.
- To warn translators against specific translation traps.

3. NATIONAL ADAPTATIONS OF TEST MATERIALS

Considerable efforts were made during the development of the material to prevent the content or presentation from abnormally favouring or putting at a disadvantage students of certain countries or certain cultures. The texts and documents that were selected are of various origins; they were selected by taking into account as much as possible the likely common interests and concerns of 15-year-olds everywhere in the world. Moreover, the National Centre of each participating country was asked to have its committee of national experts evaluate the texts and documents, and to indicate to the International Centre those items that the national committee judged not to be well suited to the culture of the country.

It should be said that to reach perfect cultural neutrality for the tests is probably not possible, nor even desirable. How literate would the youth in our modern societies be if their competence did not allow for insight into cultures other than their own?

Therefore, the general principle will be to restrict national adaptations to those cases in which they are truly required, to avoid needlessly disconcerting the students by confronting them with expressions or concepts that are much less familiar to them than to students in other countries, but without affecting the very substance of the text or the items.

However, the majority of the PISA stimuli are texts that the student would normally be likely to encounter in daily life when reading books, newspapers, magazines etc. Make it a rule to have national adaptations only in those cases where it would appear usual to make them in your country, should the text be published in a magazine or occur in a school textbook.

WHAT SHOULD BE ADAPTED, WHAT SHOULD NOT BE ADAPTED?

Never adapt the format of an item.

Items that are open-ended items in the source version should never be turned into multiple-choice items in your national version, or vice-versa. Never change the order or the content of headings of responses presented as columns in tables.

For example, do NOT invert “Yes/No” or “True/False” categories into “No/Yes” or “False/True”.

Similarly, do NOT change the order of answer categories “Strongly agree, Agree, Disagree, Strongly disagree” into “Strongly disagree, Disagree, Agree, Strongly agree”, or the reverse.

Do not include explanatory notes nor additional instructions

In some cases, you may be tempted to add a footnote to explain the meaning of a particularly difficult word (or to include a brief explanatory parenthesis next to a difficult word in the text). Please note that such adaptations must imperatively be submitted for approval, and that virtually none of them was considered as acceptable by the test developers in previous PISA assessments.

In the same order of ideas, avoid adding extra instructions (e.g. to explain where the students should write their answers, or to indicate that *two* answers are requested). Again, such adaptations are usually considered as not acceptable in PISA.

Adapt mathematical symbols, abbreviations and formulas.

Adapt the abbreviations of units of measurement and the symbols used for operations, with particular attention for all information provided in graphs (captions, units of measure, coordinates of the axes, graduations etc.)

- For example, it can be quite disconcerting for students to read, in a question relating to angles:

$\hat{\quad} \quad \hat{\quad}$
Compare ABC et BCD,
 if the notation used by their teachers is:
Compare $< ABC$ and $< BCD$.

Pay particular attention to the following:

- Units sometimes have differing written forms (“cl, ml”, vs “cL, mL”).
- The symbol used for multiplication can be (.), or (*) or (x).
- Approximation is sometimes noted (~) or (+/-).
- Hours are usually noted as 7:20 AM, 4:30 PM in some countries, while other countries would use 7h20 and 16h30.
- Decimal period (e.g.: 2.5) must be replaced with decimal comma in many languages (2,5).
- The captions, units of measure, coordinates of the axes must practically be translated every time.

Adapt mathematical and scientific vocabulary.

Mathematical and scientific language is far less “universal” than it is often said to be. For example:

- What is called a *Bunsen burner* in some countries is called a *hot plate* in others.
- In certain languages 0°C is the *freezing point of water*, while in other languages it is the *fusion point of water*.
- Diagrams named *pie charts* and *scatter plots* in certain countries are called *cheeses* and *clouds of points* in others.
- A *mathematical expression* may be called a *formula*, an *expression*, and an *equation* in different school systems.

From this viewpoint, it is essential for the translation team to include members who are competent in mathematics and science, or for its work to be carefully checked by specialists (in particular, math and science teachers teaching at the level attended by the majority of 15-year-old students).

Check the use of scientific and mathematical terminology vs “common language”.

Expert advice will be needed, in particular, to know whether you should adapt or not the “real life” vocabulary in a number PISA units, where mathematical or scientific concepts are expressed in the English source versions through non-technical, “every day” language. While in many countries this would reflect the actual language of 15 years old students, it may happen that certain expressions are considered as too “inaccurate” or too “careless” to be acceptable in certain school systems.

- Check, for example, whether in your Mathematics classes it would be acceptable to use the word *round* instead of *spherical* (“Round stones” in M454 Flowerbed), or *circles* instead of *Venn diagrams* (M478 and M469).

Check “common language” quantitative expressions

Be particularly careful in translating all “common language” quantitative expressions when translating Mathematic items. Many of these expressions are less equivalent than they seem to be across languages. Differences occur, for example, in the way various languages count certain things, and on whether the first (or last) item should be included in the total:

- In French and in English(USA), a “*three storey building*” is, actually, a building with four levels, since the first level (ground floor) is not included in the sum. In many other languages the equivalent expression would be “*a four-floors building*”.
- “*Students above the age of 12*” is a tricky expression, since in certain languages this includes students aged 12 or more, while in other languages students should be at least 13 years old to be included.
- Similarly, in items where intervals are specified (such as “*Temperatures from 60 °C to 74 °C*”, or “*From –12 °C to 4 °C*”), the boundary values (e.g. 60 and 70°C) are considered as included in the interval in certain languages, but as excluded in other languages.
- In M503 Copy Machine, a too literal translation of the expression “*The number of copies would be twice as much*” would result in certain languages into three times that number (i.e. the initial number, plus two times the same number).

Do not adapt the “zed” currency

The PISA materials (particularly the Mathematics and Science materials) sometimes refer to a fictional country (*Zedland*) or a fictional town (*Zedville*), where a fictional currency is used: the *zed*. Please **never adapt** these names to national location names or to real currencies, because the adaptation of currencies would be likely to seriously affect the arithmetic demands of the item. The use of these fictional numbers will guarantee that students from all countries are faced with the same computational requirements.

In contrast, currencies other than *zeds* can be adapted to your national currency in most of the Reading materials, where no computational task is asked of the student. Most often, a translation note will indicate explicitly what aspects can be adapted or not.

Names of persons and locations can usually be adapted

In most of the test units, you can use names and locations that are familiar to the students in your country.

- It may help, however, that you use national names that start with the same capital letter as in the source version. For example, in M407 Sharing the bill, the three friends mentioned in the stimulus were deliberately named *Amy*, *Betty* and *Cindy*, so that the student can use in his/her computations the letters A, B and C as abbreviations.
- Having the same starting capital letters as in the source version will also help avoiding errors when translating complex lists of names, or distractors in multiple choice questions (for example, in the table indicating which pairs of players will play the various matches in Rounds 1 to 3 at Tables 1 or 2 in M421 Table Tennis Tournament).

Some names should NOT be adapted

In **literary texts** the names of characters, places, currencies are part of the universe described by the author; therefore, the only acceptable changes are those slight spelling adaptations that are generally used by your country's translators.

- It would be ridiculous to translate “*the brothers Karamazov*” into “*the Thom(p)son brothers*”.

Adapt the **biographical names** of famous characters **only** in case they are usually adapted in your country.

- Both in English and French the name of the Portuguese explorer *de Magalhães* would be adapted as *Magellan*.

Do not adapt the **names of institutions or agencies** (unless there is a well-known national version of the name); rather leave the name of the institution or agency in the original language and add its translation in square brackets if some of the information it contains is deemed important for understanding the stimulus.

Similarly, please leave in the original language any references presented in a note or at the bottom of the text (such as **author and title of the text**). Add a translation in square brackets if needed).

In **geographical maps**, translate **only** those proper names, for which it is customary to have them translated in your country's most recent atlases. Leave all other names in the original language.

In **newspaper articles**, only translate or adapt what you would expect to see translated or adapted if the article were published in one of your country's daily newspapers or periodicals. As a rule, the spelling of proper names used in the country's most popular newspapers or weekly magazines should be followed.

- Current English newspaper articles would write, for example, *Beijing* and *Mumbai* rather than *Peking* and *Bombay*.

In **functional texts** (advertisements, instruction manuals, catalogues etc.), make sure that you comply with the *Translation notes*: they generally specify which aspects can be adapted – to bring the document “closer” to advertisements, instruction manuals, catalogues etc. such as those which the student is likely to encounter in his daily life – without making amendments that could be harmful to proper item functioning.

Check that adaptations are consistently applied

It often happens that adaptations are entered in a unit in an inconsistent way. If you change a proper noun or the name of a currency, do not forget to do it **every time** these terms appear in the text or in the items, or in any illustration accompanying the text. If decimal commas need to be used in your national version rather than decimal periods, please check **all** decimal numbers in your materials.

Think of possible other adaptations that may be needed in your country.

The most common adaptations are well known by the teams responsible for the development of international tests. Most of them will be described in the *Translation notes*. However, not everything can be provided for. Your team of translators will probably be confronted with new problems requiring deviations from the source versions.

- For example, in countries where the school week goes *from Monday to Saturday*, it would be necessary to modify the headings of a possible school timetable, which would go *from Monday to Friday* in the source version. Of course, one must be careful that this modification does not affect the item(s) in any way.
- A number of Mathematics items ask for numeric responses, such as in M803 Labels: “*How many of the shipped cans are likely to be damaged? Answer:cans*”. In Slavic languages, one of two different forms of plurals will be used for the word “*cans*”, depending on whether the expected numeral is less than 5, or 5 or more. The prompt would then need to be adapted (e.g., into “*Number of cans:*”) in order to avoid giving the student undesirable cues on the magnitude of the expected answer.

If you happen to have any queries related to translation or adaptation, please do not hesitate to e-mail them to the PISA translation referee (see contact address *Material Preparation FT09 document*).

Document all national adaptations made in the test materials.

It is mandatory that **all** of the adaptations done to the materials (whether required by a translation note or recommended by one of the directions above, or added by the translators or reconcilers for whatever reason) be documented in the *Test Adaptation Spreadsheets (TAS)* and submitted to the Consortium.

What about the PISA test material that was submitted by your country?

Some of the material included in PISA may have been submitted by your country. **Please never use your original submission without carefully comparing it against the source versions.** Test developers may indeed have introduced changes in the text or in the items that had been submitted. In such cases, the same modifications will have to be added to the original when including it into your national version.

If a **literary text** has been submitted by your country, in principle the English and the French source versions are scrupulously faithful to the original. Please report to the Consortium any deviation you might observe and which could be liable to affect the equivalence between the original version and the ENG or FRE source versions. The latter will be corrected if needed, so that **you can use the author’s original text without having to amend it.**

If a literary text included in the PISA test material stems from an author in a different language than yours and if there exists a version of that text translated in your language, you may use that translation, provided that:

- Permission has been granted by the owners of the copyright and the references are duly quoted; and
- The translated version’s equivalence with the English and French source versions has been carefully verified. In case of diverging versions, the PISA source version will be the reference.

4. TRANSLATION AND ADAPTATION OF QUESTIONNAIRES.

Translation of Questionnaires and Manuals pose slightly different problems. The aim of these instruments is not, as in the tests, to assess competencies. During the process of translating an international *test*, it is necessary to ensure that the difficulty of the questions is equivalent across countries, so that students with an equal competence have the same opportunities of answering them correctly. In order to properly translate a *questionnaire* or a *Manual*, the key issue is to be perfectly understood: the questions about the educational context and the instructions on the implementation of the survey must be as **transparent** as possible and have the **same meaning** in every participating country.

Keep in mind that some respondents will misunderstand anything that can be misunderstood

The smallest ambiguity in the formulation of a question can make the interpretation of the answers difficult. Try to anticipate the problems, which could arise in your country, and to formulate the question in a way that will prevent them.

- For example, the notion of *language spoken at home* may lead to confusion. In the IEA/RLS survey, no less than 26% of the young Italians answered that they *did not speak Italian with their parents*, probably because many of them speak one of the many Italian *dialects*.

Take care with the vocabulary and the syntactic turns of phrase used. Remember that the student questionnaire must be well understood even by those students whose reading skills are poor.

- An American study on the reliability of answers to questionnaires (Levine & al., 1998) showed that, among the grade 8 students questioned, some thought that “*vocational program*” indicated a “*summer school*”.

Avoid translating complex English questions into “translanese”

English is probably one of the most flexible languages in the world for the construction of interrogative sentences. Any English adjective or adverb can be used as a basis for a wh-phrase (“*How old...?*”, “*How strong...?*”, “*How valuable...?*”, “*How often...?*”, “*How likely...?*”, and so on), while many other languages have a much more limited list of interrogative words or expressions.

In addition, the English syntax allows interrogative expressions to be embedded in other interrogative phrases (e.g., “*Which of the following factors determine whether students are admitted to your school?*”), while these complex expressions are not always possible in other languages. This is probably one of the reasons why questionnaire materials used in international studies often suffer from awkward or contorted translation of the question stems.

Please do your best effort to keep the wording of question stems as simple and fluent as possible in your national language. If needed, you may want to split very complex questions into two different sentences, rather than producing a too literal translation, which might be confusing or difficult to understand for the respondent.

Do not change the question layout

Length of translated materials can be a problem in questionnaires just as in test materials. It may happen that one-page questions in the source versions are spread over two subsequent pages in the

translated version. In some cases, this can be solved by slightly changing the width of some text columns, or by slightly reducing the font, in order to keep the question on just one page as in the original text.

If this solution is insufficient and you really need to use two pages for a question, please never forget to replicate on the top of the second page the column headings containing the question answer categories, so that the respondent doesn't have to flip back and forth to remember the meaning of each box or letter he has to tick.

Like in test materials, please make sure that any word or expression that is emphasised in the original text by using **bold** characters or **underlining** has exactly the same emphasis in your target version.

Pay attention to the **lettering** in the questions. In particular, never change into capital letters A, B, C, etc. the letters a), b), c) etc. that appear on the left of the items. Neither should you do the reverse: do not change A, B, C into a, b, c. Capital letters are used in the PISA questionnaires to indicate the various *answer categories* (usually the respondent must tick or circle only one of A, B, C, etc. Lower case letters are used as *item codes* (the respondent usually has to provide an answer for each of the items a), b), c) etc.

For the sake of **uniformity** across questionnaires, it may help to list the instructions that appear several times in the questions (such as “*Please circle one answer only*”, “*Please circle the appropriate answer for each statement*”) or the most common answer categories (such as “*Strongly disagree/Disagree/Agree/Strongly agree*”, or “*Never/ 1 or 2 times a year/3 to 5 times a year/Once a month/More than once a month*”). Translators should decide since the beginning on how each of them will be translated. This will prevent possible errors or inconsistencies across the materials, particularly if more than one person is in charge of the translation.

Pay attention to omissions that may change the item format or the item scope

Omitting the italicised instructions accompanying questionnaire items (such as “*Please tick only one box in each row*” or “*Please circle as many boxes as apply in each row*”) may result in ambiguities about the type of answers required, which would seriously affect the distribution of responses.

Never omit from the stem of the question any word or expression that gives the respondent some indication on the degree of approximation required in the answer, such as “*Approximately, what percentage of...*”; or “*In average, how many...*”, or “*In a typical school week...*” or “*During your lessons, do you usually...*” or “*During the last full school year...*”. Omitting such expressions is likely to lead to inaccurate answers or/and to an increased rate of omissions.

- In PISA 2003, STQ33 asked about the time spent each week by the student in “*studying and doing homework outside regular Mathematics classes*”. In the source version, the word “*Mathematics*” was repeated in each of the items included in the question (“*Homework set by your Mathematics teacher*”; “*Working with a Mathematics tutor*”, “*Other Mathematics activities*”, etc. In one national version, however, the translator omitted to repeat “*Mathematics*” in each item. At the data cleaning phase, the students in that country appeared to spend disproportionate time in those activities, probably because they summed up *all* hours spent in homework, tutoring, etc., for *all* courses, rather than for *Mathematics only*. Missing data had to be imputed for this question in that country.

Check the level of generality of the words used.

A general notion such as “*remedial activities*” may correspond in your school system to a range of possible activities – e.g. small group “*catching up*” courses where the teacher re-explains difficult notions, or specialised help for dyslexic children, or extra language courses for immigrant students, or

private individual lessons, etc. – each of which may have specific names in your language. Using a too specific term (such as “*catching-up lessons*”) rather than a generic expression covering all types of remedial activities may seriously alter the information captured by the item.

Similarly, in some languages, the notion of “*homework*” corresponds to separate terms referring on one hand to “*doing written homework*” and on the other hand to “*studying lessons*”. Please make sure that the wording used does not reduce the information conveyed by the item to only one of these activities.

Pay special attention to the wording of answer categories.

Imperfect translation of answer categories (like “*Strongly Disagree, Disagree, Agree, Strongly Agree*” or “*Not at all, Very little, To some extent, A lot*”) can have considerable impact on the responses. Please be particularly careful in finding words or expressions that are as equivalent as possible to those in the source version.

- For example, the answer category « *More than once a month* » should NOT be translated into « *Several times a month* », since the latter suggest a slight higher frequency than “more than once”.
- In items where the answer categories are numeric intervals, check that the interval limits are equivalent in your language, so that the categories do not overlap. For example, answer categories such as “*1 to 2 hours/2 to 4 hours/4 to 5 hours/5 hours or more*” are unambiguous in some languages where the upper limit of an interval is always implicitly *excluded*. In other languages, the upper limit is implicitly *included*, thus the respondents whose answer would be “*2 hours*” or “*4 hours*” would be confused: answer “*2 hours*” could be included in both category “*1 to 2* and category “*2 to 4*”, and answer “*4 hours*” could be included in both category “*2 to 4*” and category “*4 to 5*”.

Never change the order of answer categories.

Like in test items, never invert the order the order of “*no/yes*” questions into “*yes/no*”, nor the order of “*boys/girls*” questions into “*girls/boys*”, nor the order of answer categories for any other type of items, such as “*Not at all/Very little/To some extent/ A lot*”, or “*Strongly disagree/Disagree/Agree/Strongly agree*”. At the data entry phase, these categories will be coded as 1/2 or 1/2/3/4. Inverting the order would result in variables with a wrong orientation, meaning exactly the reverse than intended.

In some languages, there may be a problem in using “*No/Yes*” answer categories for some questions, since just answering “*yes*” or “*no*” may be considered as “*rude*”, or “*impolite*” in the national culture. In case this problem exists in your language, please submit possible adaptations to the Consortium.

Do not introduce mismatches between answer categories and the wording of items.

Please make sure that, when translating multiple choice questionnaire items, the various proposed answers are **syntactically ‘aligned’** with the stem of the question.

- For example, the following question would sound awkward:

In your school, how do you usually group students?

- We do not use grouping.*
- Age.*
- By ability.*
- Separating undisciplined students.*

When translating questions that ask the respondent **how often** something happens (at home, or at school, or in the classroom), avoid including in the translation of the various items any indication of **frequency**.

- For example, in a question where the answer categories are “*Never, In some lessons, In most lessons*”, do not add in the items any time adverbs or adjectives, such as in “*Students often spend time in laboratories*” or “*There is constant noise and disorder in my class*” or “*My teachers would regularly help the students who need it*”).

Similarly, in “**How much...?**” questions, avoid adding in the wording of the items any reference to **quantity**.

- For example, in a question where the answer categories are “*Not at all/Very little/To some extent/A lot*”, avoid adding quantitative expressions in the items, such as “*Does your school have huge rates of students’ absenteeism?*”.

In the same order of ideas, any overt grammatical **negation** should be avoided in the wording of items that have **answer categories containing negative expressions**, such as “*Yes/No*”, or “*Not at all ...A lot*”, or “*Strongly agree...Strongly disagree*”.

- For example, a Likert-type item such as “*Learning advanced science topics would be difficult for me (Strongly agree... Strongly disagree)*” would become rather confounding for the students if it is translated into “*Learning advanced science topics would not be easy for me*” in your national version.

In translating attitude items, pay attention to words that convey affective nuances

Questionnaire items used to collect information on *opinions* or on *affective dimensions* such as self-concept or professional satisfaction scales are particularly sensitive to this specific kind of vocabulary issue.

- For example, in the Mathematics Self-Efficacy scale used in PISA 2003, respondents were asked to what extent they felt “*confident*” in solving a number of Mathematics problems. Particular care had to be devoted to the word used to translate “*confident*” in a number of languages, where potential corresponding words were likely to convey either a higher or a lower degree of certainty/uncertainty than the English word, which could have some impact on the distribution of responses.
- Items such as “*Teachers are supportive and encouraging towards the students*” heavily rely on the adequateness of the translation of the words “*supportive*” and “*encouraging*”.
- Finding a proper translation for the word “*reasonable*” would be crucial for items such as “*A priority in this school is to help the weakest students to attain reasonable levels of achievement*”.

Avoid increasing the social desirability of self-reported responses

When answering questionnaire items that ask for self-reports, many respondents tend to conform to what they think are socially ‘acceptable’ or ‘desirable’ positions (i.e., they tend to report more ‘positive’ characteristics than what the real characteristics actually are).

- This pattern is partly dependent on the ‘positive’ or ‘negative’ connotations of the vocabulary used in the items. Note that items containing particularly ‘positive’ or ‘negative’ words (e.g. “*In our school, teachers work with enthusiasm*”; “*School buildings are inadequate*”) are more

sensitive than others to this kind of artefact. When translating, please try to choose words that are as equivalent as possible in terms of ‘positive’, ‘negative’ or ‘neutral’ connotations.

Correctly adapt all <bracketed> terms and expressions that relate to school organisation.

In order to help your translators, all terms and expressions that require adaptation in many countries (sometimes in every country) appear in the source version between <clamps>, with references to explanations and translation notes attached. Please make sure that all these terms are translated into expressions, which are actually used by the students and the teachers of your country rather than into technical terms used only by specialists.

- For example, you will have to replace <Test language> with the name of the language used for instruction in your country (*Spanish* in Spanish-speaking countries, *Arab* in the Arab-speaking countries, etc.)
- The international categories <ISCED 1>, <ISCED 2> etc. should be replaced with the designations by which the countries refer to the study levels covered by those categories. Please conform to the international classification conventions adopted by your country (and described in detail in the OECD brochure *Classifying Educational Programmes Manual for ISCED-97 Implementation in OECD Countries*, 1999 Edition, OECD).
- The term <grade> must be replaced with the current term used in your country to refer to the grade attended (e.g. in France, grade 1 = *CPI*, grade 6 = *sixième*, grade 9 = *troisième*. The equivalent *form 1*, *6th form* and *9th form* are called differently in the UK, in Ireland or in New Zealand. In Germany, although there are slight differences across Länder, grade 1, 6, 9 = *Jahrgangsstufe 1, 6, 9*).
- If <Science> is divided up into separate subject matters in your country, you may have to reword certain questions in the Student Questionnaire (e.g.: “*In the last full week you were in school, how many class periods did you spend in <Science> subjects?*”) by specifying the various subjects (“... *in Physics, Biology, Chemistry*”).
- Make sure that correct course names are used. For example, in English-speaking countries, students usually attend “*Earth and Space Sciences*” courses, while this course name is “*Sciences de la Terre et de l’Univers*” [*Earth and Universe Sciences*] in a number of French-speaking countries.

When implementing such adaptations, please bear in mind that the formulation selected for the national version(s) should both be easy to understand by the respondent and enable the response-coding scheme to remain consistent across countries.

- For example, there is no easy solution in some languages when translating English expressions that include various subject matters in a single term, such as “*Social studies*” (which cover *History, Geography, Economy* and the like), or “*Language studies*” (which cover both *Mother Language* and *Foreign Language* courses). Please use the *Questionnaire adaptation spreadsheets* to accurately describe any such potential translation problem and to discuss with the Consortium’s questionnaire developers the adaptations needed in your national version.

It may also occur that the NPM wishes to incorporate additional items into the School or Student Questionnaire for the purpose of national analyses. Should this be the case, please bear in mind that:

- All additional questions that the NPM may want to implement as national options must be submitted to the Consortium for prior approval; and
- The timing of the Student Questionnaire session will have to be modified accordingly, depending on the time the students will need to respond to those additional items.

It might be useful to pilot your student questionnaire in a few classes, asking the students to circle the words and items that they did not understand. Check whether any translation problem was the source of the difficulties encountered.

Document all Questionnaire adaptations.

The adaptations done in the questionnaires have important implications in terms of data cleaning and data management. For this reason, the ***Questionnaire Adaptation Spreadsheet (QAS)*** describing all adaptations that your country intends to introduce in the Questionnaires must be submitted first, for approval, to the Consortium's persons who are in charge of questionnaire issues. Only when the changes and adaptations are approved by the Consortium, they can be entered in the materials.

5. TRANSLATION AND ADAPTATION OF MANUALS

Notes to NPMS

A number of national adaptations must be made to the *School Co-ordinator* and *Test Administrator Manuals*, to make sure that the instructions given to the School Co-ordinators and Test Administrators are consistent with decisions made by the NPM (for instance concerning the testing schedule). In the source version of both, text boxes titled "***Notes to NPMS***" draw your attention to those passages, in which it is recommended that the NPM make an adaptation.

Because NPMs need to adapt the Manuals to reflect the situation in their own countries, but in a way that is internationally consistent, it is very important that the Consortium reviews all adaptations and approves them before the Manuals are printed and circulated. The list below identifies a number of procedures, and indicates whether they can be modified or not. NPMs are encouraged to contact the Consortium if they have any questions about other proposed modifications.

Procedures that should NOT be changed in the Manuals

- Coding information required on the tracking instruments;
- The timing of the sections of the Assessment Booklet;
- The text of the script (after materials have been distributed) and the Booklet directions;
- The security of the items, and the importance of maintaining that security;
- The prohibition against the Test Administrator being a reading, mathematics or science teacher of students in the assessment;
- The requirement that a trained person administer the session.

Procedures that may be changed or adapted

- Separation of responsibilities between the School Co-ordinator and the Test Administrator – these positions may be combined or responsibilities interchanged.
- Definitions of special education needs and instructions regarding students to be excluded from the assessment – it is very important that exclusions be kept to a minimum. The wording may be changed, but the concepts should not be. ACER will review carefully how countries modify the exclusion categories;
- Procedures for notifying teachers, students, and parents about the assessment;
- How the Assessment Booklets and Student Questionnaires are packaged and shipped from the National Centre;

- Length of break between parts of the Assessment Booklet (up to a maximum of 5 minutes) and between the Assessment Booklet session and the Student Questionnaire session;
- The requirement that a follow-up session be held if more than 5 students (or more than 15% of the cluster size being used) are absent. NPMs may "suggest" that this be done "if at all possible" or they may delete the requirement. The goal is to increase student participation. Follow-up sessions are one way to increase student participation, but if NPMs think they will hurt participation, they should modify these procedures.

Documenting adaptations for Manuals.

It is mandatory that **all** of the adaptations done to the Manuals be documented in the appropriate ***Manual Adaptations Spreadsheet*** and submitted for approval to the Consortium.

References

Allalouf, A., Hambleton, R., & Sireci, S. (1999). Identifying the causes of translation DIF on verbal items. *Journal of Educational Measurement*, 36, 185-198.

Grisay, A. (2002). Translation and cultural appropriateness of the test and survey material. In Ray Adams and Margaret Wu (Eds.), *PISA 2000 Technical Report* (pp. 57-70). Paris: Organization for Economic Co-operation and Development.

O'Connor, K.M., and Malak, B. (2000). Translation and cultural adaptation of the TIMSS instruments. In M.O. Martin, K.D. Gregory, & S.E. Stemler (Eds.). *TIMSS 1999 technical report* (pp. 89-100). Chestnut Hill, MA: Boston College, 89-100.

Kelly, D. L. and Malak, B. (2001). Translating the PIRLS Reading Assessment and Questionnaires, in Progress in International Reading Literacy Study, (PIRLS), *Survey Operations Manual – Main Survey* (Doc. Ref.: PIRLS 01-0001). Chestnut Hill, MA: PIRLS International Study Center, Boston College.

OECD (1999). *Classifying Educational Programmes. Manual for ISCED-97 Implementation in OECD countries*. 1999 Edition. Paris: OECD.

Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229-224.

Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45, 153-171.

Hambleton, R.K. 2002: Adapting achievement tests into multiple languages for international assessments. In Porter, A.C. and Gamoran, A., editors, *Methodological advances in cross national surveys of educational achievement*. Washington DC: National Academy Press.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, 1, 1-30.

Jeanrie, C., & Bertrand, R. (1999). Translating tests with the International Test Commission's guidelines: Keeping validity in mind. *European Journal of Psychological Assessment*, 15, 277-283.

APPENDIX A .

CHECKLIST FOR THE VERIFICATION OF PISA TEST MATERIAL

*Verification of test units***1. Instructions to the students** (intro in italics, at the beginning of a unit)

Check whether these intros are present, whether they are located at the same place as in the source version and contain the same information.

2. Stimulus (= Text or document)

2.1 Text heading: If the source text has a title or heading, verify whether it has been translated. If it does not have any, please make sure no heading or title was added.

2.2 Text body

2.2.1 Is the *presentation (layout)* the same as that of the source version? (E.g. column layout, font size, text on one or several pages, location of graphics or illustrations etc.). If the lines are numbered in the source text, check whether they are also numbered in the national version you are verifying.

2.2.2 Is the text significantly *longer or shorter*?

2.2.3 How difficult is the *vocabulary* used (About the same as in the source text? Or is it harder i.e. are there more rare or abstract words? Or is it easier i.e. are there many cases where more common or concrete words were used?)?

2.2.4 How difficult is the *syntax* used (About the same as in the source text? Or is it harder i.e. are sentences longer, with too many phrases/subordinate clauses, is passive voice used more often, or complex negatives? Or is it easier i.e. are there many cases where the translation eliminates one or several difficulty factors?)?

2.2.5 Respect of *style and tone* of the source text (casual, elevated...).

2.2.6 *Content-related* errors (mistranslation, incomprehensible passages, forgotten passages, incorrect vocabulary).

2.2.7 *Form-related* errors (spelling, syntax, punctuation).

2.3 Graphics (if any):

2.3.1 Check whether graphics comply with the source version's (size, location, contrast of the shades of grey etc. The graph or illustration must be imported from the source file, not redrawn).

2.3.2 Check whether all keys and captions that belong with the graph are translated, whether they are reproduced at the right place and make sure that none of those keys and captions are missing.

2.3.3 Should certain questions refer to information contained in the graphs, verify whether the translation of acronyms, symbols, keys etc. do not make it harder or easier to find the correct answer.

2.4 Translation notes (if any):

2.4.1 Double-check whether the translator(s) actually complied with translation note(s).

3. Items (= Questions)

3.1 In a general way, it is very important to make sure that the wording would not make it harder or easier to understand the stem (due to its syntax, its vocabulary etc.) than the item's *stem* in the source version.

3.2 Check whether the question is presented the same way and that no facilitating elements were added or withdrawn (e.g. the translation should not add a prompt for the answer to an open-ended question if the source version does not propose any).

3.3 In questions that include several sub-items — presented as a table or list — the same format must be maintained in the translated version. In particular, please make sure that column headings do not contain more or less (specific) information than in the source version.

3.4 (Remark) In questions that include several statements whereby the student is requested to circle "Yes" / "No" or "True" / "False" or "Present" / "Absent" etc., in the source version this choice is printed on the left

hand-side of the different statements. Should this order be inverted in the translation (statements on the left, answer choice on the right), then the question would still work the same way; as a consequence, this type of adaptation is acceptable.

- 3.5 Verify whether the stem (part of the item that poses the question) contains the same information as the source version does. For instance if the source version states that there could either be *one or several* correct answers [this is the case for Questions in English that begin with “*Which of the following...*”), check whether the translation brings this ambiguity over (it should neither suggest that there is only *one* correct answer, nor should it suggest that there are *several*; it must indicate that there may be *one or several* correct answers).
- 3.6 If the stem refers to line numbers from the text/stimulus, double-check whether the reference still relates to the corresponding lines in the translation, where line numbers are likely to be different.
- 3.7 Both in the stem and in the remaining parts of the question, verify whether (possible) links with the relevant passage from the text are the same in the translation and in the source version: if the source question uses words or expressions identical to those used in the text/stimulus, this must also apply for the translation (literal match). However, if the source question uses synonyms or paraphrases, then the translated version should *not* use literal match. The nature of the connection must remain the same.
- 3.8 In multiple choice items, make sure that both the number and the order of the answer choices are the same as in the source version. As far as possible, the proportional length of the different answers should not differ too much from the source version. In particular, please make sure that the correct answer is not by far the longest if this is not the case in the source version. If certain distractors (incorrect answers) contain expressions identical to those used in the text/stimulus, the translation should reflect the same correspondence.
- 3.9 In open-ended questions, verify whether the number of lines (half-lines) provided for the student to fill in his/her answer is the same as in the source version. If these lines are numbered, they must also be numbered in the translated version. If they are not, make sure no numbers were added in the translation.

4 Scoring rubric

- 4.1 Check whether the scoring rubric has indeed been translated and make sure each question has its own rubric.
- 4.2 As far as possible, please check whether translations of the recurrent expressions in the scoring rubric were translated in a consistent way across all of the material.
- 4.3 In multiple choice items, make sure that the answer listed as correct answer is indeed the right one, and verify whether its wording is identical to that of the item.
- 4.4 In open-ended questions, double-check the relevant values for the different scores and see that they do not contain mistakes.
- 4.5 Verify whether the translation does not extend the scope of possible correct answers (due to a more vague description in the translation) or, conversely, make sure that this scope is not made too restrictive. For instance, when the scoring rubric indicates that an answer can be regarded as correct even if it does not *explicitly* contain a given element, provided that this element can clearly be *inferred* based on the content of the answer, check whether this indication is also stated in the translated version.
- 4.6 If the scoring rubric refers to line numbers from the text/stimulus, check whether those line numbers are correct in the translation.
- 4.7 If the scoring rubric quotes the text/stimulus, check whether the quotations are consistent with the text.

5 National Adaptations

- 5.1 First, check whether you received the *Test Adaptations Spreadsheet (TAS)* file with information for each and every unit translated. If the file appears to be missing, you must advise your verification co-ordinator, who will request the file from the country’s NPM (National Project Manager).
- 5.2 Verify whether the adaptations listed in the file comply with the guidelines stated in the *Translation Guidelines* and/or translation notes occurring within the unit itself. Please report adaptations that appear to go beyond authorised cases. When in doubt, consult the co-ordinator. Tricky cases will have to be submitted to Béatrice Halleux.
- 5.3 Also be watchful for possible adaptations to a unit that might not be listed on the *TAS* file and report such events.
- 5.4 In the same way, report cases where adaptations requested or advised have not been made in the translation.

Verification of questionnaires

- 6.1 Check the layout of the questions. If a question is spread over two pages, check that the headings are repeated at the top of the second page.
- 6.2 Check the question stems and items for possible awkwardness.
- 6.3 Verify whether any piece of information was omitted. In particular, check that none of the italicised instructions is missing.
- 6.4 Pay special attention to the translation of answer categories.
- 6.5 Check the match between answer categories and item wording.
- 6.6 In attitude items, check that affective expressions are consistently used across the item set (e.g., verify that the same words are used in all items that contain “*I like...*”, and that a different word is used in those that contain “*I enjoy...*”).
- 6.7 Check that all <bracketed> expressions were appropriately and consistently adapted. Cross-check on whether they correspond to the adaptations described in the Questionnaire Adaptation Spreadsheet.

Verification of the layout/presentation of the booklets (final optical check)

In a general way, the purpose is to check all elements not seen previously or requiring adaptation, and to use the opportunity to verify last-minute edits.

- 7.1 Double-check the contents of the flyleaf (title/cover page). More specifically, it must contain the same titles, subtitles and student identification box as the source versions.
- 7.2 Check the order and the content of the sections of the booklet. Make sure that no part is missing and, conversely, that the booklet does not contain optional tasks which are not foreseen by the source version. Verify that the section break between Part 1 and Part 2 has been correctly inserted.
- 7.3 Verify whether the general instructions are listed at the beginning of the booklets, and look for the presence of other instructions given to the student during the course of the assessment task or at the end of the booklet. The translation of those guidelines/instructions must be correct and easy to understand. In particular, ensure correctness of instructions such as “on the following/previous pages”.
- 7.4 Verify whether the final layout of each task is consistent with that of the source version. In particular, when stimulus/text and items/questions are presented opposite to one another in the source version, the same must apply in the translated version (i.e. the student should not have to turn the page to cycle back through the text while s/he is responding to the questions, unless this is also the case in the source version).
- 7.5 Verify final rendering of graphic elements (correct shading, light/dark contrast, etc.)
- 7.6 Verify all item codes, page numbering and footers.
- 7.7 Perform random checks to verify whether indispensable corrections are duly implemented.

Other Remarks

1. Possible inconsistencies between the English and French source

Report the inconsistency to the Verification Co-ordinator.
The English version must be considered as the reference version.

2. Verification Report

We expect to receive, on completion of your work,

- a) The original electronic copy in the target language, on which you will have noted points that require improvement and correction suggested. The co-ordinator will forward those documents to the relevant NPM and archive one copy.
- b) The Test Adaptation Spreadsheets, with your remarks on the adaptations suggested by the country’s translation team, and documentation on your suggested corrections.
- c) A short verification report. Please use the template that has been circulated at the Verifier’s training session to summarise the main strong points and weaknesses of the national version you will have verified.

NB: Please note that the whole of the PISA Test Material must be regarded as strictly confidential.