



5

Translation and Verification of the Survey Material

Development of source versions	90
Double translation from two source languages	91
PISA translation and adaptation guidelines	92
Translation training session.....	92
Testing languages and translation/adaptation procedures	92
Testing languages	94
International verification of the national versions	94
Summary of items deleted at the national level, due to translation, printing or layout errors	104



One of the important aspects of quality assurance in PISA is to ensure that the instruments used in all participating countries to assess students' performance provide reliable and comparable information. In order to achieve this, PISA implemented strict procedures for adaptation, translation and verification of national versions of all survey instrumentation.

These procedures included:

- development of two source versions of the instruments, in English and French (except for the financial literacy and reading component skills options and for the operational manuals, provided only in English);
- double translation design;
- preparation of detailed instructions for the translation of the instruments for the Field Trial and for their review for the Main Survey;
- preparation of translation/adaptation guidelines;
- training of national staff in charge of the translation/adaptation of the instruments; and
- verification of the national versions by independent verifiers appointed by the Consortium.

DEVELOPMENT OF SOURCE VERSIONS

Part of the new test materials used in PISA 2012 was prepared by the Consortium test development teams on the basis of submissions received from the participating countries. Items were submitted by approximately 21 different countries, either in their national language or in English, for either or both of mathematics or problem solving. The other part of the material was prepared by the test development teams at: ACER, the University of Melbourne (both in Australia); aSPe (University of Liege, Belgium); DIPF (Deutsches Institut für Internationale Pädagogische Forschung), IPN (Leibniz-Institute for Science and Mathematics Education) and Heidelberg University (all three in Germany); NIER (the National Institute for Educational Policy Research, Japan); CRP-HT (Centre de Recherche Public – Henri Tudor, Luxembourg); ILS (the Department of Teacher Education and School Research, University of Oslo, Norway); and ETS (Educational Testing Service, United States). Then, all materials were circulated (in English) to the expert groups and the National Project Managers (NPMs) for comments and feedback.

The item development teams received specific information/training about how to anticipate potential translation and cultural issues. The document prepared for that purpose was mainly based on experience gained during previous PISA survey administrations. The item developers used it as a reference when developing and reviewing the items.

The French version was developed at an early stage through double translation and reconciliation of the English materials into French, 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.

Experience has shown that some issues do not become apparent until there is an attempt to translate the instruments. As in previous PISA survey administrations, the English to French translation process proved to be very effective in detecting residual errors overlooked by the test developers, and in anticipating potential problems for translation in other languages. In particular, a number of ambiguities or pitfall expressions could be spotted and avoided from the beginning by slightly modifying both the English and French source versions; the list of aspects requiring national adaptations could be refined; and further translation notes could be added as needed. In this respect, the development of the French source version served as a pilot translation, and contributed to providing NPMs with source material that was somewhat easier to translate and contained fewer potential translation problems than would have been the case had only one source been developed.

The final French source version was reviewed by a French domain expert, for appropriateness of the terminology, and by a native professional French proof-reader for linguistic correctness. In addition, an independent verification of the equivalence between the final English and French versions was performed by a senior staff member of cApStAn who is bilingual (English/French) and has expertise in the international verification of the PISA materials, and used the same procedures and verification checklists as for the verification of all other national versions.

Finally, analyses of all national versions adapted from the French source version were conducted, using the Field Trial item statistics from the five French-speaking countries and from all English testing countries participating in PISA 2012. The primary aim of these analyses was to increase the equivalence between the two source versions. National item



parameters in both source versions were aggregated and then compared in order to detect statistical differences. Further, within each source version, item parameters were compared to identify potential outliers that might be responsible for a significant difference at the aggregated level. Two main item parameters were analysed: (i) the item difficulty and (ii) the item discrimination.

The results were used during the revision of the French and English source versions for the Main Survey in three ways:

- For each statistically different item functioning in most French versions, careful word-by-word comparison with the English version was done. If a wording issue was identified, the French source version, and consequently all French national versions, were modified. This also sometimes resulted in a change in the English source in order to increase the equivalence between the sources.
- When statistically different item functioning occurred in only one or two French national versions, the versions that worked correctly were examined to see if national adaptations or wording improvements had been made by the National Centres. If this was the case, these were implemented in the French source version and consequently in all other national versions for the Main Survey.
- Items that worked well and similarly in English and French were kept unchanged.

DOUBLE TRANSLATION FROM TWO SOURCE LANGUAGES

Back translation has long been the most frequently used way 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.

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.

PISA uses double translation from two different 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). In particular, one would wish the highest possible semantic equivalence (since the principle is to measure access that students from different countries would have to a same meaning, through written material presented in different languages). However, 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).

Some interesting findings in this respect were reported in the IEA¹/reading comprehension survey (Thorndike, 1973), which showed a better item coherence (factorial structure of the tests, distribution of the discrimination coefficients) between English-speaking countries than across other participating countries.

Resorting to two different languages may, to a certain extent, reduce problems linked to the impact of cultural characteristics of a single source language. Admittedly, both languages used in PISA share an Indo-European origin, which may be regrettable in this particular case. However, they do represent relatively different sets of cultural traditions, and are both spoken in several countries with different geographic locations, traditions, social structures and cultures.

The use of two source languages in PISA results in other anticipated advantages such as the following:

- Many translation problems are due to idiosyncrasies: words, idioms, or syntactic structures in one language appear untranslatable into a target language. In many cases, the opportunity to consult the other source version may provide hints at solutions.



- The desirable or acceptable degree of translation freedom is very difficult to determine. A translation that is too faithful to the original version may appear awkward; if it is too free or too literary it is very likely to jeopardise equivalence. Having two source versions in different languages (for which the translation fidelity/freedom has been carefully calibrated and approved by Consortium experts including through analysis of Field Trial item statistics) provides national reconcilers with accurate benchmarks in this respect, which neither back translation nor double translation from a single language could provide.

The double translation and reconciliation procedure using both source languages was recommended in PISA 2012 as in previous survey administrations. Countries testing in non-Indo-European languages received additional support: the document “The challenge of translating PISA materials into non Indo-European languages”, developed for PISA 2009, was re-circulated to countries for PISA 2012 as well.

PISA TRANSLATION AND ADAPTATION GUIDELINES

The *PISA Translation and Adaptation Guidelines*² had been extensively revised in 2009 with a view to obtaining a document that would be relevant to any PISA survey administration with relatively few administration-specific adjustments, which were made for 2012. The guidelines include:

- Instructions on double or single translation. Double translation (and reconciliation) is required for test and questionnaire materials, but not for manuals, coding guides and other logistic material. In double translation, it is recommended that one independent translator uses the English source version while the second uses the French version. In countries where the NPM has difficulty appointing competent translators from French/English, double translation from English/French only is considered acceptable; in such cases it is highly recommended to use the other source version for cross-checks during the reconciliation process insofar as possible.
- Instructions on recruitment and training.
- Security requirements.
- References to other documents, including technical guides for translating computer-based materials.

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.
- Instructions on how to adapt the test material to the national context.
- Instructions on how to translate and adapt questionnaires and manuals to the national context.
- The checklist used for the verification of PISA material.

As explained in the previous section, a separate document containing additional guidelines for translation into non Indo-European languages was also provided to countries.

After completion of the Field Trial, instructions on how to revise national version(s) were provided as a separate document known as *Material preparation Main Survey 2012*.³

TRANSLATION TRAINING SESSION

NPMs received sample materials to use when recruiting national translators and training them at the national level. The NPM meeting held in October 2010 in Budapest included a session on the Field Trial translation/adaptation activities in which recommended translation procedures, *PISA Translation and Adaptation Guidelines*, and the verification process were presented in detail.

At this meeting, countries were offered the opportunity to participate in a half day translation training workshop. Translators and NPMs attending the workshop received detailed information about the new PISA translation training “kit” designed to help National Centres implement PISA translation requirements in a more systematic way and were taken through presentations and hands-on exercises.

TESTING LANGUAGES AND TRANSLATION/ADAPTATION PROCEDURES

NPMs had to identify the testing languages according to instructions given in the *School Sampling Preparation Manual*⁴ and to record them in a sampling form for agreement.



Prior to the Field Trial, NPMs had to fill in a translation plan (describing the procedures used to develop their national versions and the different processes used for translator/reconciler recruitment and training). Information about a possible national expert committee was also sought. This translation plan was reviewed by the Consortium for agreement.

Figure 5.1 summarises the Field Trial translation procedures for cognitive materials, as described in the confirmed country translation plans.

The distribution of the translation procedures used for the questionnaires was quite similar (with some shift) within and between countries. For example, the questionnaires in German were produced using the double translation from English and French sources while the cognitive units were produced using the double translation from English with cross-checks against the French version.

There is a domain effect in the procedure changes compared to PISA 2009. Some countries (e.g. Germany and Norway) that used double translation from both English and French sources in 2009 chose double translation from English source with cross-checks against the French source version in 2012 because they could not find translators from French with good experience in the mathematics domain.

■ Figure 5.1 ■

Field Trial 2012 translation procedures

Activity	Paper-based instruments	Computer-based instruments*
Double translation from English and French source versions	18	14
Double translation from English source version with cross-checks against the French source version	10	7
Double translation from French source version with cross-checks against the English source version	1	0
Double translation from English source version only	20	7
Adaptations in one of the source versions	18	15
Adaptations made in a borrowed verified version	13	16
Adaptations made in a verified common version	4	4

*Note that some countries did not participate in any of the computer-based assessments. Note also that in some cases, countries such as Germany and Belgium (Flemish Community) borrowed part of the computer-based instruments (problem solving) but produced their own computer-based mathematics instruments because the country from which they borrowed the problem-solving version did not participate in the computer-based assessment of reading literacy option.

Countries sharing a testing language were 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. There is evidence from all previous survey administrations (PISA 2000 through PISA 2009) that high quality translations and high levels of equivalence in the functioning of items was best achieved in the three groups of countries that shared a common language of instruction (English, French and German) and could develop their national versions by introducing a limited number of national adaptations in the common version. Additionally, a common version for different countries sharing the same testing language implies that all students instructed in a given language receive booklets that are as similar as possible, which reduces cross-country differences due to translation effects.

Co-operation between countries sharing a same language was hence fostered and facilitated by the Consortium: workable models were designed so that verified versions from one country could be adapted by a second country.

- In the case of the Chinese-speaking participants Hong Kong, Macao, Shanghai and Chinese Taipei, a common base version was developed jointly, verified centrally, and then adapted for local usage by each of the four entities. Each national version was then verified separately.
- The model followed by German-speaking countries was (again) highly efficient: the German version of each of the components of the assessment material was double translated and reconciled by one of the countries, then verified, then adapted by the other countries who administered that component. The adapted versions were then verified.
- Spain and Mexico were the Spanish-speaking countries with the earliest testing window, and they developed their national versions following the standard PISA procedures. Their national versions were verified by cApStAn and then, once Mexico and Spain had accepted/rejected verifier interventions as needed, “clean” verified versions were made available to the other Spanish-speaking countries. Costa Rica used the verified Mexican version; Argentina, Chile and Uruguay used the verified Spanish version from Spain to prepare their national version. Colombia and Peru developed their own versions and used the verified version from Spain for cross-checks.



- Other co-operation models involving the use of borrowed versions included: Greece and Cyprus⁵ for Link items; Estonia and Latvia for the Russian version of the financial literacy option; the Russian Federation made its materials available to Estonia and Latvia; Italy and the Canton Ticino in Switzerland each translated from one source language and then held a meeting to reconcile and implement national adaptations; Montenegro made its version available to Serbia; and the United Arab Emirates borrowed the verified Jordanian version of paper-based Mathematics units.

TESTING LANGUAGES

A total of 98 national versions of the materials were used in the PISA 2012 Main Survey, in 46 languages. The languages were: Albanian, Arabic, Basque, Bulgarian, Chinese (traditional script), Chinese (simplified script), Catalan, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hebrew, Hungarian, Icelandic, Indonesian, Irish, Italian, Japanese, Kazakh, Korean, Latvian, Lithuanian, Malaysian, Norwegian (Bokmål and Nynorsk), Polish, Portuguese, Romanian, Russian, Serb (Ekavian and Yekavian variants), Slovak, Slovenian, Spanish, Swedish, Thai, Turkish, Valencian, Vietnamese and Welsh.

International verification (described in section below) occurred for 85 national versions out of the 98 used in the Main Survey.

International verification was not implemented when a testing language was used for minorities that make less than 10% of the target population or when countries borrowed a version that had been verified at the national level without making any adaptations. This concerned 18 versions across the following countries: Belgium (German), Finland (Swedish), Hong Kong-China (English), Ireland (Irish), Italy (Slovene and German), Liechtenstein (German), Lithuania (Polish and Russian), Macao-China (Portuguese), Montenegro (Albanian), Romania (Hungarian), Serbia (Hungarian), the Slovak Republic (Hungarian) – for which however financial literacy was verified, Slovenia (Italian), Spain (Valencian), Sweden (English) and the United Kingdom (Welsh).

Note that among these 18 versions, only three (Irish, Valencian and Welsh) were only verified at the national level. All other versions were prepared using internationally verified versions.

INTERNATIONAL VERIFICATION OF THE NATIONAL VERSIONS

As in previous PISA survey administrations, one of the most important quality control procedures implemented to ensure high quality standards in the translated assessment materials for PISA 2012 was to have an independent team of expert verifiers, appointed and trained by the Consortium; verify each national version against the English and/or French source versions.

The Consortium undertook international verification of all national versions in languages used in schools attended by more than 10% of the country's target population. For a few minority languages, national versions were only developed (and verified) in the Main Survey phase.

The main criteria used to recruit verifiers of the various national versions were that they had:

- native command of the target language;
- professional experience as translators from English or French or from both English and French into their target language;
- as far as possible, sufficient command of the second source language (either English or French) to be able to use it for cross-checks in the verification of the material (note that not all verifiers are proficient in French, but this is mitigated by the fact that the cApStAn reviewer and the translation referee have command of French);
- as far as possible, familiarity with the main domain assessed (in this case, maths);
- a good level of computer literacy; and
- as far as possible, experience as teachers and/or higher education degrees in psychology, sociology or education.

Verifier training sessions were held prior to the verification of both the Field Trial and the Main Survey materials. Attendees received copies of the PISA information brochure, *Translation Guidelines*, the English and French source versions of the material and a *Verification Checklist*. The training sessions focused on:

- presenting verifiers with PISA objectives and structure;



- familiarising them with the material to be verified, the verification procedures, and the software tools to be used (in particular, the Open Language Tool (OLT) software used for computer-based materials);
- reviewing and extensively discussing the *Translation Guidelines* and the *Verification Checklist*;
- conducting hands-on exercises on specially “doctored” target versions;
- arranging schedules and dispatch logistics; and
- security requirements.

Verification procedures have been continually improved throughout each PISA round, based on the experience and learning from previous rounds. In the following subsections we review the “state of the art” of procedures for the different components subject to verification. These included in 2012: the (first phase) verification of new test units (both paper-based and computer-based), of the “booklet shell” (paper-based), of practice units/widgets (computer-based), of questionnaires (including an optional online School Questionnaire); the “convergence check” of link units (both paper-based and computer-based); the final check of assembled test booklets, questionnaire booklets, computer-based test units, and coding guides; and the verification of (selected parts of) operational manuals.

Verification of paper-based test units

Since the PISA 2000 Main Survey, verifiers enter their suggested edits in Microsoft Word® files (mostly in item pool format, i.e. including coding sections) using the track changes mode. This facilitates the revision of verified materials by the National Centre, which can directly “accept” or “refuse” the edits proposed.

Since the PISA 2003 Main Survey, the mainstay of the verification procedure for test units has been the Test Adaptation Spreadsheet (TAS). Figure 5.2 shows a sample Test Adaptation Spreadsheet from the PISA 2012 Field Trial. The aim of this form is to function as an aid to translators, reconcilers, and verifiers (through the increasing use of item-specific translation/adaptation guidelines); as a centralised record of national adaptations, of verifier corrections and suggestions; as a way of initiating discussions between the National Centre and the Consortium referee; as a way of recording the implementation status of “key corrections” in test booklets; and as a tool permitting quantitative analysis of verification outcomes.

Some points of note are:

- Since PISA 2003, and increasingly so with each round, the column “Consortium Recommendation” is used to list item-specific translation/adaptation guidelines. These complement the general translation/adaptation guidelines and the translation notes embedded in source unit files with additional advice covering recommended, allowed or proscribed adaptations, literal or synonymous matches to be maintained, other psychometric characteristics to be considered (e.g. relative length or other patterns in multiple choice responses), desirable register of terms to be maintained, emphasis to be echoed, tips for the translation of difficult or idiomatic terms, etc. The verification co-ordinators consider that the generalised use of item-specific guidelines (for the attention of both translators and verifiers) is a significant breakthrough for translation quality assurance and quality control.
- Since PISA 2006, verifiers are instructed to document their “significant” verification interventions in the test adaptation spreadsheet, with a view to formalising the process by which a) the Consortium translation referees are informed of important issues and can liaise, when needed, with the test developers; b) if there is disagreement with the National Centre, a back-and-forth discussion ensues until the issue is resolved; c) key corrections in test materials are pinpointed so that their implementation can be checked at the Final Optical Check (FOC) phase. In the PISA 2000 and PISA 2003 rounds, this process was less structured.
- As of the PISA 2009 Main Survey, a conceptual change was introduced with regard to defining “significant” verification interventions tracked in the test adaptation spreadsheet. It was deemed desirable to reduce variability in the choice that verifiers make whether to report an intervention in the test adaptation spreadsheet or only correct in track changes in the unit, and to ensure that all potentially serious corrections are included in the test adaptation spreadsheet. This so that they may acquire “key correction” status and be checked during the Final Optical Check (FOC). The criterion was thus revised from “distinguish between purely linguistic issues and those that may potentially threaten equivalence” (used formerly) to “distinguish between minor errors or suggestions for improvement (that would not really affect the instruments if not corrected) and serious or potentially serious errors that require action.”

■ Figure 5.2 ■

Sample Field Trial Test Adaptation Spreadsheet (TAS) for a new PISA 2012 mathematics unit

UNIT, LOCATION	ENGLISH SOURCE VERSION	CONSORTIUM RECOMMENDATION (ITEM-SPECIFIC GUIDELINE)	COUNTRY COMMENT (DESCRIPTION + JUSTIFICATION OF ADAPTATION, ENG. TRANSL. OF NATIONAL VERSION)	VERIFIER INTERVENTION	VERIFIER COMMENT	CONSORTIUM REFEREE	CORRECTION STATUS	COUNTRY'S POST-VERIFICATION COMMENT	FINAL CHECK	COMMENTS RELATED TO FINAL CHECK
PM900hjump	metre (m), centimetres (cm)	Retain metric units throughout the unit	OK	OK						
Stimulus				Minor linguistic defect	Wrong use of capital letters in the title of the table. Changed by verifier	Please consider carefully the verifier correction	for NC to decide	OK, we will use the verifier translation		
PM900Q02	two similarities, changed	Use bold (or equivalent emphasis)	OK	OK						
PM900Q02 Scoring	question intent, description			Consistency		Question intent: Please consider carefully the verifier correction.	for NC to decide	OK, we will use the verifier translation		
PM900Q02 Scoring Code 2	The table lists a number of acceptable similarities.			OK	"similarities" was translated differently than in other occurrences of this word in PM900Q02 Scoring. Changed by verifier.	Please accept the verifier correction	KEY CORRECTION			
Code 2 dotpoints 1-4	AND	Use UPPERCASE (or equivalent emphasis)	OK	OK						
PM900Q03	per year, men's	Use bold (or equivalent emphasis)	OK	OK						
PM900Q03 Scoring				Errata	Incorrect scoring code in both item label and scoring section; Changed by verifier.	Please accept the verifier correction	KEY CORRECTION	OK	YES, corrected	YES, corrected NO, not corrected
Code 11	women	Use bold (or equivalent emphasis)	OK	OK						
Reserved for verifier: any other corrections in unit, entered in track changes but not listed above?				NO						

NO
YES, LESS THAN 4
YES, 5 OR MORE

OK
Added Info
Missing Info
Layout/Visual issues
Grammar/Syntax
Consistency
Register/Wording
Adaptation issue
Mistranslation
Untranslated text
Minor linguistic defect
Errata

- Since the PISA 2006 Main Survey, an innovation in the test adaptation spreadsheet is that verifiers use a scroll-down menu to categorise issues. As before, an additional comments field allows verifiers to explain their intervention with a back translation or description of the problem. The purpose of the categorisation is to reduce variability in the way verifiers document their verification; to make it easier for the Consortium referee to judge the nature of an issue and take action as needed; and to provide an instrument to help assess both the initial quality of national versions and the quality of verifiers' output.
- In the PISA 2012 Field Trial verification, there were 12 intervention categories for the verifier to choose from. Four new categories were introduced since PISA 2009: "OK", "Minor linguistic defect", "Untranslated text" and "Errata". The "OK" category implies a formal commitment from the verifier: s/he confirms having checked linguistic correctness and equivalence of the text element/segment and, if applicable, its conformity to an item-specific translation/adaptation



guideline. The category “Minor linguistic defect” was intended for minor issues that the verifier may wish to mention. “Untranslated text” was included as a separate category for cases in which part of the text remained in source language, for example in graph captions. “Errata” was introduced to differentiate between errors originating from the translation process and errors originating from defects in the source version.

- In training verifiers, special attention has been given since PISA 2009 to harmonising comment-writing practices. The life cycle of a comment makes it necessary to express it in such a way that it will be useful for both the National Centre and for the Consortium referee. Furthermore, the Final Optical Check (FOC) reviewer, who is not always the same person as the verifier, must be able to verify at final check whether a correction has been implemented or not. The following guidelines were set for verifier comments:
 - comments should be understandable to the Consortium referee who does not always understand the target language and preferably only looks at the test adaptation spreadsheet and the source version when reviewing comments;
 - specify in what way the target deviates from the source (rather than giving instructions on how to fix it, quoting the source, or explaining how the text has been corrected);
 - mention whether the verifier has made the correction or not for and why (e.g. because the verifier is unable to do it, or is not sure how to go about it);
 - comments should be factual and written in a clear and unemotional way and opinion statements should be kept to a minimum; and
 - each comment should relate to the category label selected.

For the PISA 2012 Field Trial verification, all translated versions were submitted for verification in item pool format (one unit per Word® file, including related coding sections), while a majority of national versions adapted from the English or French source versions were submitted in booklet format, i.e. as assembled test booklets (without coding sections). A special TAS was created to accommodate this, and later customised TAS were produced for the separate verification of coding guides. Customised TAS were also developed for other “homolingual” national versions adapted from a previously verified version in the same language.

Verifiers charged with homolingual versions were given modified instructions, with a focus on checking whether the adaptations made by the National Centre are appropriate and do not introduce errors or inconsistencies, checking for possible undocumented changes, and checking whether the National Centre has missed making any needed or recommendable adaptations to suit local context.

Verifiers for respectively English and French versions (and the cApStAn reviewer charged with co-ordinating and reviewing their work before delivery) were further instructed to consult a list of “tips”, including spelling and other adaptable elements but especially errata, errata candidates and “quasi-errata” (suggestions for improving the source) built up with each successive verification.

Main Survey verification of paper-based test units

Main Survey verification is, in essence, a second verification of materials already verified once before the Field Trial. In PISA 2009, Main Survey materials were fully (re-)verified, while in PISA 2012, the Main Survey units mostly underwent a *focused* verification: the verifiers concentrated on what had been or should have been changed since the Field Trial.

For paper-based assessment items, the assignment specification of verifiers included (i) checking correct implementation of Field Trial to Main Survey Consortium changes; (ii) assessing whether changes proposed by countries are acceptable by PISA standards; (iii) verifying whether those changes were correctly and consistently implemented; (iv) addressing issues from dodgy item reports; (v) checking whether the country used its final Field Trial version to prepare its Main Survey version; (vi) carrying out a selective full (re-)verification of earmarked units and, in some problem cases, of the entire national version. We refer to steps (i) to (iv) as “focused” verification, as opposed to “full” verification (sentence-by-sentence verification of national vs. source version on the entire materials), which is the norm for Field Trial verification.

For computer-based materials, as we shall see later, it was possible to perform an entirely “safe” focused verification i.e. there was no need for selective re-verification of parts that were effectively guaranteed to be final Field Trial versions in which no Field Trial to Main Survey changes were made or needed. Technological solutions implemented in the Translation Management System (TMS) allowed to a) ensure that countries would start out from correct “base” national versions; and b) mark segments that were (or should have been) revised by countries.

Paper-based materials were submitted for verification in cluster format, with the exception of a small group of early-testing countries which submitted units for verification in item pool format. The TAS were organised by cluster. The purpose of these spreadsheets was to list all Consortium changes; to provide countries with a file in which to list the changes they would want to introduce in their Main Survey version to address issues detected through the analysis of Field Trial results; to document verifier interventions and the follow-up of these interventions as indicated by the Consortium's translation referees.

■ Figure 5.3 ■

Field Trial to Main Survey Consortium changes documented in the TAS

UNIT, LOCATION	CLUSTER	Field Trial (FT) > Main Survey (MS) CHANGES			COUNTRY COMMENT (Description and/or justification of change, plus English translation of new national version)
		ENGLISH SOURCE VERSION (in case of national change)	NATIONAL FT VERSION (or English FT source version)	NATIONAL MS VERSION (or English MS source version)	
PM967Q03	PM7B-12	1 st response option in table	α must be an even number	α must be an even number of degrees	Consortium change (addition of 'degrees')
		2 nd response option in table	360° divided by α must be a whole number. Yes / No	360° divided by α must be a whole number. Yes / No	Consortium change (deletion of 2 nd response option)

The TAS were dispatched to the National Centres together with the final source versions, with rows pertaining to scoring sections greyed out (indicating that scoring sections would not be verified at the same time as items, see later).

Countries were asked to document in the TAS any Field Trial to Main Survey changes they wished to make in their units. Such changes were expected to be relatively rare, since the instructions to the National Centres were to “refrain from over-revisions just aimed at finding more “elegant”, “literary” or “efficient” words and syntactical structures. Such preferential or cosmetic modifications might affect the item difficulty in unexpected ways, perhaps introducing flaws in items that had no problem in the FT.” Verifiers were instructed to gauge national changes in light also of the above consideration. For each national Field Trial to Main Survey change, the National Centres were asked to enter the Field Trial version, the suggested revised Main Survey version, and to explain in English the rationale behind the change.

■ Figure 5.4 ■

Country change documented in the TAS

Field Trial (FT) > Main Survey (MS) CHANGES			COUNTRY COMMENT (Description and/or justification of change, plus English translation of new national version)
ENGLISH SOURCE VERSION (in case of national change)	NATIONAL FT VERSION (or English FT source version)	NATIONAL MS VERSION (or English MS source version)	
Statement 2: In December 2010, Locality 1 had more than half of its total population employed.	Τον Δεκέμβριο του 2010, περισσότεροι από το μισό πληθυσμό του δήμου 1 ήταν εργαζόμενοι.	Τον Δεκέμβριο του 2010, πάνω από το μισό του συνολικού πληθυσμού του Δήμου 1 ήταν εργαζόμενοι.	Slightly rephrased and addition of the word «total», missing in FT

Prior to the start of the verification, cApStAn received selected information from the Field Trial item analysis. This mainly concerned items for which countries suspected that a translation or adaptation issue might have affected response patterns. So, items were singled out whenever they showed differential item functioning in the Field Trial and either the country or the verifier felt that a national Field Trial to Main Survey change might address the problem. Information about such “dodgy” items was incorporated in the TAS of the affected versions prior to sending the materials to the verifier.

For those language versions that were deemed to be of unsatisfactory overall quality during the Field Trial verification, a full verification of all units was performed before the Main Survey. This was the case for three national versions.

Although there were random checks (six units were earmarked for full verification in each version), there was no systematic check whether National Centres had used the final Field Trial versions as a basis when implementing Field Trial to Main Survey changes, and no systematic check whether countries had made undocumented edits in addition to documented edits. However, verifier interventions in Main Survey units revealed this type of problem in some language versions: one country implemented Field Trial to Main Survey changes in unverified financial literacy units, and the same was observed in Mathematics clusters from two countries. Some countries testing in English submitted Main Survey source versions for verification (i.e. ignoring all national adaptations made during the Field Trial verification). Appropriate corrective action (e.g. re-implementation of key corrections or national adaptations from the Field Trial stage) was taken in such cases.



Verification of the booklet shell

Since PISA 2006, the booklet shell has been handled as a separate component subject to verification. In PISA 2012, the booklet shell included the *Booklet Cover*, the *Formula Page*, the *General Directions*, sources/copyright acknowledgements, and the *Calculator Use and Effort Survey*. For countries taking the financial literacy option, the latter was replaced by two different forms of a short “*Questions about your experiences on money matters*” survey in the financial literacy booklets. It was dispatched together with a booklet adaptation spreadsheet which has the same structure as the test adaptation spreadsheet, and is verified following the same procedure as the test units.

Verification of computer-based test units

Of the 66 countries in PISA 2012 Field Trial, 43 participated in the computer-based assessment (CBA) taking at least the problem solving core component. Of these 43 countries, 29 countries participated also in the optional computer-based mathematics assessment, and 14 countries participated also in the optional digital reading assessment. Nineteen countries had participated to the digital reading assessment (DRA) within the framework of the PISA 2009 Main Survey. All other countries were “new” in administering computer-based materials.

Computer-based units were translated and verified using the Open Language Tool (OLT) software on XLIFF (tagged XML Localisation Interchange File Format) files which were exchanged, previewed and archived on the ACER Translation Management System (TMS), a web-based upload-download platform.

To perform the verification task, verifiers used a version of the OLT software especially customised for PISA by DIPF. They were instructed to verify the text segments one by one, comparing target version appearing on the right side of the OLT interface to source version appearing on the left side, consulting previews and the comment window to see item-specific guidelines and National Centre comments. They made corrections as needed, documenting their interventions in the comment window within OLT, including selection of the appropriate intervention category using a drop-down menu. The basic verification task, the checklist and the intervention categories were the same as for the paper-based materials. However, there was no TAS (replaced by the comment field carried within each XLIFF file and later in the process by the differences report). Also, as there is no “track changes” facility in the OLT, verifiers used the segment status marker (a feature of OLT) to differentiate edited versus unchanged segments.

Once a domain was verified, reviewed and “finalised” on the TMS, the Consortium Referee obtained an automatically generated “differences report” in Excel® format from the TMS. This report was used as the TAS for the paper-based units for a key corrections process (but using the international contractor’s website MyPISA, <http://mypisa.acer.edu.au>, for various phases).⁶ Key corrections were negotiated between Consortium referee and National Centre, the National Centre uploaded revised XLIFF files on the TMS for final check, the final check reviewer checked the correct implementation of key corrections and either released the files to ACER for national version construction or released them back to the National Centre for last-minute corrections.

Arrangements for the verification of homolingual versions – computer-based units adapted from the English or French source versions, from a common version (e.g. Chinese) or from a borrowed verified same-language version (e.g. Chile borrowing from Spain) were similar to those for paper-based cognitive units. In particular, English and French versions benefited from a co-ordination process similar to the one implemented for cognitive materials: a list of “tips” for verification of computer-based units, including spelling and other possibly adaptable elements, and especially errata, errata candidates and “quasi-errata” (suggestions for improving the source) was maintained, built up, and used in each successive verification.

As already mentioned, in the case of computer-based units a breakthrough was achieved in PISA 2012 regarding a “safe” exclusively focused verification at Main Survey phase. Different from paper-based materials, the process for CBA ensures that countries started out from their verified and finalised Field Trial versions, and prevents undocumented Field Trial to Main Survey changes that may escape verification. Further, the segment status used for marking unchanged segments (“Approved”) also “lightly locks” such segments: both the National Centres and the verifiers could not change these segments by accident. This helps both in discouraging “cosmetic” changes by National Centres and verifiers (they have to unlock “Approved” segments where they purposefully choose to make a change), and in easing the post-verification review by National Centres (at that stage, they can ignore the “Approved” segments in verified XLIFF files).

Verification (convergence check) of coding sections/coding guides

For the Field Trial, coding sections were verified either together with stimuli/items – in Word® files submitted in item pool format – or separately, the latter being the norm for versions adapted from the French or English source versions and for



the computer-based (CB) coding guides (problem solving, mathematics, reading). The process was the same as for other paper-based (PB) test materials, but there was no final check of key corrections related to coding sections at Field Trial stage. It was the countries' responsibility to implement key corrections in the coding guides and to echo edits that arose from the coder training meeting in Rome in February 2011, which were reflected in the final "post-Rome" version of the coding guides released in early March 2011. This meant differential treatment between countries that submit their units for verification earlier or later: early-testing countries (other than those testing in English or French) had their coding sections verified before the final source version of the coding guides was released.

For the Main Survey, consolidated PB and CB Coding Guides were verified (or "convergence-checked", see next paragraph) separately from assessment items, after the source version of coding guides was finalised in March 2012, incorporating late edits arising from the coder training meeting in Salzburg. Introductions were fully verified. For the coding sections, the procedure was the same as for the clusters: countries document in the TAS any national FT (Field Trial)>MS (Main Survey) changes they want to make, and verifiers are asked to check that such changes are acceptable and correctly and consistently implemented in the guide, in addition to checking correct implementation of Consortium FT>MS changes. Because in the Field Trial the coding guides were verified prior to the coder meeting for most language versions and post-coder training coding guides had not been subject to verification, checking the correct implementation of a sample of late FT changes was included in the Main Survey verification procedure.

Verification (convergence check) of link units

Link units were verified for three new versions from Cyprus⁷ (Greek and English) and Viet Nam, which had not participated in PISA 2009. Cyprus⁸ adapted their Greek units from the version used by Greece in 2009 and their English units from the English source version, while Viet Nam provided new translations of all items. For these countries, link units were verified following the same procedure as the new mathematics or financial literacy units (full verification at Field Trial, focused verification at Main Survey).

For countries and economies that had already participated in PISA, the Main Survey process included a special focus on checking identicalness between national versions of link items (which were not field-trialled) versus the versions used previously. This was assessed by means of a semi-automated convergence check, whereby the versions the countries submitted in 2012 were compared to the final PISA 2009 Main Survey versions that countries had uploaded to MyPISA in 2009. It remains a concern that a vast majority of countries either failed to use the final 2009 versions to assemble the 2012 versions or introduced a number of changes without documenting them, often dismissing these modifications as unimportant – whereas the literature indicates that even minute changes can potentially affect response patterns.

A convergence check was organised to compare the content of the nine link clusters (three for science, three for mathematics and three for reading) submitted by countries for the PISA 2012 Main Survey with the content of the corresponding clusters administered in the PISA 2009 Main Survey. To avoid jeopardising the trend, no changes were allowed, except if it seemed indispensable to correct an outright error. For such exceptions, National Centres had to describe the rationale of each edit in a special "link TAS" prior to submission of the link clusters, which was made available only to those countries that requested it. Each of these edits was first commented by the verifier, then by the Consortium referee, and if needed by the test developers. Each change request was ultimately approved or rejected, following a negotiation phase with the National Centre if needed.

A selection of 5 science link units (out of 18), 7 mathematics link units (out of 25) and 5 reading literacy link units (out of 13) was earmarked for a comprehensive convergence check. The selection of units was not arbitrary: (i) based on TAS statistics per unit from the PISA 2009 Main Survey, units with the highest number of verifier interventions in 2009 were selected; (ii) units from PISA 2009 that underwent last minute changes were also included. The National Centres did not know which units were selected for convergence check.

The convergence check was performed using a combination of an automated PDF[®] comparison functionality and a manual check. The verifier documented all discrepancies in the link TAS. The verifier was (in a first stage) not asked to assess the significance of identified changes. If it became apparent that the country had made a great number of changes, the convergence check was interrupted and the National Centre was asked (i) to check whether the correct version of link units had been used to prepare the PISA 2012 Main Survey link clusters; (ii) if yes, to either make formal, well-documented requests for these changes; or (iii) to revert back to the 2009 version. In case the National Centre insisted on making changes, the full negotiation procedure was launched: the link TAS was submitted with a description of the rationale for each change; the verifier commented on them, then the Consortium referee, and then the test



developers. The TAS was then sent back to the country, and the country was asked to re-submit link clusters, in which *only* those changes that had been approved should be implemented. In this second convergence check, at least one (randomly selected) unit per domain was added to the convergence check.

For 11 out of 77 language versions, the convergence check did not reveal any changes from PISA 2009 to PISA 2012. For 10 language versions, changes were requested *before* the convergence check. The link clusters submitted for the convergence check were rejected for 11 language versions due to a high number of undocumented changes. In the case of Kazakhstan (both Russian and Kazakh versions) too many changes were found also in the second convergence check. As a result, these two versions were fully verified. In a majority of language versions, 1 to 10 discrepancies were identified in the units selected for convergence check. All discrepancies were regarded as “key corrections”, meaning that compliance with the Consortium’s recommendation for each discrepancy documented in the TAS is checked at final check stage.

The convergence check procedure of the link units revealed two major weaknesses: firstly the National Centres often failed to comply with the requirement that the link units used in 2012 should be fully identical to the units used in 2009. Compliance with the procedure was achieved only by a minority of National Centres. Secondly; it seems that many National Centres found it difficult to locate the correct, final versions of their 2009 units. This may be due to changes in national teams since 2009 or to defective version management policies.

The scoring sections of the new mathematics units were verified separately in the form of consolidated coding guides; therefore also the convergence check of the scoring sections of link units took place separately from the convergence check of clusters. The procedure was the same as for clusters: the same “link TAS” was used and the selection of units subject to this check was the same. Very few countries documented the changes they wished to make, and most countries made some changes. As the scoring sections are less sensitive in terms of impact on maintaining the link, minor changes (punctuation issues, typos corrected, and changes in the labelling of scoring rubrics made for consistency across domains or with new units) were mostly approved by the Consortium referee, even if the country had not specifically requested permission to make such changes. At final check of coding guides it was checked if residual issues in the coding sections of the link units were correctly addressed.

Verification of questionnaires

Questionnaires are submitted for verification together with an agreed Questionnaire Adaptation Spreadsheet (QAS). The purpose of the QAS is to document all content-related deviations from the international reference versions. Such national adaptations are subject to clearance by the questionnaire team before the material is submitted for verification.

The verifiers’ brief (successively refined throughout PISA survey administrations) is now defined as checking whether target questionnaires are linguistically correct and faithful to either the source version (when no adaptation is made) or the approved English translation of the national version (when an adaptation is made). With a view to this, verifiers are instructed:

- to check whether the back translation of the agreed adaptation is faithful;
- to check whether the agreed adaptation is correctly reflected in the questionnaire;
- to check the questionnaires for undocumented adaptations (deviations from the source not listed in the QAS) and report them; and
- to check linguistic correctness (grammar, spelling, etc.) of the entire target version.

In the same manner as for test units, corrections are entered in the questionnaires using the track changes mode (exception: the optional online school questionnaires, for which a modified procedure was used to track corrections), while verifier comments are entered in the verifier column of the QAS.

In essence the Field Trial verification procedure for questionnaires did not change significantly since PISA 2009. The main changes consisted in (i) using MyPISA for each milestone; (ii) introducing verifier intervention categories in questionnaire verification; (iii) having a structured key corrections check, similar to the final check process for booklets.

Although there were no special “homolingual” procedures for the verification of questionnaires (which are very extensively adapted even when sharing a common language), English and French versions benefited from a co-ordination process similar to the one implemented for cognitive materials. A list of “tips” for verification of questionnaires, including



spelling, possibly recurring adaptation issues, and especially errata, errata candidates and “quasi-errata” (suggestions for improving the source) was maintained, built up, and used in each successive verification.

There was also an increased effort, versus previous PISA rounds, to harmonise the verification feedback for different language versions of questionnaires used in the same country (e.g. German, French and Italian for Switzerland, or the five language versions for Spain). Such versions are by necessity entrusted to different verifiers, but insofar as possible, cApStAn’s verification reviewers made a point of reviewing and delivering such versions together, striving to harmonise e.g. verification interventions on adaptation issues common to the different language versions.

Main Survey verification of questionnaires

Similarly to the procedure adopted as of 2009, to significantly save time during the reviewing process, ACER provided a customised QAS for each language version in which all the Field Trial national adaptations were imported for items that were left unchanged from the Field Trial.

The Main Survey QAS was designed to track the particular Main Survey translation/adaptation and verification process. Next to the Field Trial column documenting the English translation of the national version in Field Trial (which was locked), a column was added in which the National Centre was required to either confirm the unaltered Field Trial version or record the intended revised adaptation for Main Survey. If a source element had been altered by the Consortium, or if it was “dodgy”, i.e. it worked in an unexpected way during the Field Trial at national or international level, the National Centre was required to provide the Main Survey adaptation. In the QAS, source items removed from the Field Trial version were deleted without tracking, but for the items that were modified; “strikethrough” and coloured text were used to indicate the changes. Rows subject to verification were identified via colour-shading.

Once the negotiation of Main Survey adaptations was concluded, cApStAn could launch the verification. For the Main Survey (focused verification), the verifiers were instructed to: check that all the Consortium changes are correctly implemented, check that all national changes are appropriate and implemented as documented, verify dodgy items completely, re-check whether confirmed adaptations are correctly implemented, verify completely questions that were modified or added; and fully (re-)verify a selection of items.

Similarly to the process for paper-based cognitive materials, verifiers entered corrections in track changes mode in the Word® files and documented the verification in the QAS. The QAS was used like the TAS for paper-based units for a key corrections process (but using MyPISA for various stages).

As was the case for cognitive materials, verification revealed occurrences of undocumented Field Trial to Main Survey changes or Main Survey changes introduced in non-finalised (e.g. pre-verification) Field Trial instruments. Such occurrences were treated on a case per case basis. Verifiers re-implemented the still relevant Field Trial key corrections in the Main Survey version, and then performed the verification of target versus source as per Main Survey procedure.

Final (Optical) Check of test booklets, questionnaire booklets, computer-based test units, coding guides

As in all previous rounds, test booklets and questionnaire forms were checked, at both Field Trial and Main Survey, page-by-page for correct item allocation, layout, page numbering, item numbering, graphic elements, item codes and footers. This phase continues to prove essential in locating residual flaws, some of which occur only during assembly of materials.

The final check process for test booklets and questionnaire booklets in the PISA 2012 Main Survey was mostly unchanged from the Field Trial and from PISA 2009, though a more formalised Final Optical Check (FOC) review process was put in place. A FOC reviewer carefully reviews the draft FOC report (systematic check of all comments listed by the FOC verifier) plus transfers any issues concerning key corrections from the various adaptation spreadsheets to the FOC report, using the category “Residual issue at content level” (since PISA 2009, this is the process used to enforce key corrections). The FOC review process was formalised as a distinctive step in the PISA 2012 Main Survey, along the lines of what was done for paper-based clusters, computer-based units, questionnaires, and coding guides: a reviewer stands behind each verifier.

In practice, the FOC review leads to a “lighter” FOC report. FOC verifiers are trained to spot and report every possible layout deviation, and they do not necessarily have knowledge on specific languages nor are they acquainted with



all verification steps, in particular with the convergence check of link items. Thus a major aspect of FOC review is the “weeding out” or “qualification” of comments. In particular, “qualification” is often required when it comes to comments concerning link items, for which the FOC may reveal deviations versus the source version – rather than versus the previous national version. When minor issues or deviations had previously been approved (or had previously passed unchallenged) in link items, the FOC reviewer suitably modifies the comment.

Computer-based units were also subject to a final check process focusing on the correct implementation of key corrections. After verification and post-verification review by the country (including possible negotiation of key corrections with the Consortium referee), a final check reviewer consults the differences report and launches previews of the latest target versions uploaded by the country and of the English (or French) source versions as needed. If all key corrections are found to be correctly implemented, the final check reviewer documents this in the differences report and finalises the unit on the TMS, indicating it is ready for national construction. If a “key correction” is found not implemented (or only partially implemented, or defectively implemented), the final check reviewer describes the issue in the differences report and “un-finalises” the unit on the TMS, making it accessible again to the National Centre for re-editing.

For CBA units (as for Questionnaires but differently from PBA clusters, coding guides, booklets), the final check was part of a MyPISA task.

For the Main Survey (only), after verification and post-verification processing by the National Centres (including possible negotiation of key corrections with the Consortium referee), it was required to submit coding guides in PDF[®] format for a final check.

The final check of coding guides consisted in: a page-by-page check for completeness; a systematic check of correct implementation of key corrections from the verification of the general introductions and the verification and/or convergence check of coding sections; a random check of key corrections from the verification of stimuli + items, to determine whether final versions of stimuli and items were correctly used to assemble the coding guides (this part of the process was sometimes carried out at verification phase); a random check of late Consortium Field Trial changes in the coding section of new mathematics units, to determine whether the final Field Trial mathematics coding guide was correctly used as basis for updating to Main Survey – not forgetting that for most countries, the correct implementation of such changes was not checked at Field Trial when cognitive materials were verified in unit format, before the release of the final post-coder-training-meeting Field Trial coding guides (this part of the process was also sometimes carried out at verification phase).

The final check of coding guides did not include a layout check as such. Different from test booklets, countries are not required to echo the source version pagination and may even change the order of presentation of the material (e.g. some countries prefer cluster order rather than the unit order used in the source). Nonetheless, the completeness check often reveals minor layout or formatting defects (e.g. incorrect formatting of credit labels, descriptors, or sample student responses), which are then reported. In the case of one national version with changed order of presentation (Belgium-Dutch), it revealed an entire missing unit in the paper-based mathematics coding guide.

A final check report in Excel[®] format was sent to each National Centre, detailing any recommended corrections; these were infrequent and mostly minor (see above). Differently from the final check of test booklets or questionnaires, the procedure for coding guides was kept lighter in that National Centres were not asked to return the report with confirmations of their follow-up (selection for each issue of “Corrected” or “Not corrected” plus explanation in the latter case). In process terms it may be questioned whether this differential treatment is justified.

Similar to analyses of the test adaptation spreadsheet, cApStAn conducted quantitative analyses of FOC reports both at Field Trial and Main Survey phases, which gave good estimates of the number and types of residual errors found in assessment booklets.

Verification of operational manuals

The verification process for Manuals in PISA 2012 is basically unchanged from PISA 2009: ACER vets the national adaptations, following which cApStAn verifies a list of key components called “specified parts”. One difference is that in 2012 it was decided that cApStAn would perform its verification only in the Main Survey – whereas in 2009 verification by cApStAn was performed in the Field Trial and then repeated in the Main Survey for “problematic countries”, i.e. those for which significant manuals-related issues were identified at FT.



In practice the bulk of manual verification work is carried out or anyway extensively prepared by cApStAn staff, with national verifiers consulted only as needed. Extensive explanations have been provided in previous verification reports on the reasons for involving verifiers the least possible in manual verification.

Quantitative analyses of verification outcomes

In PISA 2000 and PISA 2003, verification reports contained qualitative information about the national versions and illustrative examples of typical errors encountered by the verifiers. As of the PISA 2006 Main Survey, the instruments used to document the verification were designed to generate statistics, and some quantitative data is available. The verification statistics by item and by unit yielded information on translation and adaptation difficulties encountered for specific items in specific languages or groups of languages. This type of information, when gathered during the Field Trial, could be instrumental in revising items for the Main Survey but would also give valuable information on how to avoid such problems in further survey administrations. The verification report includes all data and country names and is a confidential document reviewed by the technical advisory group. Each country received its own report and data.

This information also makes it possible to detect whether there are items that elicited many verifier interventions in almost all language groups. When this occurs, item developers would be prompted to re-examine the item's reliability or relevance. Similarly, observing the number of adaptations that the countries proposed for some items may give the item developers additional insight into how difficult it is for some countries to make the item suitable for their students. While such adaptations may be discussed with the Consortium, it remains likely that extensively adapted items will eventually differ from the source version (e.g. in terms of reading difficulty).

The verification reports for the PISA 2012 Field Trial and PISA 2012 Main Survey include sections with quantitative analyses conducted on verification and assessment booklet FOC outcomes. They also contain pointers and directions for further work that could be carried out in this direction. NPMs have shown a keen interest in this type of analysis.

SUMMARY OF ITEMS DELETED AT THE NATIONAL LEVEL, DUE TO TRANSLATION, PRINTING OR LAYOUT ERRORS

In all cases when serious flaws were identified in the functioning of specific items as administered in individual countries, the NPMs were asked to review their translation of the item and to provide the Consortium with possible explanations of the malfunction.

Across all administrations of the mathematics, reading, science, problem solving and financial literacy items used for PISA 2012, approximately 47 instances were identified for which data were excluded from the analysis for reasons related to a translation issue (38 cases), printing issue (3 cases) or layout issue (6 cases).

Some 24 countries had one or more items affected by such errors, and an additional three instances affected several countries using a common language (two items affecting several of the German-speaking countries, and one item affecting some of the countries using Chinese).

Notes

1. International Association for the Evaluation of Educational Achievement.
2. Available at www.oecd.org/pisa.
3. Available at www.oecd.org/pisa.
4. Available at www.oecd.org/pisa.



5. Note by Turkey: The information in this document with reference to “Cyprus” relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the “Cyprus issue”.

Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

6. The MyPISA website contains the source versions of instruments, manuals and other documents and information relating to National Centres.

7. See note 5.

8. See note 5.

References

Grisay, A. (2003), “Translation procedures in OECD/PISA 2000 international assessment”, *Language Testing*, No. 20 (2), pp. 225-240.

Thorndike, R.L. (1973), *Reading Comprehension in Fifteen Countries*, Wiley, New York, and Almqvist & Wiksell, Stockholm.