

Chapter 5

Translation and Verification of the Survey Material

INTRODUCTION

This chapter explains the procedures used for translation, adaptation and verification for both paper-based (PBA) and computer-based (CBA) materials in PISA 2018.

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, strict procedures for the localisation (adaptation, translation and validation) of national versions of all survey instrumentation were implemented in PISA 2018 as in all previous PISA rounds.

These procedures included:

- optimising the English source version for translation through translatability assessment
- developing two source versions of the instruments, in English and French (except for the financial literacy and for the operational manuals, provided only in English)
- implementing a double-translation design with a final reconciliation
- preparing detailed instructions for the localisation of the instruments for the field trial and for their review for the main survey
- preparing translation/adaptation guidelines
- training of national staff in charge of the translation/adaptation of the instruments
- validating the translated/adapted national versions: verification by independent verifiers, review by cApStAn staff and the translation referee or the Questionnaires team, countries' post-verification review and "technical" and linguistic final checks.

DEVELOPMENT OF SOURCE VERSIONS

Translatability assessment

The translatability assessment was an effort to combine linguists' expertise with that of item developers to bridge the gap between a draft item written in the source language, and an actual source version of that item, suitable for translation/adaptation.

While item writers are increasingly aware of localisation issues, they are rarely in a position to identify some of the challenges translators will be confronted with. In line with the trend to do more upstream work, i.e. work before the start of the actual translation process, a methodology was developed to identify and document potential translation and adaptation difficulties in draft PISA 2018 items before the source versions were finalised. This process, referred to as the translatability assessment, was first implemented in PISA 2015.

The translatability assessment consists of submitting draft versions of new items to a pool of experienced linguists covering a broad range of language groups. These individuals were selected among the international verifiers and were trained to use a set of 13 translatability assessment categories to report on potential translation, adaptation and cultural issues that could affect translatability. For both questionnaire items and new reading items, the items were submitted in batches. The work was organised so that at least three linguists, from

different language groups, would comment on each item. Note that in PISA 2018 there were no new items for the science or mathematics assessments.

The approach was for each linguist to first mentally translate each item allocated to him/her. When the item appeared straightforward to translate, the item was classified as “straightforward.” When the linguist found an item somewhat difficult to translate/adapt or identified a potential cultural issue, s/he went through the exercise of (i) producing a written translation of that item; (ii) selecting the relevant translatability category (such as “Unnecessarily complex” or “Potential cultural issue”); (iii) describing the issue; and (iv) proposing an alternative wording or a translation/adaptation note to circumvent the problem. It should be noted that the translations produced in category (i) were not intended for further use; they were used to help the linguists identify and describe the translation and adaptation challenges that translators might face if no pre-emptive action were taken.

The feedback from the different linguists was then collated by a senior linguist at cApStAn or, in some cases, by the translation referee. The referee reformulated the comments so that similar issues were processed in a consistent way; selected or rewrote proposals for alternative wording that addressed all the issues identified and drafted translation/adaptation notes when applicable. When several linguists working in different languages pointed out similar issues in a given item, special attention was given to the wording of that particular item. The senior linguist produced a *Translatability Report*, which was then sent to the item developers for review. Using the Translatability Report, item developers took this opportunity to eliminate ambiguities, e.g. Anglo-Saxon idiosyncrasies that may be difficult to render in certain languages, double-barrelled questions, cultural issues or unnecessary complexity. Overall, an attempt was made to fine tune the initial version of the items so that it became a more translatable source version.

Production of the second source version in French

Since the inception of PISA, it has been a requirement in the PISA Terms of Reference that the international contractor should produce an international French source version of the data collection instruments. Experience has shown that some issues do not become apparent until there is an attempt to translate the instruments into a second language. As in previous PISA survey administrations, the English-to-French translation process proved to be very effective in detecting issues not detected or overlooked by the item writers, and in anticipating potential problems for translation into other languages. In particular, a number of ambiguities or pitfall expressions could be spotted and therefore avoided in the source versions by slightly modifying both the English and French source versions. As a result, the list of aspects requiring national adaptations could be refined; and further translation notes could be added as needed.

The French source version was produced through the double-translation and reconciliation process, followed by a review 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 using the same procedures and verification checklists as for the verification of all other national versions.

Both the translatability assessment and the development of the French source version contributed to providing national project managers (NPMs) with source material that was easier

to translate and contained fewer potential translation problems than would have been the case had only one source been developed without a translatability assessment.

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 second approach is a double-translation design (i.e. two independent translations from the source language(s), and reconciliation by a third person).

This second approach 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. On the other hand, in a back-translation design 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.

Both back-translation and double-translation designs have a potential disadvantage 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. 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 recommended approach in PISA therefore builds on the strengths of the double-translation approach by using double translation from two different source languages.

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. 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 second 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, with clear guidelines on the amount of translation fidelity/freedom,

provides national reconcilers with accurate benchmarks in this respect, which neither back translation nor double translation from a single language could provide.

As in previous PISA cycles, the double-translation and reconciliation procedure were a requirement for all national versions of test and questionnaire instruments used in the assessment. It was possible for countries to use the English source version for one of the translations into the national language and the French source version for the other. An efficient alternative method was to perform double translation and reconciliation from one of the source languages, and extensive cross checks against the second source language. For the option Financial Literacy units were double translated from English only, as there was no French source version of these units.

PISA TRANSLATION AND ADAPTATION GUIDELINES

PISA Translation and Adaptation Guidelines were produced to guide the national teams in the adaptation work of the instruments. The guidelines included:

- Instructions on double or single translation: Double translation (and reconciliation) was required for test and questionnaire materials, but not for manuals, coding guides and other logistic material. In double translation, it was recommended that one independent translator use the English source version while the second use the French version. In countries where the National Project Manager (MPM) has difficulty appointing competent translators from French and English, double translation from English or French only was considered acceptable; in such cases it was 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 and reconciling computer-based materials.
- 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.

In addition to the generic translation and adaptation guidelines, the translators and reconcilers were given item-specific guidelines within the monitoring sheets that accompanied the materials throughout the localisation process. These guidelines provided help for specific translation and adaptation challenges. The item-specific guidelines were produced based on a thorough review first of the English source, then of the comments arising from the translatability assessment and then of those arising from the production of the French source version.

Centralised trend material transfer

Cognitive units were administered in paper-based format (MS Word) until PISA 2012. In PISA 2015, the majority of participating countries switched the mode of administration from PBA to CBA, but there were still some countries that remained with the PBA. In PISA 2018, some of those countries also switched to CBA.

As the trend contents need to remain identical, the transfer of trend contents from PBA to CBA, i.e. from Word to XLIFF, was centrally managed, as it was in PISA 2015. To do this operation, a semi-automated process (different from the more manual process applied in 2015) was adopted. National centres were then asked to review their transferred units using the preview widget on the PISA portal and report any transfer error or residual issues identified in the trend materials using change request forms (in Excel format). Approved changes were then centrally implemented by the contractors.

Questionnaire Adaptation negotiation and verification.

Questionnaire verification before the Field Trial sought to ensure cross-linguistic equivalence of 106 national versions of the data collection instruments. This process began with the negotiation of national adaptations documented in the Questionnaire Adaptation Spreadsheet, referred to as QAS in this report.

In the questionnaires, national adaptations are defined as intentional deviations from the source, aiming to reflect the national context and to keep the comparability on the international level at the same time. A set of these national adaptations was compulsory, such as country-specific responses options in a question that asks about education levels, types of school, or language spoken at home. Beyond these "forced adaptations", countries could propose requests for additional adaptations in the QAS.

Countries proposed their adaptations to new items in the QAS, provided a back translation in English and justified it, as needed. Based on the back translation and the explanation, the questionnaire team expressed agreement to the proposed changes, or asked the National Centre to further adjust the translation to correspond to the source and ensure across-country comparability. This dialogue between the National Centre and contractors took place in the QAS until an agreement was reached.

The National Centre implemented the agreed adaptations in their national versions. CBA countries encoded the translations directly in the Questionnaire Authoring Tool (QAT).

After having tested the different scenarios (rules and filters) advised by Core A (ETS Data Management), countries uploaded the QAS documenting the negotiation and released the national questionnaires for the next step in the workflow, i.e., verification.

TRANSLATION TRAINING SESSIONS

National project managers received sample materials to use when recruiting national translators and training them at the national level. The NPM meeting held in March 2016 in Prague included sessions 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 separately for each component of the survey (questionnaires, global competence units, new reading literacy units, trend units).

TESTING LANGUAGES AND TRANSLATION/ADAPTATION PROCEDURES

National project managers had to identify the testing languages according to instructions given in the School Sampling Preparation Manual and to record them in a sampling form for agreement.

Prior to the field trial, national project managers had to complete 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 translation referee for discussion/approval.

Table 5.1 summarises the field trial translation procedures for tests and questionnaires, as described in the confirmed translation plans. The figures in the table include minority language versions that represented less than 10% of the target population and were not verified internationally.

Table 5.1: Translation procedures reported by national centres in the translation plan

Type	Cognitive	Questionnaires
Double translation from English and French source versions	19	22
Double translation from English source version with cross-checks against the FRE source version	10*	12*
Double translation from English source version only	16	35
Adaptations in one of the source versions	26	28
Adaptations made in a borrowed verified version or "base" version	46	32
Total	88	88

Note that for the Catalan, Galician and Basque versions, the cross-checks were made against the verified Spanish version of Spain. Also, the totals do not always match between cognitive items and questionnaires, because while most PBA countries had no translation to do for the cognitive domains (trends only), they all had to translate the 2018 School and Student questionnaires.

The lower number of questionnaire versions adapted from a verified or base version versus the same number for tests is largely explained by the fact that a Spanish, Chinese and Arabic base version of the tests was produced, but there was no such base versions of the questionnaires. Therefore, countries that could adapt the Spanish, Chinese and Arabic base version were responsible for translating the questionnaires themselves.

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. It has been found in previous survey administrations that high-quality translations and high levels of equivalence in the functioning of items were achieved in countries that shared a common language of instruction and could develop their national versions by introducing a limited number of national adaptations in a 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 potentially reduces cross-country differences due to translation effects.

Co-operation between countries sharing a same language was therefore fostered and facilitated. To this effect, workable models were designed so that verified versions from one country could be adapted by another country.

Particular instances of this sharing are the following:

- As in previous cycles, the model followed by German-speaking countries was 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, and adapted by the other countries who administered that component. The adapted versions were then verified.
- A Spanish base version of the new test materials was produced by an independent contractor and shared by nine Spanish-speaking countries (Chile, Colombia, Costa Rica, Dominican Republic, Mexico, Panama, Peru, Spain and Uruguay)
- An Arabic version of the new test materials was produced by an independent contractor and shared by two Arabic-speaking countries (Morocco and UAE).
- A Chinese version of the new test materials was produced by an independent contractor and shared by 3 Chinese-speaking participants (Chinese Taipei, Hong Kong (China), Macao (China), then the version for China (People's Republic) was adapted from the verified Hong-Kong version, which was converted from Traditional to Simplified Chinese before the National Centre could adapt it.
- The Russian version from Russia was fully verified and then adapted by Azerbaijan (Baku), Belarus, Estonia, Kazakhstan and Latvia.
- Finally, Bosnia and Herzegovina, Montenegro and Serbia shared the translation effort translating each one part of the assessment and then adapted the verified versions to their local contexts.

Translation of coding guides for open-ended items was not included in the translation plan because, for PISA 2018, the recommended procedure was to single-translate from one source version with cross checks against the other. Some countries produced translated coding guides in one national language only (Spain), while some used the English source (Sweden) or French source (Morocco) without translation.

Table 5.2: Test languages used in PISA 2018

Country/Economy	Language	Country/Economy	Language	Country/Economy	Language	Country/Economy	Language
Albania	Albanian	Dominican Republic	Spanish	Luxembourg	German	Russia	Russian
Argentina	Spanish	Estonia	Estonian		French	Saudi Arabia	Arabic
Australia	English		Russian	Macao (China)	English	United Kingdom (Scotland)	English
Austria	German	Finland	Finnish	Macao (China)	Chinese (trad)	Serbia	Serbian (Ekavian)
Baku City (Azerbaijan)	Azerbaijani	France	French	FYROM	Macedonian	Singapore	English
	Russian	Georgia	Georgian		Albanian	Slovak Republic	Slovak
Belarus	Belarussian	Germany	German	Malaysia	English	Slovenia	Slovenian
	Russian	Greece	Greek		Malaysian		Slovenian
Belgium	Dutch	Hong Kong (China)	Chinese (trad)	Malta	English	Spain	Catalan
	French	Hungary	Hungarian		Maltese	Spain	Basque
Bosnia and Herzegovina	Bosnian	Iceland	Icelandic	Mexico	Spanish	Spain	Castilian
	Croatian	Indonesia	Bahasa Indonesia	Moldova	Romanian	Spain	Galician
	Serbian	Ireland	English		Russian	Sweden	Swedish
Brazil	Portuguese	Israel	Arabic	Montenegro	Montenegrine	Switzerland	German
Brunei Darussalam	English		Hebrew	Morocco	Arabic		French
Bulgaria	Bulgarian	Italy	Italian	Netherlands	Dutch	Chinese Taipei	Chinese
Canada	English	Japan	Japanese	New Zealand	English	Thailand	Thai
	French	Jordan	Arabic	Norway	Nynorsk	Turkey	Turkish
Chile	Spanish	Kazakhstan	Kazakh		Bokmål	United Arab Emirates	Arabic
B-S-J-Z (China)	Chinese (simpl)		Russian	Panama	Spanish		English
Colombia	Spanish	Korea	Korean	Peru	Spanish	Ukraine	Ukrainian
Costa Rica	Spanish	Kosovo	Albanian	Philippines	English	United Kingdom (excl. Scotland)	English
Croatia	Croatian	Latvia	Latvian	Poland	Polish	United States	English
Cyprus ¹	English		Russian	Portugal	Portuguese	Uruguay	Spanish
	Greek	Lebanon	English	Qatar	Arabic	Viet Nam	Vietnamese
Czech Republic	Czech		French		English		
Denmark	Danish	Lithuania	Lithuanian	Romania	Romanian		

CENTRALISED MANAGEMENT OF CHANGES IN TREND

In PISA 2018, a centralised management approach for trend content was implemented for both test and questionnaire materials. The cornerstone of this approach is that all changes to trend content requested by countries/economies went through a strict review and approval process; approved changes were then implemented centrally so that countries/economies did not have editing rights at any stage of the process. This approach prevented unnecessary, undocumented or unverified changes in the trend materials, and thus will allow both more reliable comparability across cycles, and a detailed record of all changes made in trend materials.

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 2018 was to have an independent team of expert verifiers, appointed and trained by the international contractors, verifying each national version against the English and/or French source versions.

International verification was carried out for all national versions in languages used in schools attended by more than 10% of the country's target population.

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 and/or French into their target language
- if 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
- if possible, familiarity with the main domain assessed, in this case, reading literacy
- a good level of computer literacy and experience with computer-aided translation tools (CAT tools)
- if possible, experience as teachers and/or higher education degrees in psychology, sociology or education.

A verifier training seminar was held prior to the verification of the field trial materials. For those who could not attend the training seminar, webinars were organised. The training sessions focused on:

- presenting verifiers with PISA objectives and structure
- familiarising verifiers with the material to be verified, the verification procedures, and the software tools to be used, in particular, the two CAT tool software used for computer-based materials
- conducting a review and extensively discussing the translation guidelines and the verification checklist
- conducting hands-on exercises on specially “doctored” target versions in which typical errors (linguistic issues, adaptation issues, or errors related to guidelines not being followed) had been planted

- arranging schedules and dispatch logistics
- reviewing security requirements.

Verification procedures have been continually improved throughout each PISA round, based on the experience and learning from previous rounds.

VERIFICATION OF NEW COMPUTER-BASED TEST UNITS

Of the 79 countries/economies participating in PISA 2018, nine participated in the paper-based assessment (PBA). The remaining 70 countries/economies participated in the computer-based assessment.

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 PISA portal, a web-based platform that allows the files to travel through a predefined workflow. Another Open Source CAT tool called OmegaT was used for a small group of "pilot" countries/economies and verifiers.

To perform the verification task, the verifiers were instructed to verify the text segments one by one, comparing the target version appearing on the right side of the OLT interface to the source version appearing on the left side, while consulting previews on the portal and the test adaptation spreadsheet (TAS) to see item-specific guidelines and comments from the national centres. They made corrections as needed, documenting their interventions in the test adaptation spreadsheet, including selection of the appropriate intervention category using a drop-down menu.

Once a domain was verified, reviewed and “finalised” on the portal, the translation referee was able to download the test adaptation spreadsheet annotated by the verifier. The referee would then go through each verifier and country comment, and label as “requires follow-up” any crucial issues that could potentially affect equivalence or item functioning.

Changes labelled as “requires follow-up” were negotiated between the referee and the national centre. The national centre then uploaded revised XLIFF files and TAS on the portal for final check. The final check reviewer checked the correct implementation of any changes “requiring follow-up” and either released the files for layout check and national version construction by the international contractors or released them back to the national centre for additional corrections.

Since the PISA 2003 main survey, the central element and repository of the entire translation, adaptation and verification procedure for test units has been the test adaptation spreadsheet. Figure 5.1 shows a sample test adaptation spreadsheet from the PISA 2018 field trial. The spreadsheet functions as:

- an aid to translators, reconcilers, and verifiers through the increasing use of item-specific translation/adaptation guidelines
- a centralised record of national adaptations, of verifier corrections and suggestions
- a way of conducting discussions between the national centre and the translation referee
- a record of the implementation status of “requires follow-up” in test units
- a tool permitting quantitative analysis of verification outcomes.

Figure 5.1: Sample of a test adaptation spreadsheet (TAS) from the PISA 2018 field trial

ENGLISH SOURCE VERSION	ITEM-SPECIFIC TRANSLATION / ADAPTATION GUIDELINE	COUNTRY COMMENT (ADAPTATION, DOUBTS)	VERIFIER INTERVENTION	VERIFIER COMMENT	CONSORTIUM REFEREE COMMENT	CORRECTION STATUS	COUNTRY POST-VERIFICATION COMMENT	FINAL CHECK
Refer to "... " on the right. Type your answer to the question.	Recurring instructions	OK	Inconsistency	1st instruction harmonised with SC645, seg 4	Please make sure to keep the verifier correction		OK	
			Register/Wording	"Stress builds up..." translated as "Stress creates..." Verifier thinks translation in the meaning of accumulating/increasing is more appropriate. Changed by ver.	Please keep the verifier correction	REQUIRES FOLLOW-UP	OK	OK

Verification of homolingual versions

Whenever a country adapted their national version from the English or French source, base version, or same-language verified version of another country, this was considered a homolingual version. The resulting national version was verified using a special procedure for these versions. There were in total 50 CBA homolingual versions that were verified using this process.

The essential difference between the "full" verification of translated national versions and the "focused" verification of homolingual versions is that in the latter, the verification concentrates on the changes made by the country versus the base, source or borrowed version. Automatically-created difference reports were used to identify all such changes in a reliable way.

Verification of paper-based test units and booklet shell

Since no new paper-based units were developed for PISA 2018, PBA countries that had participated in previous cycles did not have anything new that required translation or adaptation. For these countries, the units only went through the centralised change-management process where the country had the opportunity to request corrections to errors, and these – when accepted by the translation referee – were then implemented centrally by the verifiers.

Paper-based countries that were new in PISA 2018 or that had not participated in one or more of the relevant cycles had to translate or adapt units they had not administered before. These were verified following the same process as described above for computer-based materials. The only essential difference was that the verifiers implemented the changes in the MS Word files using the "track changes" functionality, rather than in the online system. The test adaptation spreadsheet was used the same way as in the computer-based verification.

VERIFICATION OF QUESTIONNAIRES

Questionnaires were submitted for verification together with an agreed questionnaire adaptation spreadsheet (QAS). The first purpose of the questionnaire adaptation spreadsheet was to document all content-related or ‘structural’ deviations from the international reference versions. Such national adaptations were subject to clearance by the questionnaire team before the material was submitted for verification. Subsequently, the spreadsheet served the same objectives and followed the same logic as the test adaptation spreadsheet for test units (see above). Figure 5.2 shows a sample questionnaire adaptation spreadsheet from the PISA 2018 field trial.

Figure 5.2: Sample of a questionnaire adaptation spreadsheet (QAS) from the PISA 2018 field trial

National Centre to complete			Questionnaire Team to complete			Verifier to complete		
8a	8b	9	10	11	12	13	14	15
English translation of the national version	Proposed target version in national language	Justification for proposed changes; national centre comments	Questionnaire team Comments	Recode suggestion or other	Agreement status	Verifier intervention category	Verifier comments	Verifier target version
Number of lessons per week in Slovak language	Pocet vyučovacích hodín slovenského jazyka a literatúry týl denne				AGREED	OK	Complete translation in 8a should be: "Number of lessons per week in Slovak language and literature." which is the whole correct name of the subject. Considered OK by ver.	
Number of lessons per week in mathematics	Pocet vyučovacích hodín matematiky týl denne				AGREED	OK		
Number of lessons per week in school science	Pocet vyučovacích hodín prírodovedných predmetov týl denne				AGREED	OK	Adapted as: Number of lessons per week in science subjects. Considered OK by ver.	

The verifiers’ brief was to check 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). In light of this, verifiers were instructed:

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

For the paper-based questionnaires (Student and School questionnaires for countries administering paper-based assessment, Parent Questionnaire for all countries/economies taking this option), verifier interventions were entered in the questionnaires using the track changes mode, while verifier comments were entered in the verifier column of the questionnaire adaptation spreadsheet.

For computer-based questionnaires administered on the questionnaire authoring tool (QAT) platform, the verifier interventions were inserted in the spreadsheet in a separate “Verified

target version” column, in addition to documenting the rationale for the change. The verifiers did not have editing access to the new questionnaire items on the platform. If the change was agreed, the country implemented it in the spreadsheet. In paper-based questionnaires, the verifier introduced the changes in the Word files using the track changes functionality, and documented the changes in the QAS.

As with test units, any more significant changes were labelled as “requires follow-up” by the questionnaire team, and after negotiation with the country teams, their correct implementation was checked by verifiers during final check.

There were no special “homolingual” procedures for the verification of questionnaires since differences in education systems mean that these are very extensively adapted even when sharing a common language. Nevertheless, English and French versions benefited from a co-ordination process similar to the one implemented for test materials. A list of “tips” for verification of questionnaires, including spelling, possibly recurring adaptation issues, and especially errata (errors identified in the source version after release to the countries/economies) and “quasi-errata” (suggestions for improving the source) was maintained, built up, and used in each successive verification.

As in previous cycles of PISA, there was also an increased effort 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 four language versions for Spain). Such versions are by necessity entrusted to different verifiers, but as frequently as possible, cApStAn’s verification reviewers made a point of reviewing and delivering such versions together, striving to harmonise verification interventions on adaptation issues common to the different language versions.

VERIFICATION OF CODING GUIDES

In PISA 2018, the coding guides were verified in a separate process from the test items, and at a later time. This was necessary since many additions and improvements were made to the master versions after the coder training meetings, long after preliminary versions of the guides had been made available to countries/economies. As in PISA 2012 and 2015, the scoring sections were not made available for translation at the time of the unit dispatch. There was one coding guide per trend domain (mathematics, science and reading).

For CBA countries/economies, there was, in addition, one coding guide for new reading units, and for those countries/economies that opted for financial literacy, there was a separate coding guide for this domain.

The overall procedure was the same as for paper-based test units: verifier corrections were made in track changes in the MS Word files, and documented in the monitoring sheets in Excel format. For countries/economies that had participated in previous cycles, trend coding guides underwent a similar controlled change request process as the test units.

MAIN SURVEY VERIFICATION

In previous cycles, the instruments were revised to some extent between the field trial and main survey and were then re-verified in this revised form before the main survey. In PISA 2018, no major changes were made in the master versions after the field trial (apart from entire units or

items being dropped), only few updates had to be made and were registered as errata. Verification therefore consisted of verifying changes that countries/economies requested to their FT instruments, for example based on poor performance or differential item functioning in the FT, or the detection of residual “outright errors” (the latter, in particular, for questionnaires) and implementation of the FT-to-MS errata, which included errata discovered after the last release of the FT errata document and central FT>MS updates. This process was similar to the centralised change management used to control changes in trend: countries/economies requested changes, and the verifiers implemented centrally those changes that were approved by the translation referee. The countries/economies did not have editing access to their units or questionnaires at this stage.

Quantitative analyses of verification outcomes

In PISA 2018, the instruments used to document the verification were designed to generate statistics, thus providing some quantitative data on the frequency of different types of issues identified. 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 gives valuable information on how to avoid such problems during the main survey administration.

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