Chapter 6
Translation: The Preparation of National Language Versions of Assessment Instruments

Paulína Koršňáková, Steve Dept, and David Ebbs

Abstract To ensure the quality of its large-scale comparative studies, the International Association for the Evaluation of Educational Achievement (IEA) has created technical standards and guidelines, and developed a wealth of experience related to instrument production. The content experts who served as test editors and the professional linguists working toward achieving equivalence of assessment materials during the verification processes informed this chapter: it provides an overview of the rationale and implementation of linguistic quality control procedures for the instruments used in IEA studies, covering the general steps, key aspects, vocabulary, and discourse related to the translation of the study instruments. There are many challenges related to fine-tuning translations and preparing the national instruments used in IEA's large-scale comparative studies in education. The use of technology and development of tailored online translation systems can contribute here, both by guiding and streamlining the workflow and by guarding consistency. The numerous consecutive steps required to produce any national version of a large-scale international comparative assessment that strikes a good balance between faithfulness to the source and fluency in the target language are increasingly regarded as a collaborative effort between international and national experts and stakeholders rather than a system of checks and balances.

P. Koršňáková (✉) · D. Ebbs
International Association for the Evaluation of Educational Achievement (IEA),
Amsterdam, The Netherlands
e-mail: p.korsnakova@iea.nl

D. Ebbs
e-mail: d.ebbs@iea.nl

S. Dept
cApStAn Linguistic Quality Control, Brussels, Belgium
e-mail: steve.dept@capstan.be

6.1   Introduction

Producing comparable national versions of the international source instruments[1] is a key methodological issue in international comparative studies of learning outcomes such as those conducted by IEA. Once the validity and reliability of the international source version is established, procedures need be put in place to ensure linguistic equivalence of national versions with a view to collecting comparable data. As outlined in Chap. 2, the assessment landscape has changed considerably, not only in terms of the absolute numbers of participants (national and sub-national) but also in terms of their linguistic, ethnic, and cultural heterogeneity. In this chapter, our focus is on understanding the complexity of the issues as they relate to translation and the methodological response to this changing landscape.

English is a relatively concise Indo-European language with a simple grammatical and syntactic structure. A straightforward translation of a question into a language with a more complex structure may result in an increased reading load. When measuring education outcomes in, for example, mathematics or science, a fair translation should not increase reliance on reading proficiency. To achieve this level of fairness, a subtle alchemy of equivalence to the source version, fluency in the target language, and adaptations devised to maintain equivalence needs to be applied and monitored. Language components such as sentence length, quantifiers, direct speech, use of interrogatives, reference chains, active or passive voice, use of tenses, use of articles, idiomatic expressions, use of abbreviations, foreign words, or upper case may all need to be treated differently depending on the target language. For example, Slavonic languages do not use articles, in Chinese the context provides additional information, there is no upper case in Thai or in Arabic, there are more possible forms of address in Korean: all these elements have to be balanced so that comparability across language versions can be maximized.

6.2   Translation Related Developments in IEA Studies

The first notes on translation in IEA studies can be found in the results of an international research project undertaken between 1959 and 1961 on educational achievements of thirteen-year-olds in twelve countries (Foshay 1962). Although translation was of great concern for the participants, it was not the main focus of the study, so they agreed to leave to each participant the translation of the items into their own language. The most interesting feature of the recorded procedures followed in translating (mainly already existing tests originally developed in England, France, Germany, Israel, and the United States) into eight languages was the role of

---

[1]Instruments here refers to the test and questionnaire items, including the test and questionnaire booklet covers, introductory texts, and any scripts read by test administrators. Other important assessment materials, such as manuals and scoring guides, are reviewed as part of the international quality control measures.

a test editor, who reviewed any criticism and suggestions gathered through a pre-test (involving a small number of children in each country) and approved the test prior to its duplication and circulation, including approval of any alterations in the substance of items (such the change in units of measure to conform with the custom of the country). As expected, some difficulties in translation were found and reported, but these were "small in number and so scattered as to be insignificant" (Foshay 1962, p. 19), and there was no evidence that these would have influenced the national scores.

In the late 1960s and in 1970s, however, researchers began to realize that the operation of translating an assessment (or, as a matter of fact, any data collection instrument) into different languages, for use in different cultural contexts involved a cluster of challenges, each of which had implications on fairness and validity. Articles in the literature emerged claiming that translation and adaptation changes test difficulty to the extent that comparisons across language groups may have limited validity (Poortinga 1975, 1995). As a consequence, linguistic quality control methods were introduced in the 1970s, mostly to check the linguistic quality and appropriateness of tests translated from English and, more importantly, their semantic equivalence versus the English source version.[2]

IEA has implemented different translation procedures over time. Ferrer (2011) highlighted three careful steps: (1) comparative curricular analysis, and creation of the conceptual frameworks and specification tables; (2) cooperative procedures like collaborative item development, and multiple discussion and review stages; and (3) rigorous procedures for (back) translation of the items into the different languages of the participating countries. The last step was in line with the initial focus on validation of the translated versions of the assessment through a back translation[3] procedure (Brislin 1970, 1976). It should be noted, however, that Brislin pointed out the limitations of back translation.

With increasing interest and scrutiny in cross-national studies, IEA's Third International Mathematics and Science Study in 1995, embarking on a cyclical trend assessment that was to become known as the Trends in International Mathematics and Science Study (TIMSS), made ensuring the validity of translations a priority and commissioned in-depth research into translating achievement tests (Hambleton 1992). With Hambleton's report as a guide, TIMSS 1995 established the basis for all IEA's translation procedures. To verify the translations, TIMSS 1995 relied on multiple forward translations, translation review by bilingual judges (translation verifiers), and, due to concerns about the limitations of a back translation approach, a final statistical review. Reasons for not using back translation include the resources needed for back translation and the concern that flaws in the translation can be missed if the back translator creates a high quality English back translation from a

---

[2]In this chapter, the term "source version" systematically refers to the language version from which the materials are translated into other languages, resulting in "target versions." The source language is the language of the source version; the target language is the language of a specific target version.

[3]Back translation is a three-step procedure. The test is translated from English into the target language; a different translator translates that version back into English, and finally an English-speaking person compares the original test with the back-translation (see Hambleton 1992).

poor quality initial translation, thus resulting in a non-equivalent translated national version (Hambleton 1992). Thus, back translation was not used in TIMSS 1995 (Maxwell 1996) nor in any later IEA study.

6.3 Standards and Generalized Stages of Instrument Production

In IEA's technical standards (Martin et al. 1999), the task of translating and verifying translations is mentioned under standards for developing data collection instruments. In addition, when discussing standards for developing a quality assurance program, Martin et al. (1999) acknowledged that IEA studies depend on accurate translation of materials (e.g., tests, questionnaires, and manuals) from the source version (usually English) into the target languages of participating countries. Consequently, IEA makes every attempt to verify the accuracy of these translations and ensure that the survey instruments in target languages conform to the international standard, and that "no bias has been introduced into survey instruments by the translation process" (Martin et al. 1999, p. 27).

Verification of the translations of instruments into the different languages of the participating countries is seen as a measure ensuring that the translated instruments will provide comparable data across countries and cultures, or more concretely "that the meaning and difficulty level of the items have not changed from the international version" (Martin et al. 1999, p. 32). The standard is set as follows: "When translating test items or modifying them for cultural adaptation, the following must remain the same as the international version: the meaning of the question; the reading level of the text; the difficulty of the item; and the likelihood of another possible correct answer for the test item" (Martin et al. 1999, p. 43). In addition, verification is supposed to keep the cultural differences to a minimum, and retain the meaning and content of the questionnaire items through translation.

The survey operations documentation provides guidelines and other materials for the participating countries that describe the translation and cultural adaptation procedures, the process for translation verification, and serve as a means to record any deviation in vocabulary, meaning, or item layout. During translation and adaptation, participants can submit their adaptations to the international study center (ISC) for approval. Upon completion of the translation and prior to finalizing and using the instruments, countries submit their translations to the ISC for verification of the translations and adaptations of the tests and questionnaires. Professional translators also assess the overall layout of the instruments: "[t]he professional translator should compare each translated item with the international version and document the differences from the international version" (Martin et al. 1999, p. 44).

After the completion of the verification, the national center receives the feedback and addresses the deviations (e.g., incorrect ordering of response options in a

multiple-choice item, mislabeling of a graph that is essential to a solution, or an incorrect translation of a test question that renders it no longer answerable or indicates the answer to the question).

While this chapter provides a simplified overview of the process of securing high quality translations (Fig. 6.1), determining and conveying what this process actually aims to achieve is a far more complex endeavor.

In the standards, field testing is (among its other benefits) seen as a way to produce the item statistics that can detect and reveal errors in the translation and/or adaptation processes that were not corrected during the verification process and check on any flaws in the test items. Through the use of item statistics from the field test, items may be either discarded or revised and corrected, minimizing the possibility of translation errors in final versions of the test instruments.

While some problems arising from the translations of the English versions of attitudinal and value statements in the questionnaire scales were already noted in relation to the First International Mathematics Study (FIMS; Ainley et al. 2011), few further details were provided. Van de Vijver et al. (2017) conducted some further analyses using IEA data that indicate that part of the problems attributed to translations could be related to response styles and cultural differences rather than translation errors.
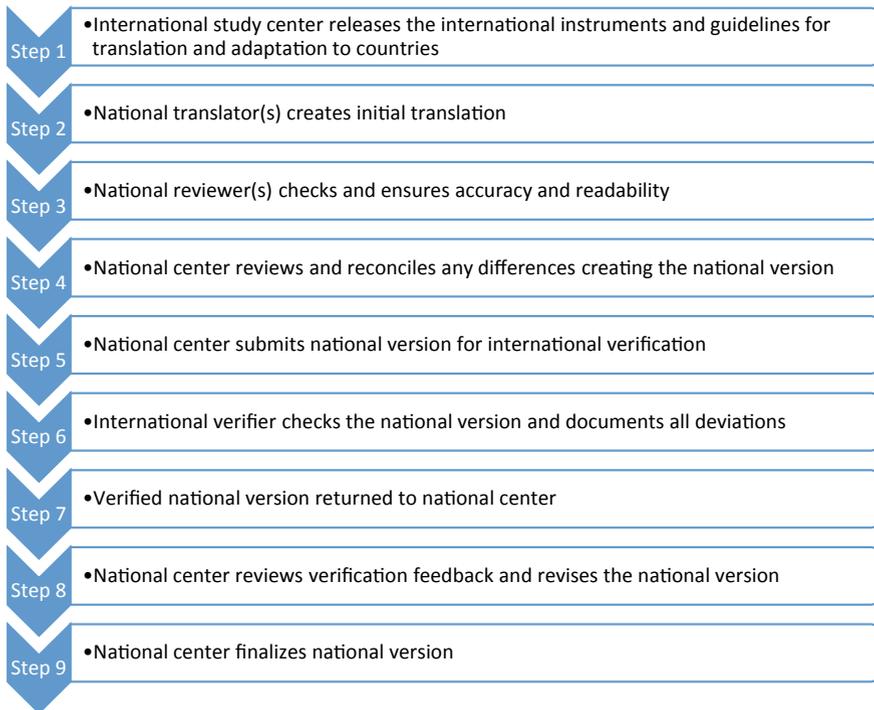

Step 1 • International study center releases the international instruments and guidelines for translation and adaptation to countries

Step 2 • National translator(s) creates initial translation

Step 3 • National reviewer(s) checks and ensures accuracy and readability

Step 4 • National center reviews and reconciles any differences creating the national version

Step 5 • National center submits national version for international verification

Step 6 • International verifier checks the national version and documents all deviations

Step 7 • Verified national version returned to national center

Step 8 • National center reviews verification feedback and revises the national version

Step 9 • National center finalizes national version

Fig. 6.1 Generalized and simplified stages of instrument production

6.4   Source Version and Reference Version

Nowadays, in international large-scale assessments (ILSAs), it is widely accepted that there is an international source version that creates a base for development of national instruments. In IEA studies, this is crafted in a collaborative effort by non-native and native speakers of the source language, which is English. By the time that this source version is ready for translation, it has been through a number of revisions, piloted, and/or gone through one or several rounds of cognitive pre-testing. By the time the source version is released for translation, it is regarded as a mature draft of the data collection instrument.

In this context, translation can be viewed as an attempt to mirror the source version in the target languages, under the assumption that the highest degree of faithfulness to the source version will be conducive to the highest degree of functional equivalence. If the quantity and quality of the information present in the source version is scrupulously echoed in the target version, the translated instrument should function the same way as the original. This, however, is not a given: Hambleton and Patsula (1999) and Hambleton (2002) described some common myths in what they refer to as test adaptation.

6.4.1   Terms Used: Translation Versus Adaptation

It should be noted that authoritative authors propose different definitions of "test translation" and "test adaptation." They share the view that the term "translation" is too narrow to capture the scope of the challenges that need to be addressed when producing multiple versions of assessment instruments while considering the objectives of functional equivalence and cross-linguistic comparability. Joldersma (2004) deemed the term "translation" too restrictive to describe the process of culturally adjusting a test rather than just translating it literally. Hambleton et al. (2005) suggested that the term "test adaptation" is preferable. Harkness (2003, 2007) and Iliescu (2017) regarded test translation as a subset of test adaptation. Iliescu (2017) explained that test translation is linguistically driven (content over intent), while test adaptation is validity-driven (intent over content), and this is certainly a workable distinction. In this chapter, however, we shall use the term translation in a broad sense and adaptation in a narrower sense, because we view the latter as an integral part of translation process.

The aim of a test translation should be to minimize the effect of language, culture, or local context on test difficulty. A straightforward translation process would not ensure this fairness. To prevent a given target population or culture being placed at an advantage or a disadvantage, it is necessary to deviate from the source version to some extent. If, for example, a general text contains references to July and August as summer months, the translator will need to consider whether, for countries in the southern hemisphere, it would be preferable to keep July and August but refer

to them as winter months; to change July and August to January and February; or to translate literally and explain in a note that this text refers to the northern hemisphere. In this light, we use the following working definition for adaptation, used by the Organisation for Economic Co-operation and Development (OECD) in its Programme for International Student Assessment (PISA):

> An adaptation is an intentional deviation from the source version(s) made for cultural reasons or to conform to local usage (OECD 2016, p. 3)

An adaptation is needed when there is a risk that respondents would be (dis)advantaged if a straightforward translation were used. While general guidelines for test translation may prescribe that each translated item should examine the same skills and invoke the same cognitive processes as the source version, while being culturally appropriate within the target country, this is a tall order. One of the myths described in Hambleton (2002) is that "translators are capable of finding flaws in a test adaptation."

No honest linguist or psychometrician can claim that the combination of a robust translation design and an expert translation verification will ensure that items examine the same skills or elicit the same cognitive processes in the source version and in the target versions. However, IEA procedures have been established to maximize comparability and ensure that the most egregious errors or misunderstandings are avoided.

In comparative assessments, cross-linguistic, cross-national, and cross-cultural equivalence is not only an objective but also a fundamental requirement without which the whole notion of quantitative cross-cultural comparison is invalidated (Dept et al. 2017). Even in translations produced by the most experienced professionals, verified by local subject matter experts, by teachers, and by trained reviewers, readers may still observe language-driven meaning shifts and/or culture-driven perception shifts. While dictionaries may provide a direct equivalent for the word "coffee" in most languages, the cultural context will lead to a different semantic loading: it is hardly possible to sip on an Italian coffee for half an hour and it would not be safe to gulp down a mug of American coffee in four seconds (Eco 2003).[4]

The concept of "mother tongue" translates as "father tongue" in some languages, and as "language of the ancestors" or "language of the fatherland" in others, with all the different connotations that this implies (Banks 2006).

Therefore, maximizing cross-language and cross-cultural comparability is a subtle balancing exercise, whereby (1) different players work together to strive for a balance between faithfulness to the source version and fluency in the target version; and (2) at the test and questionnaire design stage, it is desirable to identify concepts, terms, or contextual elements that will need to be adapted for which an intentional deviation is required or desirable to maintain equivalence, while a literal translation might jeopardize this equivalence.

---

[4]The example taken from Umberto Eco's *Experiences in translation* illustrates the challenge of semantic portability in general and is not related to IEA assessments.

6.4.2 Collaborative Efforts

In parallel with the growth in awareness as outlined in the previous section, IEA studies like TIMSS (see Fig. 6.2) and the Progress in International Reading Literacy Study (PIRLS; see Fig. 6.3) also accommodate a growing number of participating countries, and, even more importantly, a growing number of the national sets for verification, exceeding the number of participating countries. The latter shows that implementation of IEA assessments at the national level became more inclusive of different language minorities. It is important to point out that for the number of languages listed, all versions of a language (e.g., English) are combined and counted as one language, making the number of languages lower than the number of the countries involved. The different versions of a language (e.g., British English, American English, and Australian English) are accounted for in the number of verified national sets.

As in every translation procedure, the quality of the professionals who are involved in the process is one of the key determinants of a high quality result (Iliescu 2017). Considering that the majority of ILSAs (and IEA studies are no exception) have adopted a decentralized translation approach whereby national research centers are responsible for the translation of assessment instruments into their language(s) of instruction, it is important to agree on as many procedural aspects as possible to reduce disparities. Different national study centers (NSCs) may have different approaches to translation: some may outsource the translation to language service providers, while others will produce the translation in-house with more involvement of subject matter experts than professional linguists. With a view to keeping the
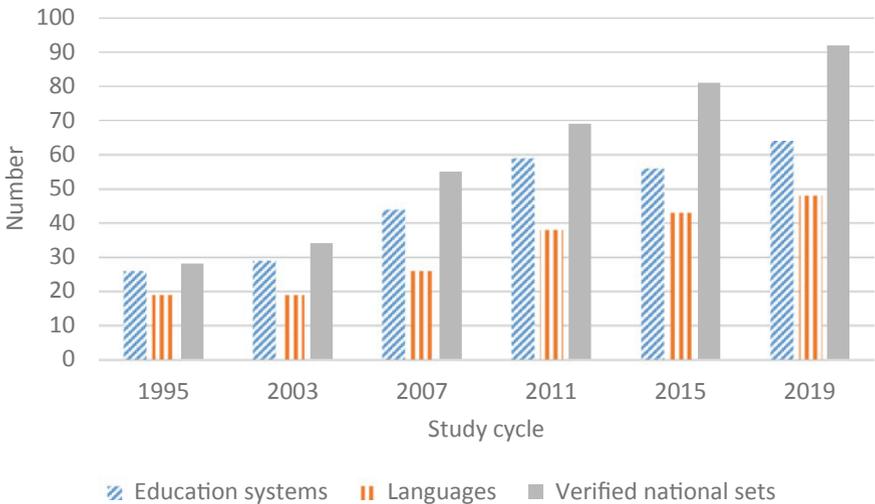

Fig. 6.2 Number of education systems, languages, and verified national sets involved in each cycle of the TIMSS grade 4 from 1995 to 2019
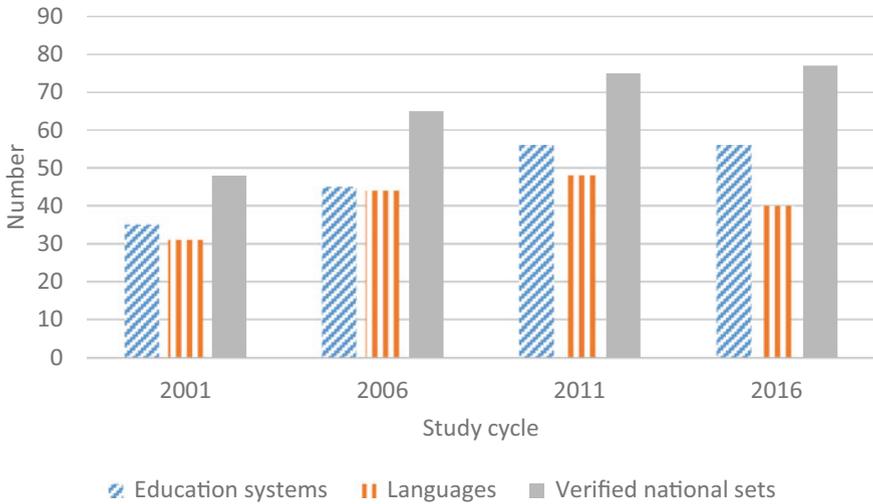
Fig. 6.3 Number of education systems, languages, and verified national sets involved in each cycle of the PIRLS from 2001 to 2016

"translator effect" in check, IEA prepares comprehensive translation and adaptation guidelines for the different instruments and tools and offers extensive technical support to NSCs during the translation and adaptation process.

While working on translations, some countries engage in collaborative efforts, like producing jointly translated instruments or sharing translations. These efforts improve quality (more faithful and fluent target versions of instruments), because such translations undergo additional reviews and the collaboration involves discussions that can reveal differences in understanding and facilitate clarification.

As the studies developed, IEA has engaged in additional efforts. In TIMSS 2007, Arabic became the largest linguistic community; IEA therefore prepared an Arabic reference version of the international instruments for Middle East and North African countries (which was based on the international source version and prepared after its release), providing Arabic-speaking countries with an initial translation of the instruments that could be easily adapted or used as a starting point for creating their national instruments. IEA oversaw and managed the collaborative process of creating the Arabic reference version in cooperation with cApStAn (an independent linguistic quality control agency in Brussels, Belgium) and staff at the TIMSS & PIRLS International Study Center at Boston College in the United States.

The process of creating the Arabic reference version began with the creation of an initial translation produced by a skilled team of translators from different Arabic-speaking countries. Following the IEA translation and adaptation guidelines, each translator produced a separate translation that a reviewer checked and compared against the other translations. The reviewer selected the best translation from the translators for use in the Arabic reference version. Upon completion of the translation, a panel of experts with experience and knowledge of school subjects at the

target grades reviewed the translation. In addition to reviewing the translation, the experts checked the consistency and correctness of terminology and commented on possible translation and adaptation issues. Based on the feedback from the experts, the translation underwent further revisions. Then the revised translation was sent to the TIMSS & PIRLS International Study Center for production of the instruments to be released to the countries (For an example of the overall translation and translation verification procedures in TIMSS 2015, please refer to Ebbs and Korsnakova 2016).

6.5 Translation and Adaptation

As indicated earlier in this chapter, we use the term translation in its broadest sense and with full awareness of the limitations of literal translations. Nevertheless, using different scripts, implementing spelling reforms, working within the grammatical constraints of the target language, and trying to achieve a subtle balance between faithfulness to the source and fluency in the target language can all reasonably be regarded as the remit of a trained professional translator. Conversely, determining which intentional deviations are acceptable, desirable, required, or ruled out should be regarded as the remit of the test authors. The latter may, of course, seek advice from cultural brokers and subject matter experts in the target culture or language. In this context, the working definition we have adopted here for adaptation (see Sect. 6.4.1) is relevant in the context of IEA assessments.

In its international studies, IEA prepares guidelines for the adaptation of test instruments and questionnaires, in which the focus is on clear prescriptions about adaptations that are required, desirable, acceptable, and/or ruled out.

The following general information is based on the guidelines for IEA studies. It aligns with the most recently published chapters on translation and verification for the completed studies (see, e.g., Malak et al. 2011; Noveanu et al. 2018; Yu and Ebbs 2012).

The two distinct steps within the production of national instruments (translation and review) at the participating countries level are designed to build up the comparability as well as linguistic quality of the national instruments. When translating (and adapting) the study instruments, NSCs are advised to pay attention to the following:

- finding words/terms and phrases in the target language that are equivalent to those in the international version;
- ensuring that the essential meaning of the text and reading level do not change;
- ensuring that the difficulty level of the items does not change;
- ensuring correspondence between text in the stem/passage and the items;
- ensuring that national adaptations are made appropriately; and
- ensuring changes in layout due to translation are minimized.

When NSCs review their translations, they are advised to use the following guidelines to evaluate the quality of their national translations:

- the translated texts should have the same register (language level and degree of formality) as the source texts; if using the same register could be perceived as inappropriate in the target culture, then the register needs to be adapted and this needs to be documented;
- the translated texts should have correct grammar and usage (e.g., subject/verb agreement, prepositions, or verb tenses);
- the translated texts should not clarify or remove text from the source text and should not add more information;
- the translated text should use equivalent social, political, and historical terminology appropriate in the target language;
- the translated texts should have equivalent qualifiers and modifiers appropriate for the target language;
- idiomatic expressions should be translated appropriately, not necessarily word for word; and
- spelling, punctuation, and capitalization in the target texts should be appropriate for the target language and the country's national context.

For assessment materials, some words or phrases might need to be adapted in order to ensure that the students are not faced with unfamiliar concepts, terms, or expressions. Common examples include reference to the working week, units of measurement, and expression of time. For questionnaires, some words and phrases require adaptation to the country specific context. To aid NSCs in identifying the words and phrases requiring adaptation, the text is placed in carets (angle brackets) in the international source version. Examples of such required (sometimes referred to as forced, obligatory, or compulsory) adaptations are < language of test >, < target grade >, and < country >.

Examples of acceptable adaptations include: fictional names of people and places that can be changed to other fictional names; measurement units that can be changed from imperial to metric or vice versa with correct conversions/numerical calculations (e.g., 3000 feet to 900 m); time notation (e.g., 2:00 p.m. to 14:00); the names of political institutions (e.g., parliament to congress) or representatives that may need to be adopted to the local context; and the names of school grades (fourth grade to year 5), programs, or education levels.

The above examples of adaptation illustrate the requirement to conform to local context and usage as necessary. However, it is useful to note that, in international assessment instruments, there should also be a requirement to deviate from the international source version each time a straightforward, correct translation is likely to put the respondent at an advantage or at a disadvantage. If a well-crafted, linguistically equivalent translation elicits different response strategies, test designers should consider adapting the translation to approach functional equivalence. For contextual questionnaires, where the notion of advantage or disadvantage does not apply, deviations from the international source version need to be considered when a straightforward translation is likely to introduce a perception shift and could affect response patterns.

In this sense, a reference to a student's boyfriend or girlfriend may become a more potent distractor in a predominantly Muslim country, for example. If this can be adapted to the student's cousin or niece without changing the information needed to respond to the question, this may be a desirable adaptation.

Likewise, if a country proposes to use a music instrument as a < country-specific wealth indicator > , there needs to be a clear assessment establishing how owning a music instrument is a socioeconomic status marker or if the term "wealth" could have been perceived as "cultural wealth" rather than "economic wealth," leading the translation to assume a different underlying construct.

6.6 Decentralized Translations and Adaptations

IEA studies have adopted the decentralized approach where the NSCs are responsible for translating and adapting the instruments into their language(s) of instruction. To aid the NSCs, the ISC always releases documents and manuals intended to guide NSCs through the processes and procedures of instrument preparation. Activities covered in the documents include:

- translating and/or adapting the study instruments;
- documenting national adaptations made to the study instruments;
- international verifications (translation, adaptation and layout); and
- finalizing the national instruments for administration.

For the process of translating and adapting the study instruments, the advice given to NSCs of early IEA studies was to have multiple translators create translation that would be consolidated into a single version by the NSC. Over the years, the advice has evolved to the use of at least one translator and one reviewer per language. The recommended criteria for the translator(s) and reviewer(s) are (see Ebbs and Friedman 2015; Ebbs and Wry 2017; Malak et al. 2011; Noveanu et al. 2018; Yu and Ebbs 2012):

- [an] excellent knowledge of English;
- [an] excellent knowledge of the target language;
- experience of the country's cultural context;
- [a familiarity with survey instruments], preferably at the level of the target grade; and, if possible,
- experience in working with students in the target grade.

The translator creates the initial national version by translating and adapting the international study instrument according to the translation and adaptation guidelines provided by the ISC. If an NSC uses more than one translator to create multiple translations of a single language version, it is the NSC's responsibility to review the translations, reconcile the differences, and produce a single version of the

instruments in that language. Upon completion of the initial national version, the reviewer proofreads and checks that the translation is of high quality, accurate, and at an appropriate level for the target population. If an NSC uses more than one reviewer, the NSC is responsible for reviewing all feedback from the reviewers and ensuring the consistent implementation of any suggestions or changes. If an NSC prepares translations in more than one language, they are advised to use professionals that are familiar with the multiple languages to ensure consistency across the national language versions. Before submitting their national language version(s) for international translation verification, the NSCs are advised to perform a final review of the language version(s) in an effort to reduce and prevent errors in the instruments that will be verified.

6.7 Centralized Verification

The international verification consists of three steps: adaptation verification, translation verification, and layout verification. The order in which these verification steps are conducted has changed over the years.

These quality control steps are centralized. For example, during the translation verification (TV) stage the ISC may: (1) entrust this step to an external linguistic quality assurance (LQA) provider; (2) perform an internal review of the feedback provided by this LQA provider; (3) send the feedback to the NSCs, who have the opportunity to review, accept, reject, or edit the LQA interventions; and (4) perform a formal check, including a layout check of the final version after TV and review of TV are completed.

Prior to TIMSS 1995, the national versions of study instruments underwent national verification procedures but did not undergo international verification procedures. The main reason was related to limited resources for conducting the studies. This resulted in a dependence on the data analysis for identifying and removing non-comparable items from the database based on discrimination and item functioning. With increased funding and requirements for verifying and ensuring the comparability and quality of the data collected, international verification procedures were put in place to support the quality and comparability of the national instruments during the field test stage and prior the main data collection.

In TIMSS 1995, the international translation verifiers conducted layout, adaptation, and translation verification at the same time (Fig. 6.4). The initial international verification procedure required the international verifiers, professional translators, to check the layout of the national instruments, followed by comparing the national translation against the international source version. If the national versions differed in any way, the translation verifiers documented the deviations. Upon completing the verification, the verifiers reported their findings to the international coordinating center, ISC, and NSCs. The NSCs reviewed the verification report and implemented
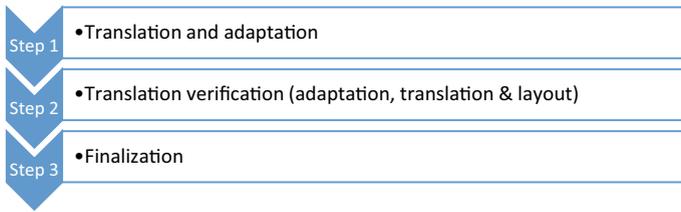
Fig. 6.4 Translation related steps in TIMSS 1995

the comments and suggestions that improved their national instruments. In addition, the verification reports were consulted during the data analysis when anomalies were found for possible translation related explanations.

With each new study and cycle, IEA's focus on ensuring the quality and comparability of the data evolved, leading to changes in the procedures. Starting with PIRLS 2006 (Malak and Trong 2007) and the Second Information Technology in Education Study (SITES) 2006 (Malak-Minkiewicz and Pelgrum 2009), the responsibility for layout verification shifted from the translation verifiers to the ISCs. This change in procedure (see Fig. 6.5) occurred to allow the translation verifiers to focus more on ensuring the quality and comparability of national translations.

Starting with TIMSS 2007 (Johansone and Malak 2008), the ISC also assumed the responsibility for adaptation verification. This change allowed the translation verifiers to further concentrate on the linguistic aspects.
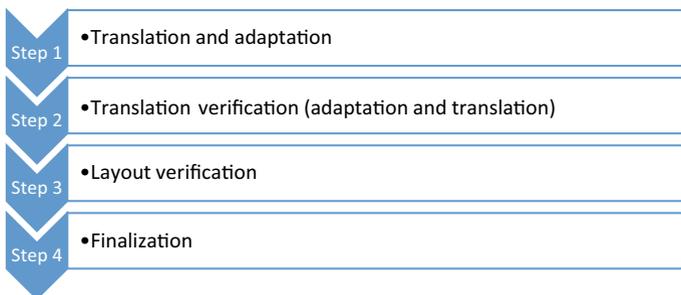

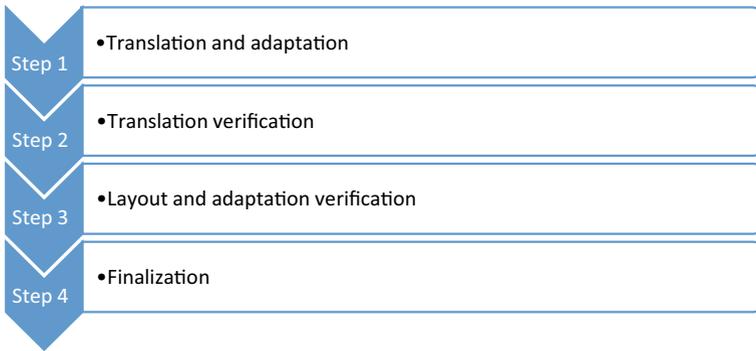**Fig. 6.5** Translation related steps in PIRLS 2006 and SITES 2006

Fig. 6.6 Translation steps in TIMSS 2007

The separation of adaptation verification from translation verification led to the creation of two different pathways for the verification procedures. In the first case (e.g., as followed in TIMSS 2007; see Fig. 6.6), translation verification is conducted first (for information how this situation is handled by verifiers, please see the code 4 description in Sect. 6.8), followed by layout and adaptation verification. This option requires fewer resources and allows for a shorter timeline for completing the verification steps. One concern with this path involves situations when a national adaptation is not approved during adaptation verification and requires further changes. Upon approval of the revised adaptation, the sole responsibility for ensuring the quality of the revised translation resides with the NSC.

The second case (e.g., as followed in ICCS 2009; see Fig. 6.7) starts with adaptation verification, then translation verification, followed by layout verification. This path requires more resources and time than the first path, but ensures that all adaptations are approved and the translations revised prior to translation verification. During translation verification, the translation verifiers review the approved adaptations ensuring the correctness of the documentation and implementation of the adaptation in the national instruments.
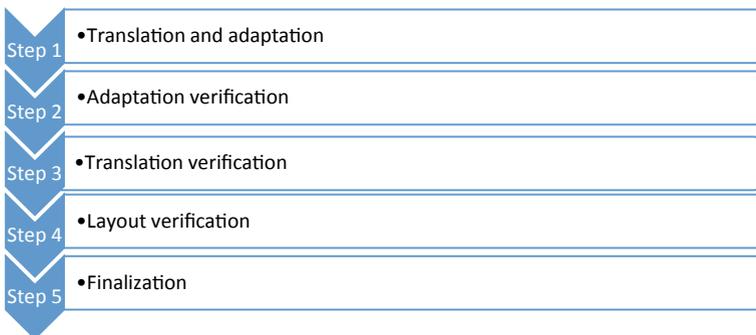

Fig. 6.7 Translation steps in ICCS 2009

6.8   Translation Verifiers

Unlike the test editors in early IEA studies, the external reviewers are linguists, not domain/content nor measurement experts, and so it is important that they judge on the linguistic aspect rather than making decisions on whether the changes that occur are appropriate.

The current IEA severity codes used by translation verifiers are:

Code 1 Major change or error: These changes could affect the results. Examples include incorrect ordering of choices in a multiple-choice item; omission of an item; or an incorrect translation that indicates the answer to the question.

Code 2 Minor change or error: These changes do not affect the results. Examples include spelling and grammar errors that do not affect comprehension; or extra spacing.

Code 3 Suggestions for alternative: The translation may be adequate, but the verifier suggests a different wording.

Code 4 Acceptable changes: Used to identify and document that national conventions have been properly documented and implemented.

If in doubt, verifiers are instructed to use Code 1? as an annotation so that the error or issue is referred to the ISC for further consultation.

If the translation verifier finds a change or error in the text while verifying the national version of the instruments, they are instructed to correct the error or add a suggestion and document the reason for the intervention. Included in the documentation of the intervention, the verifiers are to assign a code to indicate the perceived severity of the change or error corrected. A concern with the use of the severity code relates to their subjective nature. An example of this relates to the possibility that one verifier could list a grammar issue as a Code 1 error and another verifier could consider the same grammar issue to be a Code 2.

When reviewing the verifier feedback, the severity code does not inform the reader about the type of intervention performed, but does indicate the possible influence the error could have had on the item. For more information about the intervention, additional comments are added following the severity code.

IEA prepares clear and concise translation verification guidelines for the verifiers. Differences in verification style need to be kept to a minimum to avoid compounding the translator and verifier effects. It is the LQA provider's responsibility to: (1) adopt a coherent, prescriptive stance on procedures and their implementation; (2) supplement the IEA's guidelines with a face-to-face or a web-based training session for verifiers; (3) provide continuous technical and procedural support to verifiers; and (4) review the feedback provided by the verifiers, clear residual issues, and check the consistency of verifier interventions within and across instruments before returning the feedback to IEA for a second review.

The face-to-face or web-based training sessions for verifiers typically consist of:

- a general presentation on the aims of IEA studies, and on IEA's general and project-specific standards as regards translation and adaptation;
- a presentation of the materials to be verified;
- a presentation of the verifiers' tasks;
- hands-on exercises based on a selection of items from the study under verification (here a variety of errors and/or controversial adaptations may be introduced and the verifiers asked to identify the problem and assign the appropriate code); and
- information about the technical characteristics of the tools, formats, or environment in which the verification needs to be performed.

After the training, when the national version is dispatched, the package sent to verifiers includes the project-specific verification guidelines made available by IEA, a link to a recorded webinar, and a link to a resource page with step-by-step instructions on how to verify cognitive tests and questionnaires.

To further reduce disparities between commenting styles, an additional measure was implemented during the TIMSS 2019 main study translation verification. The measure required translation verifiers to select an IEA severity code and category from drop-down menus followed by using a corresponding standardized comment for each problem spotted (see Table 6.1).

Table 6.1 Examples of standardized comments in each linguistic category

Category	Example of predefined standardized comment
Formatting	Text not bolded/italicized/underlined/capitalized as in source. Changed by verifier
Grammar	Prepositional error. Corrected by verifier Syntax error. Corrected by verifier Incorrect tense. Corrected by verifier Alignment/agreement error. Corrected by verifier
Inconsistency	Translation inconsistent within item. "…" Harmonized by verifier Translation inconsistent between items. "…" Harmonized by verifier
Missing translation	Omission of text. "…" Added by verifier Untranslated text. Translated by verifier
Punctuation, symbols, and spelling	Punctuation error. Corrected by verifier Incorrect use of symbol in target. Corrected by verifier Spelling error/typo. Corrected by verifier

6.9   Layout Verification

Since differences to the layout can also affect the international comparability of the data, the ISC conducts a verification of the national instrument layout. During layout verification, the national instruments are compared to the international instruments and any discrepancies are documented. The layout verifiers check items such as the pagination, page breaks, text formats, location of graphics, order of items, and response options. All differences found are documented and need to be corrected before the national instruments are sent for printing. The goal of layout verification is to ensure minimal deviations in the comparability of the layout of national instruments. Since different languages require a different amount of space and page sizes differ in some countries, the international version of the instruments is designed with extra space in the margins to accommodate the differences in text length and page sizes. These differences are taken into consideration during layout verification. In digital assessments, the layout verification is followed by the player review and then similar duplication (of USB display instead of printing the assessment booklets) and distribution.

6.10   Development Linked to Computer-Based Assessment

As digital technologies have advanced since the millennium, the demand to use these technologies for large-scale educational assessment has increased (Walker 2017).

While IEA has been investigating the role of information and communication technology (ICT) in teaching and its use by teachers and students for a long period (see Law et al. 2008), the International Computer and Information Literacy Study (ICILS) 2013 and PIRLS 2016 were the first IEA studies to administer tests and questionnaires to students on computers. IEA contracted the development of ICILS 2013 to SoNET Systems,[5] covering the costs related to the development of computer-based assessment (CBA), as well as its use for the data collection (head counts), but the ePIRLS 2016 instruments were developed in-house by IEA. From the experience gained while using the SoNET Systems platform for ICILS 2013, IEA saw the possibilities of using an online platform for instrument production and, combined with the knowledge of the process and procedures of instrument production for international studies, IEA began development of the IEA eAssessment system. For ePIRLS 2016, development began with the translation system. The goal was to create a system that was easy to use and incorporated all the basic functions needed for translating and adapting instruments through the stages of verification. From the experiences encountered during ePIRLS 2016, further improvements to the IEA eAssessment system were implemented for TIMSS 2019.

While descriptions of the immense potential and numerous advantages of technology-based assessments abound, the transition from pencil-and-paper tests

---

[5]SoNET was acquired by RM Results in 2019.

to computer-delivered assessments has also given rise to considerable challenges. In the field of test translation and adaptation, this transition may have left some old school translators behind. At the same time, it has sometimes been experienced as a step backwards by professional language service providers. This is partly due to insufficient awareness of the complexity of translation/adaptation processes by system architects, who have frequently chosen to include some translation functionalities in the platform, but without giving consideration to exploring how the power of state-of-the-art translation technology could be harnessed. When discussing translation in computer-based delivery of cognitive assessment and questionnaires, Walker (2017, p. 249) stated that "at a minimum, a method should be available to replace text elements in the source/development language with translated/target equivalents."

This minimum approach fails to recognize that translation technology evolved considerably in the 1990s, and that computer-assisted translation tools (CAT tools) went mainstream by the end of the 20th century. Translation memories, term bases, spelling and grammar checkers, and style guides are functionalities that most professional translators use on a daily basis, so that, by the time international large-scale assessments transitioned to a computer environment, it had long become standard practice to use CAT tools to translate and verify pencil-and-paper tests.

So, when an e-assessment merely accommodates multilingual content but does not offer access to advanced translation technology, language service providers have to work without the tools of their trade. This often implies that achieving consistency becomes more work-intensive when producing or verifying translations in a computer-based testing platform than in word processing applications that allow the use of CAT tools.

Regardless of the authoring format, best practice in the localization industry is to separate text from layout. Translation editors do not handle layout and style (e.g., color, spacing, border thickness, or bullet formats); they handle only text. Untranslatable elements are kept separate from the translatable text, so that they cannot be altered during translation. These elements are merged with the translation at the end of the process.

Therefore, it is desirable to use adequate technology to extract translatable text from the source version. Layout markers are represented as locked tags that are easy to insert without being mishandled. Ideally, all elements that should not be translated or modified are either protected or hidden. Under these conditions, translators and verifiers can make full use of computer-aided translation tools to produce and edit text in their language. Once the text is translated and reviewed, the technology used for export can be used to seamlessly import the translation (the target version) back into the environment from which it was extracted, so that each text segment falls in the correct location.

At the same time, it is important that linguists and national reviewers can preview both the source version and their own work at any time, without file manipulation. It is necessary to preview the items in context because the translator needs to understand the task that respondents will be asked to perform. That is, the translator needs to read the entire stimulus and try to answer each question before translating it, and go

through the exercise again with the translated unit, to make sure that it functions the same way in the target language.

Content that will be used over different publication channels (webpages, mobile apps, print, etc.) should ideally be produced independently from the delivery modes, following a single-source/multi-channel publishing approach.

The Online Translation System (OTS) in the IEA eAssessment platform is a work in progress. In its initial form, it was a repository for multiple language versions of survey instruments rather than a tool to perform quality assurance routines and equivalence checks. For ePIRLS 2016, the OTS met the minimum function of allowing editing and replacement of the text with a translated version, but did not allow for the use of CATs. In addition, the OTS had basic functions for documenting adaptations and comments to segments, quick review, and resolution of translation verifier feedback, accessing previews of the international or national version of the instruments, and exporting PDF version of the national instruments.

After ePIRLS 2016, IEA made further improvements to the OTS in preparation for TIMSS 2019. A few of these improvements included the addition of a basic export/import function that would allow for the use of CATs, improved options for documenting comments and adaptations and sharing of translations. Even though the system has improved, there is room for more improvements and the interaction between platform engineers and translation technologists aims to close the gap and gradually build up functionality that will make it possible to harness the power of state-of-the-art translation technology.

6.11   Reviewing Results of Translation and Verification Processes

In our experience, quantifying the quality of a translation is difficult. The absence of comments may imply that the translation is near perfect, but it may also mean that the translation is so poor that it would make no sense to edit it. The item statistics have their own limitations, since a differential item functioning (DIF) can result from factors that are not related to translation, for example the curriculum and/or vocabulary used in textbooks, as well as their changes overtime. In addition, although item statistics may look acceptable, they do not indicate how the conveyed meaning was understood by respondents.

This can only be achieved by means of well-designed procedures that include multiple review loops aiding the focus, clarity, and fluency of the instruments. While test materials are skillfully crafted, they are not works of art, and there are usually multiple solutions to obtaining a satisfying target version. It is important that the experts involved in any stage appreciate the work already done, and build on it rather than change it according their particular language "taste." In most languages, grammar rules allow several possibilities, particularly as languages develop.

The international versions of IEA study instruments are finalized after multiple iterations that involve country and international expert views. Assuming that the experts from participating countries have voiced their concerns, the resulting materials contain concepts and words that are translatable to the target languages and cultures. Nevertheless, it remains a demanding process to create a target version that would be true to the international source, while fulfilling other important criteria, such as the appropriateness for the target grade and the country context (including curricula represented in the textbooks and taught in schools). While the translation verifiers comment on the submitted target language versions from the linguistic perspective, and the international quality control observers comment on the instruments and manuals used in a particular target language from a user perspective (including some qualitative input from the respondents as observed during the testing sessions or interviewed school coordinators and test administrators). In IEA studies, national research coordinators (NRCs) hold the responsibility of finalizing the instruments they will use when implementing an IEA study. Therefore, the major review loop starts and ends with the dedicated NRCs.

To begin with, NRCs review and comment on the international version. Once this source is finalized and released, NRCs engage and orchestrate the production of target versions by involving linguists and educators. Upon completion, NRCs submit the target versions for international verification. The translation verifiers review and comment on the submitted target versions. Afterwards, the verification feedback is returned to the NRCs for their review and finalization of the target version. Then NRCs document their experience in survey activity questionnaires that serve as a base for further development and improvement of procedures.

In IEA's trend studies, there is another challenge: trend instruments. These instruments are administered across study cycles to measure change. Ideally, trend instruments should not change unless there is an objective reason, and language experts are discouraged to make any cosmetic or preferential changes. In TIMSS 2003, trend items made up approximately 22% of the grade 4 items and 40% of the grade 8 items. In PIRLS 2006, 40% of the passages were trend. During the TIMSS and PIRLS 2011 cycle, the amount of trend items was increased to 60% and this has persisted to the TIMSS 2019 and PIRLS 2021 cycles. To ensure that the trend instruments are not changed from one cycle to the next, during international translation verification the translation verifiers compare the text of the trend instruments against the version used in the previous cycle. The translation verifiers document all differences found between the versions for review by the NRCs and ISC. A thorough documentation of any change made during the national instrument production is key to preventing unnecessary changes.

At the same time, if, for example, the currency has changed in the target country, or if a spelling reform has been implemented, it may be necessary to update the trend item/s. At this point, based on our research into trend item modification in TIMSS 2015 and the consequent DIF implication, we see a beneficial effect of the modest actions taken by NRCs in order to cope with the changes in the educational context (Ebb et al. 2017).

With the shift to the digital environment, we expect CBA development to provide a new base of information allowing more to be learned about the actions taken by NRCs in modifying the trend items (or any wording used in previous studies cycles), as well as the impact on response rate and achievement in relation to the particular items.

6.12 Procedure Chain and Timeline

IEA studies (indeed, all ILSAs) only have limited time available for their development and implementation (Fig. 6.8). The period available starts from the moment when the source version becomes available (released to participating countries) and must be finished at the moment when the target populations of students and related respondents (such as the parents of sampled students, teachers, and school principals) respond to the instruments.

The available time is shared by all key personnel engaged in the preparation of the national language(s) version(s) of assessment instruments. This includes NRCs, the translators and reviewers that work at a national level, IEA and ISC professionals, and expert translators performing the verification tasks. The reality resembles an assembly line, where any delay or misconduct in a single step affects the following step(s). This means that any time lost by a delayed submission of a language version can be covered by more support and swift accomplishment of the following step(s).
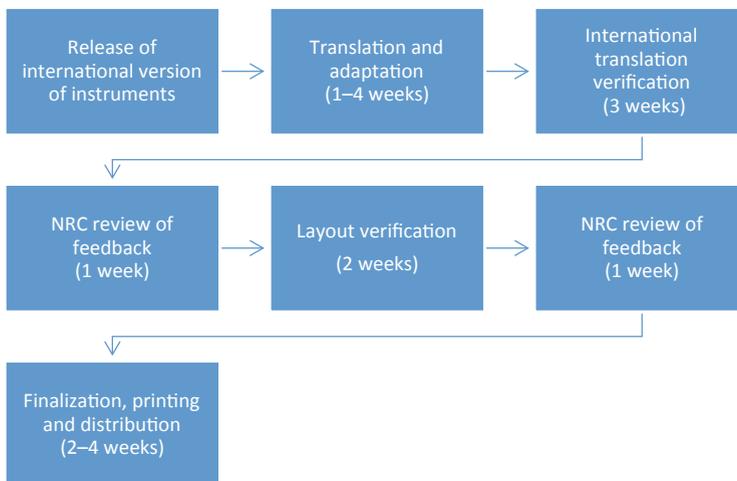

Fig. 6.8 Approximate instrument preparation timeline (based on use of printed booklets for TIMSS)
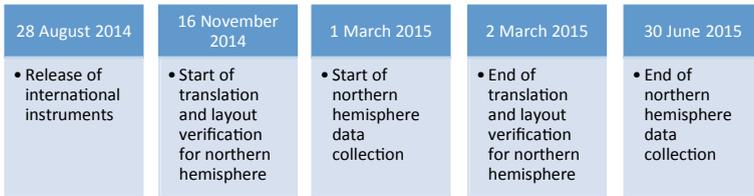
19 December 2013	6 January 2014	1 March 2014	15 March 2014	15 April 2014
•Release of international instruments	•Start of translation and layout verification	•Start of field test	•End of translation and layout verification	•End of field test

Fig. 6.9 TIMSS 2015 field test timeline

28 August 2014	16 November 2014	1 March 2015	2 March 2015	30 June 2015
• Release of international instruments	• Start of translation and layout verification for northern hemisphere	• Start of northern hemisphere data collection	• End of translation and layout verification for northern hemisphere	• End of northern hemisphere data collection

Fig. 6.10 TIMSS 2015 data collection timeline

For example, in the case of multiple national language versions of assessment instruments, these can be prepared in parallel at a country level step of translation and adaptation, and then reviewed for consistency (within and across languages) during the international translation verification. (Some examples of "real timelines" are shown in Figs. 6.9 and 6.10.)

6.13 Conclusions

Historically, participation in assessments that evaluate knowledge presupposes language proficiency, regardless of the evaluation domain. The Imperial examinations in Ancient China, a series of high-stakes tests accessible to all, were administered only in Mandarin. Acquiring advanced knowledge of the dominant language was a prerequisite toward success. A similar pattern could (and can) be observed in colonial and post-colonial states, where only the mastery of, for example, English, French, or Portuguese could open access to academic careers or high positions in the civil service: good results in admission exams taken in these colonial languages are/were regarded as a proof of competence.

In the case of IEA studies, it is of the utmost importance to diminish the impact of limitations in language proficiency that prevent participating students from demonstrating their content domain-related knowledge and their thinking skills, and hinder their engagement in responding to background and context questions expressing their own experience, attitudes, and opinions.

The review loop (see Figs. 6.1, 6.7, and 6.8), has been helpful in continuous development and improvement of the processes involved in producing the target

language instrument versions, enabling focus on the key aspects of the perceived and expected quality of the target versions of the international instruments.

The role of translation verifiers has evolved; verifiers can now concentrate on the linguistic aspects of the translations (with less need to look at layout and no need to judge the adequacy of adaptations), and, consequently, their comments have become more precise. It is also noteworthy that the numerous consecutive steps in elaborating a target version that strikes a good balance between faithfulness to the source and fluency in the target language are increasingly regarded as a collaborative effort rather than a system of checks and balances. The different players are less inclined to express their input in terms of errors, issues, or corrections, and more inclined to describe them in terms of semantic equivalence with the international source. While there is still room for improvement, it is clear that sophisticated translation, adaptation, review, and verification procedures have progressively generated a collective focus on maximizing comparability, as well as increasing reliability and validity, which has proven effective. The use of technology and development of tailored online translation systems can contribute here, by guiding and streamlining the workflow and by guarding consistency and eliciting inputs from the engaged professionals.

While engagement of particular stakeholders and professionals involved in the production of national instruments (translators, proofreaders, researchers, teachers, and administrators) is necessary, it is also important to facilitate their collaboration. In addition, cross-border collaboration has proved to be highly beneficial, since more eyes can see and eliminate more obvious mistakes, and more experience and viewpoints across different contexts can contribute to revealing incorrect assumptions and errors that would not be spotted otherwise.

Dealing with a matter as sensitive, abstract, and fluid as language benefits from a set of simple rules to deal with complexity and actions driven by common sense that all experts engaged can adopt and adhere to.

References

Ainley, J., Keeves, J., Lokan, J., Lietz, P., Masters, G., & Thomson, S. (2011). The contribution of IEA research studies to Australian education. In C. Papanastasiou, T. Plomp & E. C. Papanastasiou (Eds.), *IEA 1958–2008: 50 years of experiences and memories* (pp. 317–372). Amsterdam, the Netherlands: IEA. Retrieved from https://www.iea.nl/publications/iea-reference/iea-1958–2008.

Banks, K. (2006). A comprehensive framework for evaluating hypotheses about cultural bias in educational testing. *Applied Measurement in Education, 19*(2), 115–132.

Brislin, R. W. (1970). Back translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185–216.

Brislin, R. W. (1976). Comparative research methodology: Cross-cultural studies. *International Journal of Psychology, 11*(3), 215–229.

Dept, S., Ferrari, A., & Halleux, B. (2017). Translation and cultural appropriateness of survey material in large-scale assessments. In P. Lietz, J. Cresswell, K. Rust, & R. Adams (Eds.), *Implementation of large-scale education assessments* (pp. 168–191). Chichester, UK: Wiley.

Ebbs, D., & Friedman, T. (2015). Translation and verification of ICILS 2013 instruments. In J. Fraillon, W. Schulz, T. Friedman, J. Ainley & E. Gebhardt (Eds.), *ICILS 2013 technical report*

(pp. 55–66). Amsterdam, the Netherlands: IEA. Retrieved from https://www.iea.nl/publications/technical-reports/icils-2013-technical-report.

Ebbs, D., & Korsnakova, P. (2016). Translation and translation verification for TIMSS 2015. In M. O. Martin, I. V. S. Mullis & M. Hooper (Eds.), *Methods and procedures in TIMSS 2015* (pp. 7.1–7.16). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://www.iea.nl/publications/technical-reports/methods-and-procedures-timss-2015.

Ebbs, D., & Wry, E. (2017). Translation and layout verification for PIRLS 2016. In M. O. Martin, I. V. S. Mullis & M. Hooper (Eds.), *Methods and procedures in PIRLS 2016* (pp. 7.1–7.15). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://www.iea.nl/publications/study-reports/international-reports-iea-studies/methods-and-procedures-pirls-2016.

Ebbs, D., Desa, D., & Korsnakova, P. (2017). DIF study on the effects of item modifications based on the TIMSS trend items. Poster presentation. In *Program 7th IEA International Research Conference 2017, 28–30 June 2017, Faculty of Education, Charles University, Prague, Czech Republic* (p. 73). Amsterdam, the Netherlands: IEA. Retrieved from https://www.iea.nl/sites/default/files/2019-04/IRC%202017%20Program.pdf.

Eco, U. (2003). *Dire quasi la stessa cosa. Esperienze di traduzione.* Milan, Italy: Biompani. English translation: Eco, U. (2008). *Experiences in translation*, Toronto, Canada: University of Toronto Press.

Ferrer, A. T. (2011). Experiencing the world as an educational laboratory. In C. Papanastasiou, T. Plomp, & E. C. Papanastasiou (Eds.), *IEA 1958–2008: 50 years of experiences and memories.* Nicosia, Cyprus: Cultural Center of the Kykkos Monastery. https://www.iea.nl/publications/iea-reference/iea-1958-2008.

Foshay, A. W. (1962). The background and the procedures of the Twelve-Country Study. In A. W. Foshay, R. L. Thorndike, F. Hotyat, D. A. Pidgeon, & D. A. Walker (Eds.), *Educational achievements in thirteen-year-olds in twelve countries* (pp. 7–19). Hamburg, Germany: UNESCO Institute of Education, UNESCO. https://www.iea.nl/publications/publications/educational-achievements-thirteen-year-olds-twelve-countries.

Hambleton, R. K. (1992). *Translating achievement tests for use in cross-national studies* (Doc. Ref.: ICC454/NRC127). Report prepared for IEA, New York, NY, and NCES, Washington, DC, Laboratory of Psychometric and Evaluative Research Report No. 241, University of Massachusetts, School of Education, Amherst, MA. https://files.eric.ed.gov/fulltext/ED358128.pdf.

Hambleton, R. K. (2002). Adapting achievement tests into multiple languages for international assessments. In A. C. Porter & A. Gamoran (Eds.), *Methodological advances in cross national surveys of educational achievement* (pp. 58–79). Washington, DC: National Academy Press.

Hambleton, R. K., & Patsula, L. N. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology, 1*(1), 1–13. Retrieved from http://www.jattjournal.com/index.php/atp/article/view/48345.

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 33–56). Hoboken, NJ: Wiley.

Harkness, J. A. (2007). Improving the comparability of translations. In R. Jowell, C. Roberts, R. Fitzgerald, & G. Eva (Eds.), *Measuring attitudes cross-nationally: Lessons from the European Social Survey* (pp. 79–93). Los Angeles, CA: Sage.

Iliescu, D. (2017). *Adapting tests in linguistic and cultural situations.* New York, NY: Cambridge University Press.

Johansone, I., & Malak, B. (2008). Translation and national adaptations of the TIMSS 2007 assessment and questionnaires. In J. F. Olson, M. O. Martin, & I. V. S. Mullis (Eds.), *TIMSS 2007 technical report* (pp. 63–75). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://www.iea.nl/publications/technical-reports/timss-2007-technical-report.

Joldersma, K. J. (2004). *Cross-linguistic instrument comparability*. Unpublished manuscript, Michigan State University, East Lansing, MI. Retrieved from https://education.msu.edu/cepse/mqm/documents/KJ.pdf.

Law, N., Pelgrum, W., & Plomp, T. (Eds.). (2008). *Pedagogy and ICT use in schools around the world. Findings from the IEA SITES 2006 Study.* Hong Kong: CERC-Springer. Retrieved from https://www.iea.nl/publications/study-reports/international-reports-iea-studies/pedagogy-and-ict-use-schools-around.

Malak, B., & Trong, K. L. (2007) Translation and translation verification of the PIRLS reading assessment and questionnaires. In M. O. Martin, I. V. S. Mullis, & A. M. Kennedy (Eds.), *PIRLS 2006 technical report* (pp. 49–60). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://www.iea.nl/publications/technical-reports/pirls-2006-technical-report.

Malak, B., Yu, A., Schulz, W., & Friedman, T. (2011). Translation and national adaptations of ICCS 2009 instruments. In W. Schulz, J. Ainley, & J. Fraillon (Eds.), *ICCS 2009 technical report* (pp. 51–58). Amsterdam, the Netherlands: IEA. Retrieved from https://www.iea.nl/publications/technical-reports/iccs-2009-technical-report.

Malak-Minkiewicz, B., & Pelgrum, W. J. (2009). Translation, national adaptation, and verification. In R. Carstens & W. J. Pelgrum (Eds.), Second Information Technology in Education Study: SITES 2006 technical report (pp. 41–46). Amsterdam, the Netherlands: IEA. Retrieved from https://www.iea.nl/publications/technical-reports/second-information-technology-education-study-technical-report.

Martin, M. O., Rust, K., & Adams, R. J. (Eds.). (1999). *Technical standards for IEA studies.* Amsterdam, the Netherlands: IEA. Retrieved from https://www.iea.nl/publications/iea-reference/technical-standards-iea-studies.

Maxwell, B. (1996). Translation and cultural adaptation of the survey instruments. In M. O. Martin & D. L. Kelly (Eds.), *Third International Mathematics and Science Study (TIMSS) technical report. Volume I: Design and development* (pp. 8.1–8.10). Chestnut Hill, MA: Boston College. Retrieved from https://www.iea.nl/publications/publications/third-international-mathematics-and-science-study-technical-report-volume.

Noveanu, G. N., Friedman, T., & Köhler, H. (2018). Translation and national adaptations of ICCS instruments. In W. Schulz, R. Carstens, B. Losito, & J. Fraillon (Eds.), *ICCS 2016 technical report* (pp. 33–41). Amsterdam, the Netherlands: IEA. Retrieved from https://www.iea.nl/publications/technical-reports/iccs-2016-technical-report.

OECD. (2016). *PISA 2018 translation and adaptation guidelines.* National Project Managers' Meeting, Bangkok, Thailand, March 2013. First Meeting of the PISA 2018 National Project Managers, 14–18 March 2016, Prague, Czech Republic. Paris, France: OECD. Retrieved from https://www.oecd.org/pisa/pisaproducts/PISA-2018-TRANSLATION-AND-ADAPTATION-GUIDELINES.pdf.

Poortinga, Y. H. (1975). Limitations on international comparison of psychological data. *Nederlandse Tijdschrift voor de Psychologie, 30,* 23–29.

Poortinga, Y. H. (1995). Cultural bias in assessment: Historical and thematic issues. *European Journal of Psychological Assessment, 11*(3), 140–146.

Van de Vijver, F., He, J., & Kulikova, A. (2017). Response styles in cross-cultural perspective: TIMSS and PIRLS. In *Keynote presented at the 7th IEA International Research Conference 2017, Prague, Czech Republic [webpage].* Amsterdam, The Netherland: IEA. Retrieved from https://www.iea.nl/index.php/publications/keynotes/response-styles-cross-cultural-perspective.

Walker, M. (2017). Computer-based delivery of cognitive assessment and questionnaires. In P. Lietz, J. Cresswell, K. Rust, & R. Adams (Eds.), *Implementation of large-scale education assessments* (pp. 231–252). Chichester, UK: Wiley.

Yu, A., & Ebbs, D. (2012). Translation and translation verification. In M. O. Martin & I. V. S. Mullis (Eds.), *Methods and procedures in TIMSS and PIRLS 2011* (pp. 1–13). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from https://www.iea.nl/publications/technical-reports/methods-and-procedures-timss-and-pirls-2011.

Paulína Koršňáková has a background in natural science and mathematics education and holds a PhD in psychology. Prior to joining IEA, Dr. Koršňáková coordinated implementation of multiple IEA and OECD studies in Slovakia. The national adaptation and translation processes were part of the tasks she oversaw and carried on. Since 2010, Dr. Koršňáková worked on IEA ICILS 2013 and OECD TALIS 2013 studies at the international level, and she has overseen or directly managed translation verification processes for multiple other studies at the international level.

Steve Dept is one of the founders of cApStAn Linguistic Quality Control. He received his education in English, Dutch, French and German. He studied Classical Philology and comparative linguistics but is essentially an autodidact and a field practitioner. In 1998, he was commissioned to organize the translation review of PISA 2000 instruments in 24 languages. Within cApStAn, Steve has coordinated linguistic quality control (LQC) of PISA instruments across 8 PISA survey cycles. Other multilingual surveys or assessments for which LQC operations were supervised by Steve include TIMSS, PIRLS, PIAAC, UNESCO/LAMP, ICILS, and TALIS. Over the years, Steve's attention has shifted towards embedding linguistic quality assurance into the test development process with a view to reducing post hoc corrective action. His translatability assessment methodology has become an integral part of the adaptation design of major international large-scale assessments and multilingual surveys.

David Ebbs is a Senior Research Officer at IEA, overseeing the international translation verification process. He has used his considerable experience with IEA processes and procedures to help in the development of the IEA eAssessment system. He has a Master of Education in Educational Leadership and a Bachelor degree in Biology.