

## **Ensuring comparability of instruments in international assessments**

**Miss Bethan Burge**

Centre for International Comparisons, National Foundation for Educational Research

[b.burge@nfer.ac.uk](mailto:b.burge@nfer.ac.uk)

The Centre for International Comparisons at the NFER is responsible for administering TIMSS and PIRLS in England and Northern Ireland; PISA in England, Wales, Northern Ireland and Scotland (Scotland participate separately) and ESLC in England.

### **Abstract**

The success of large scale international surveys is dependent on gathering robust data that allows for reliable cross country comparisons. Ensuring students in all participating countries receive equivalent test and questionnaire items, regardless of the language in which they are being tested, is a vital element of every international survey. However, adaptations are sometimes necessary in order to make instruments accessible and appropriate for students in a particular participating country. Each international survey, therefore, imposes particular requirements on participating countries to ensure that such adaptations do not adversely impact on the comparability of the data gathered. This paper will explore:

- the procedures used internationally to check the comparability of both translations and adaptations
- examples of adaptations made to both test and questionnaire items in the UK (using released items from some of the international surveys, for example PISA, TIMSS and PIRLS)
- the tension between making an item suitable for use in one country and maintaining comparability with all other participating countries.

**Keywords:** international survey, translation, comparability

### **1 Introduction**

Large international surveys of educational achievement, such as the OECD PISA survey or the PIRLS and TIMSS studies organised by IEA, are an important tool for comparing student achievement and educational systems across the world. The success of these types of survey is dependent on gathering robust data that allows for reliable cross country comparisons. To help ensure this, a standardised approach needs to be adopted by each participating country. International surveys publish a Technical Standards document and a Survey Operations Manual outlining the standards that need to be met and procedures that need to be followed. These documents cover all aspects of survey administration including sampling schools and students, minimum participation rates, preparation of survey instruments, test administration and scoring of student responses. This paper will focus on one aspect of survey administration: the preparation of survey instruments.

Ensuring students in all participating countries receive equivalent test and questionnaire items, regardless of the language in which they are being tested, is a vital element of every international survey. Survey instruments, both tests and questionnaires, are usually produced in a limited number of languages by those overseeing the survey. These languages are known

as the source languages. For example, for the OECD Programme in International Student Assessment (PISA) tests and questionnaires are produced in English and French, and there are English and Arabic versions of the Trends in Maths and Science Study (TIMSS) instruments. This means that the majority of participating countries need to translate the survey instruments into the official language or languages spoken by students. Even those countries that do not need to translate the instruments may need to make changes to them so that they reflect terminology used in the country (these types of changes will be referred to as ‘adaptations’ in this paper). The challenge for all international surveys is to ensure that translations and adaptations do not result in the instruments being interpreted in different ways by participants in different countries, thus impacting on the comparability of data. Each international survey, therefore, imposes particular requirements on participating countries to ensure that such impact is limited.

## **2 Translation of Survey Instruments**

Good quality translation is crucial to the success of large scale international surveys. It is important that the process of translation does not introduce bias into the test materials which may lead to a lack of comparability. On first sight the process of translation in an international survey seems to be straightforward. However, this is not always the case. When test materials are written it is not always possible to consider how well they will translate into the myriad of languages required by participating countries. For example, in TIMSS 2011 there are 66 participating countries and the materials have been translated into 46 languages. Occasionally there are words or terms that are commonplace in one language which simply do not exist in another; in these instances the translator will need to find a suitable alternative that will not impact on the way the student interprets the item. The translator will need to consider whether the alternative word will make the item more difficult to understand or will cue the correct answer, making the item easier in that country compared with others.

In order to ensure consistency in the translation process and in the quality of the translation in participating countries, the organisations that run the international surveys have strict procedures for countries to follow as they embark on the process of translation. These procedures have varying degrees of formality.

The IEA, the organisers of TIMSS and PIRLS, issues broad guidelines to participating countries to aid the consistency and the quality of the translations, Figure 1 shows the issues that need to be considered when undertaking a translation. In addition to this, criteria are provided for countries to use as a checklist when appointing translators and reviewers. The IEA requires that translations, of the TIMSS and PIRLS materials, are read by an internal reviewer to ensure that they are at a suitable level for the target population, and are of the highest quality. It is also recommended that if a country is translating the materials into more than one language a translator familiar with the various languages is involved in cross checking to ensure the translations are equivalent.

### Figure 1: TIMSS Guidelines for translation

In translating and adapting the TIMSS 2011 achievement booklets, particular attention should be paid to the following issues:

- The translated texts should have the same register (language level and degree of formality) as the source texts.
- The translated texts should have correct grammar and usage (e.g., subject/verb agreement, prepositions, verb tenses, etc.).
- The translated texts should not clarify or take out text from the source text and should not add more information.
- 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.
- Spelling, punctuation, and capitalisation in the target texts should be appropriate for the target language and the country's national context.

It is also extremely important to keep in mind that these translations are intended for fourth-grade and/or eighth-grade students and should reflect the language level of this audience.

The OECD PISA survey and the European Commission research project the European Survey on Language Competences (ESLC) also have criteria for translators working on the survey instruments (see figure 2 for the criteria for PISA 2012). In addition, they require participating countries to check translations in a more formal way. In PISA, the test materials and questionnaires are translated by two independent translators. A third person takes the role of 'reconciler' to compare the two translations. The role of the 'reconciler' is particularly important as this person is responsible for 'merging' the two independent translations so that the final national version is as close as possible to the source version. The reconciler needs to ensure that any translation errors have been corrected and that the wording is as fluent as possible. In PISA, the Consortium develops two parallel source versions of the test and questionnaire materials, one in English and the other in French. Countries are recommended to use the English source version for one of the translations and the French source version for the other.

### Figure 2: PISA criteria for appointing translator(s)

The translators responsible for this work must have:

- a perfect command of both the source language and the target language
- a solid command of the domains covered by the test, in particular for the domains of Mathematics and Problem Solving, and (for new countries) Reading and Science. It may therefore be necessary to turn to pairs of different translators for the translation of each of the different domains
- an in-depth knowledge of the school system of the country (in particular, for the school and student questionnaires). The knowledge of the characteristics of the school systems of other countries (particularly those where the source language is used) can also be a valuable asset.

In the ESLC the only survey instruments that require translation are the questionnaires. As well as reconciling two independent translations, countries participating in the ESLC have to conduct a ‘back translation’ of the questionnaires. This means that the reconciled version is translated back to the source language, which in this case is English. The back translation is then compared to the source document to identify possible errors before the translations is finalised. In order to make the task of translation easier, the Consortium running the ESLC provides countries with a web-based translation platform called WebTrans2. This platform documents and tracks the whole translation process and also allows the Consortium to review and quality control the documents submitted by participating countries.

### **3 Adaptation of Survey Instruments**

As discussed above, in order to make valid comparisons in international surveys there needs to be an equivalence and consistency between the survey instruments used in different countries. It is, therefore, important that only limited and necessary changes are made – such as translations into other languages as described above. However, sometimes a country may feel it is necessary to make changes for cultural reasons or to conform to local usage. These changes are referred to in international surveys as ‘adaptations’. Sometimes there is a tension between making adaptations that enable students to access the survey instruments more easily and limiting the number of changes in order to maintain comparability. The PISA guidelines for adaptations recommend that any proposed national adaptations are discussed with relevant national committees or other experts (e.g. curriculum and assessment groups). Procedures such as these are likely to ensure that the proposed adaptations are both necessary and appropriate.

In the UK, the main changes made to international survey instruments (apart from the Welsh translation of the PISA materials and Gaelic translation of the TIMSS Grade 4 materials) are national adaptations. These adaptations are made in both the tests and the questionnaires for a variety of reasons. Considerable time is spent making sure that these adaptations are made consistently in all the survey instruments and that they are documented for the Consortium.

Some of the types of adaptations made to the England/ UK national versions are outlined below, with examples from TIMSS and PISA instruments.

#### **3.1 Unfamiliar vocabulary**

The primary purpose of these adaptations is to ensure that students do not have to grapple with unfamiliar text during the course of an assessment. Changing the spelling of words to nationally recognised spellings is one form of this type of adaptation. It may also be necessary to adapt names (such as those of people and places). In these cases it is important to ensure that the replacement words are equally familiar to all students and are in a similar format (i.e. length and complexity) as those used in the source versions. There may also be words, phrases and sentences that are familiar to students in some English speaking countries but are not familiar to students in the UK. A few examples of adaptations that have been made to test and questionnaire items from TIMSS are given below.

- *favorite* changed to *favourite* (this is the correct spelling in the UK)
- *Mano* changed to *Mike* (name changed to a more familiar UK one)

- *chalkboard* changed to *whiteboard* (this would be a more familiar word for students in the UK)
- *trapezoid* changed to *trapezium* (this would be a more familiar word for students in the UK)
- *'styrofoam'* changed to *'polystyrene'* (this is a more familiar word for students in UK)
- *'measure of angle DAC'* changed to *'size of angle DAC'* (this would be a more familiar term for students in the UK)
- *'raise your hand'* changed to *'put your hand up'* (this is a more familiar instruction for students in the UK)

### 3.2 National conventions

Countries often have national conventions for dates, times, measurements and punctuation. If these types of adaptations are made it is vital that this is done consistently in all questions; this is particularly important in the case of multiple choice questions where the same changes must be made to the question and the answer options. Table 1 taken from the TIMSS 2011 Survey Operation Manual (2010) shows the changes that might be made to test items to reflect these national conventions.

**Table 1**

Type of change	Specific change from	Specific change to
Punctuation or notation	decimal point	decimal comma
	place value comma	space
	quotation marks	dashes
Measurement units	kilometers (kilometres)	miles
	meters (metres)	feet
	centimeters (centimetres)	inches
	Celsius	Fahrenheit
Expressions of date or time	April 3	3 April
	2:00 p.m.	14:00

### 3.3 Terminology specific to the educational system

This is a primarily an issue that needs to be considered when preparing the questionnaire(s) for a specific national audience. Whilst it is likely that teachers and headteachers will understand the international terminology, it is best practice to make questionnaires as clear and unambiguous as possible. As a result, certain adaptations may be necessary. A few examples of the adaptations that have been made to international questionnaires are given below.

- *grade* changed to *year group*
- *high school* changed to *secondary school*
- *quiz* changed to *test*
- *principal* changed to *headteacher*

### 3.4 Age-appropriate wording

This type of adaptation involves changing the original wording of test or questionnaire items to make them easier for students to read. By using, for example, words that students in the target age range are more likely to come across in their everyday lives. It is important to recognise that making a test item more accessible in this way should not impact on the item difficulty. As well as removing words or terms that are not used by students, the construction of sentences may be simplified to make them easier for students to understand. However, it is sometimes a challenge to reword an item in a way that makes it more accessible and also retain the original meaning. Below are some successful adaptations that have been made to the PISA questionnaire for UK usage.

- *Enrol* changed to *join* ('join' would be a more familiar term for 15 year old students)
- *I reason that* changed to *I think that* (using the word 'think' makes the item easier for students to interpret and reflects language usage in the UK)
- *method of solution* changed to *method for finding the answer* ('solution' is not a word that students of this age would routinely use and therefore the adaptation helps students to interpret the question but would not make it easier for them to arrive at the answer)

## 4 Procedures for checking translations

The guidelines and procedures outlined above are designed to ensure that each country produces high quality and consistent translations and adaptations of the source materials. In each international survey the organisations running the surveys also have a part to play in the quality assurance of the national versions of the materials. They are responsible for reviewing the translations and adaptations. This process is known as verification. In order for the verification to be done systematically, each country, in addition to supplying the translated versions of the materials, must document any adaptations they have made to the source materials and explain why such changes are necessary. It is usual for those running the surveys to provide forms that can be used for this purpose.

For TIMSS and PIRLS these changes are documented on the National Adaptation Form (NAF). Once this form has been submitted to IEA, the documents are reviewed. This process requires the 'verifier' to document the comparability between a country's translated/adapted material and the international version. The verifier is responsible for documenting on the NAF any deviations from the source version and assigning a 'severity code' to each deviation. Figure 3 outlines the codes used by verifiers to give a severity rating to any differences between the translated materials and the source versions. As part of this process, verifiers may suggest alternative translations/ adaptations that they feel would improve the comparability. In TIMSS and PIRLS all national versions go through this verification process even if the materials are translated into a minority language and will only be seen by a relatively small number of students. For example, the TIMSS 2011 source materials were translated into Gaelic for use in Northern Ireland; although only a small number of Irish medium schools were sampled for the survey the Gaelic version still went through the full verification process.

**Figure 3: Severity codes used in TIMSS to rate the differences between translations and source versions**

The following severity codes, ranging from Code 1 (major change or error) to Code 4 (acceptable change), will be used:

**Code 1:** Major change or error. Examples include the incorrect order of choices in a multiple-choice item, omission of an item, incorrect translation resulting in the answer being indicated by the item, an incorrect translation that changes the meaning or difficulty of the item, and the incorrect order of the items.

**Code 2:** Minor change or error. Examples include spelling errors that do not affect comprehension.

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

**Code 4:** Acceptable change. The change was acceptable and appropriate.

In PISA the national versions of the survey instruments are only verified if they will be used in schools attended by more than 10% of the national target population. The PISA verification process is very similar to that used in TIMSS and PIRLS. Countries document the translations and adaptations that have been made on the Test Adaptation Spreadsheet (TAS) and the Question Adaptation Spreadsheet (QAS). These changes are verified by the PISA Consortium and the verifier's comments are fed back using the same forms. Where a national version will be seen by less than 10% of the national target population no international verification is required. So, for example, in the UK, the Welsh national versions of the PISA tests are not sent to the consortium for verification. In spite of this, it is still important to know whether the translation is clear and whether further adaptations may be required even though the survey instruments are being used with a very small number of students. To check this, the Welsh national version of the instruments are trialled in a few schools in order to identify any parts of the translation that were difficult for students to understand and also to establish if any changes are needed in the translation. This information is invaluable and helps to ensure that the data gathered using the translated instruments is as reliable as possible.

Once a country has received the verification feedback, the person responsible for the translation/adaptation will need to review the verifier's comments and if necessary make revisions to the national test versions. Disagreements are possible at this stage: the verifier and the translator may have different opinions about which adaptations are essential to aid students' understanding while maintaining international comparability. At this stage it may be necessary to consider whether an adaptation is actually essential or whether there are any alternatives that will not jeopardise the findings.

#### **4.1 Comparability over time**

In addition to maintaining comparability between countries, international surveys also need to demonstrate comparability over time. It is, therefore, crucial that adaptations made to items that are used in consecutive surveys, known as 'trend items', are consistent. When the organisations running the survey verify the national versions of the survey instruments they will also check that the translations/adaptations are identical to those used in previous surveys. It is unlikely that a verifier would suggest changes to a trend item unless a major

error had been made in the past – something that compromised students’ ability to understand an item - or a spelling error is detected. In spite of this, it is possible that a country may want to make changes to a trend item. For example, an item may have been adapted in a certain way the first time it was used in a survey but on reflection the adaptation may seem incorrect or unnecessary. In these cases an alternative adaptation might be suggested. The verifier would need to consider whether the change would impact on the comparability of findings over time. Figure 4 is an example extract from a Test Adaptation Spreadsheet (the form used to document adaptations to the source materials). This is not a genuine TAS extract (and does not refer to a real PISA question). However, this example serves to illustrate the type of adaptation request that may be made. This extract documents an adaptation request to a trend item and highlights that the adaptation made to a word in 2003 was no longer appropriate in the 2006 survey. In this case, the verifier agreed that this minor change would not impact on the way students would interpret the item and therefore would not impact on the comparability of data coming from the item.

**Figure 4: Extract from PISA 2006 Test Adaptation Spreadsheet**

<p><b>Requested adaptation</b></p> <p><i>In 2003 Main Survey 'Economics' was changed to lower case ('economics'). We did not make this change for the Field Trial as our students are now familiar with seeing 'Economics' capitalised. We would prefer to leave 'Economics' capitalised as in the international version. Is this ok or will it raise comparability problems between 2003/2006? (Note there was inconsistency in this for question between 2000 and 2003, which has presumably not affected comparisons over time).</i></p> <p><b>Consortium response</b></p> <p>Leave as is (capitalised).</p>
---

## 5 Discussion

The translation / adaptation guidelines produced by those running large international surveys standardise the process of translation and adaptation within countries. This standardised approach helps to ensure the comparability of survey instruments. For example, providing criteria for translators means that the quality and accuracy of translations are considered by all countries at the very start of the process. It is, however, also important that countries take responsibility for ensuring that translators are fully briefed about the importance of maintaining equivalence between the national and source versions. The use of independent translators and reconcilers or reviewers during the translation process should mean that the final translations are both accurate and appropriate for the students and teachers who will use the instruments.

The process of making national adaptations is equally important in ensuring that the final national materials reflect the cultural context and are appropriate for the respondents. As noted above, there are a number of different types of national adaptations that may need to be made to survey instruments. It is a balancing act for the person responsible for the adaptations to make sure that only a limited amount of adaptation takes place, in order to maintain comparability with other countries, whilst making all necessary changes. The tension here lies between people’s views of what is a ‘necessary adaptation’ that will aid understanding

and what constitutes re-writing the instruments. It is, therefore, important that countries review proposed national adaptations with subject, curriculum and pedagogical experts in order to get the right balance.

In summary, great care is needed when translating and making adaptations to international survey materials. It is important to balance the needs of pupils within the country with the need for equivalence in the materials used in all participating countries. It is vital that the consortiums organising the surveys are aware of the adaptations that are made and that these are documented. This documentation provides an overview of the changes made in all participating countries and helps to ensure all national instruments are comparable and are as close as possible to the international source versions.