

Maxwell, B. (1996) "Translation and Cultural Adaptation of the Survey Instruments" in M.O. Martin and D.L. Kelly (eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.

8. TRANSLATION AND CULTURAL ADAPTATION OF THE SURVEY INSTRUMENTS	8-1
<i>Beverley Maxwell</i>	
8.1 OVERVIEW.....	8-1
8.2 TRANSLATING THE TIMSS ACHIEVEMENT TESTS.....	8-2
8.3 TRANSLATION PROCEDURES AT THE NATIONAL CENTERS.....	8-3
8.4 VERIFYING THE TRANSLATIONS.....	8-6

8. Translation and Cultural Adaptation of the Survey Instruments

Beverley Maxwell

8.1 OVERVIEW

Although the international set of TIMSS instruments was prepared and distributed in English, the 45 countries participating in TIMSS represent 31 different languages, and the achievement booklets had to be translated into each of those languages. Because the inherent risk of error or inequity in the translations was obvious, translation validity was an issue from the very beginnings of TIMSS. Detailed guidelines for producing translations, the enlistment of qualified translators, and careful procedures for checking translations have resulted in high-quality instruments for all countries.

The first stage of ensuring high-quality of translations was to identify the most appropriate translation procedures for the study. In 1992, TIMSS commissioned Ronald K. Hambleton (University of Massachusetts, Amherst) to write a paper on translation procedures for international achievement instruments. That paper, *Translating Achievement Tests for Use in Cross-National Studies* (Hambleton, 1992), was the basis for the translation guidelines provided to the National Research Coordinators (NRCs) for the 1993 item pilot, 1994 field trial, and main survey in 1994-1995. In his paper, Hambleton acknowledges that the process in a study such as TIMSS is not merely test translation, but test adaptation:

Some researchers prefer the term test adaptation to test translation because the former term seems to more accurately reflect the process that often takes place: Producing an equivalent test in a second language or culture often involves not only a translation that preserves the original test meaning, but also additional changes such as those affecting item format and testing procedures. [Such

changes] may be necessary to insure the equivalence of the versions of the test in multiple languages or cultures (pp. 3-4).

The TIMSS procedures addressed all aspects of test adaptation and applied to all participating national centers. Many of the adaptations that were necessary to produce equitable items, such as changing proper nouns and units of measure, were important to all countries, regardless of the ultimate language of the test. Throughout the study, the English-speaking countries followed the same translation guidelines and verification procedures as countries producing instruments in other languages. In all of the operations manuals, the section entitled “Guidelines for Translation and Cultural Adaptation,” specifies that the guidelines include English-speaking countries.

Hambleton emphasizes the need for care in translation and for ensuring equivalence:

Unless the translation work is done well, and evidence is compiled to establish, in some sense, the equivalence of the two versions of the test, questions about the validity of the translated tests will arise. Also, the validity of comparisons among countries where different versions of the test have been administered will be in doubt until questions about the equivalence of the versions are resolved (p. 3).

Each national center was responsible for producing the instruments used in that country. The national centers had access to the best translators for the task and were familiar with the resources available within their educational systems. They could select personnel whose first language was the language into which the test was to be translated, and who had a good knowledge of the subject matter and age-appropriate language. These translations were then reviewed centrally through the International Coordinating Center (ICC) by independent certified translators who were not involved in any other TIMSS activities. Their evaluations were provided to the respective national centers to allow them to make any necessary corrections before the booklets were administered.

8.2 TRANSLATING THE TIMSS ACHIEVEMENT TESTS

Depending on their participation in pilot and field trial activities, national centers produced up to three translations of test booklets:

1. Booklets for the Population 1 and 2 item pilot in 1993
2. Booklets for the Population 1, 2, and 3 field trial in 1994
3. Booklets for the Population 1, 2, and 3 main survey in 1994-95.

Each translation was verified item by item by an independent agency. (The verification procedure is described in detail in a section below.) In the 1993 item pilot and the 1994

field trial, verification followed the administration of the instruments, because there was insufficient time between translation and administration.¹

In the main survey, verification preceded administration. Once the booklets for a country were translated, a master copy was sent to the ICC. All items were checked by certified translators, as were the Population 2 performance assessment student worksheets.² For each country submission, the items and booklets were reviewed, and the results (including suggested changes in the translations) were returned to the country. Errors or deviations discovered during translation verification were, in most cases, communicated to NRCs in time for them to make corrections before test administration. This was especially effective in minimizing errors, and was certainly an important element in ensuring and confirming good translations. In cases where the national center submitted the student test booklets for translation verification only after they were administered, the report could still be used to help resolve data anomalies, and as a *post hoc* confirmation of the quality of the booklets.

As a final confirmation, each country's Quality Control Monitor (an independent reviewer of the implementation of the TIMSS procedures, described in Chapter 11) received a copy of the translation verification report and checked that the appropriate changes had been made.

8.3 TRANSLATION PROCEDURES AT THE NATIONAL CENTERS

The recommended translation procedures applied to the item pilot, field trial, and main survey, and called for multiple-forward translations of all the test items. NRCs were asked to have the instruments translated by more than one translator and to compare the results. The expectation was that a pair of independent translations would be the same for most items. For any item where the two translations differed, the differences would be discussed, and the best translation of the item selected for the test instrument. The guidelines identified five characteristics of an appropriate translator:

- A good knowledge of English
- An excellent knowledge of the target language
- Experience in both languages and cultures
- Experience with students of the target populations
- Skills in test development

¹ In the 1993 item pilot, the translators reviewed the test items after the tests had been administered, the results analyzed, and the desirable items identified. Only the "surviving" items were reviewed. Each participating NRC received a report on the quality of the translation of each of those items. Many of those items would be included in the 1994 field trial, so the NRC had an opportunity to improve or correct the translation before it was administered again. In the 1994 field trial, reviewing all items for all countries would have been too costly, and a compromise position was taken. If a country's item pilot translation had been good, only a 25% sample of items was reviewed, to check that the same quality existed in the field trial instruments. If the item pilot translation had not been good, or if the country had not submitted item pilot instruments, all of the items were reviewed. The results of these reviews contributed to the quality of the main survey instruments.

² Population 1 tasks were not verified due to limited resources. In fact, most items were very similar to those used for Population 2, and the results for the Population 2 review could be applied to Population 1 tasks.

Although multiple translations are an excellent safeguard for validity, time and budget constraints precluded these in some countries. However, by choosing competent translators, all countries were able to produce good translations. Subsequent detailed independent evaluations of the translations confirmed the quality of the instruments. The guidelines for the translators emphasized the following operations:

- Identifying and minimizing cultural differences
- Finding equivalent words and phrases
- Making sure the reading level is the same in the target language as in the original English version
- Making sure the essential meaning does not change
- Making sure the difficulty level of achievement items does not change
- Being aware of changes in layout due to translation.

In the item development stage of TIMSS, extensive review and selection went on to ensure that the items did not introduce a cultural bias to the tests. For example, an item that requires knowledge of the rules of baseball is not acceptable in an international test. Notwithstanding this scrutiny, many changes in the questions were required from country to country because of differences in culture. Thus, measurement units, seasons, names of people, places, animals, plants, currencies, and the like were adapted to be equally familiar to all students, insofar as this is possible.

Concepts and conventions that were not common to all cultures and were not related to the substance of the question were changed. For example, a graph that showed winter clothing sales by month of the year, with increasing sales in November-December and declining sales in July-August, was sensible in the Northern Hemisphere. In the Southern Hemisphere, such a graph makes less sense. To adapt the question, one country changed “coats and sweaters” to “shorts and tee-shirts,” whereas another changed the month names on the graph. Both were good adaptations for a Southern Hemisphere climate. The meaning of the item, rather than the exact wording, was translated.

Changes in proper nouns were a necessary adaptation for many countries. These included the names of people, cities, and official titles. Changes in common nouns were also necessary, to ensure that children were equally familiar with the vocabulary and topic of the question. For example, in a question about vertebrates and invertebrates, a land-locked country replaced “clam” with “snail,” and a seaside country replaced “crayfish” with “shrimp.” In another item, where students were asked to interpret a diagram of a food web, various small animals were selected to replace the skunk shown in the English-language version.

Questions involving money were adapted in three ways. In the source instruments, the currency used was dollars. Where it was sensible to do so, the notation was directly translated into the local currency, without changing the value or the context of the item (for

example, a \$20 train ticket could be changed to a £20 ticket). In some cases, simply changing the currency resulted in inappropriate values, so the item being purchased was also changed. Finally, for some questions, some translators retained the dollar currency because there was no easy adaptation. This was limited to a few countries where students were generally familiar with the dollar as a foreign currency and where it would not affect the difficulty level of the question.

Most of the adaptations in mathematics and science notation and units of measure are generally accepted. The most straightforward adaptations were in the form of decimal notation, place value notation, and time (use of the colon or period, and use of the 12- or 24-hour clock). The test consistently used metric units of measure; however, if the context allowed, imperial measure could be substituted. This was acceptable only when the values did not also need to be changed. For example, it is acceptable to change “six bags of flour, each weighing 10 kg,” to “six bags of flour, each weighing 10 lb,” but not to “six bags of flour, each weighing 22 lb.”

In addition to their own experience and good sense, translators had two resources to inform their decisions. First, the guidelines for translators were explicit about the objectives of cultural adaptations, and provided examples of good and poor changes. Table 8.1 displays the actual examples of appropriate adaptations provided to translators. The second resource was the team of subject-matter and evaluation specialists coordinated by the International Study Center. Translators could refer uncertainties about translation to the International Study Center or the ICC. In such cases, the query was directed to the appropriate person, and a recommendation returned quickly.

Table 8.1 Examples of Acceptable Cultural Adaptations

Class of Change	Specific Change from	Specific Change to
Punctuation or Notation	decimal point place value comma	decimal comma space
Units	centimeters liters ml	inches quarts mL
Proper nouns	Ottawa Mary	Oslo Maria
Common nouns	robin elevator	kiwi lift
Spelling	center	centre
Verbs (not related to content)	skiing	sailing
Usage	Bunsen burner	hot plate

The layout of some questions also needed to be adapted in some countries. If the translated text differed in length from the English original, additional lines of text were inserted without changing the pagination of the items. This was possible because the English layout left substantial space between items. If additional space was required for an

item, NRCs could use the white space between items on that page or reduce the spacing between the lines of text.

In some items, especially questions in the negative, a word was emphasized by using all capital letters. For example, in “Which animal is NOT a mammal?”, the negative is emphasized by capitalizing the word “not.” In languages where this format would not be possible or meaningful to children, it was recommended that the word be emphasized in some other way. This was usually boldface, underlining, or italics. The objective was to ensure that the student not overlook important words or the negative form of a question.

The translators maintained records of each adaptation made in translating the achievement instruments. Before completing the translation, this information was forwarded to the International Study Center, which obtained a ruling on the appropriateness of each adaptation from a subject-matter specialist. Upon completion of the instruments, NRCs were instructed to compare them item by item with the English originals. The guidelines directed NRCs to check that the following conditions were met:

- All items were present in the correct order
- There were no misplaced graphics, incomplete texts, or incorrect options
- The translations were inserted precisely (correct spelling, no missing words)
- All variable names were correct and in order
- The graphics were printed correctly, especially those containing shading that was significant to the solution of the item.

After this comparison, the instruments were submitted for independent translation verification.

8.4 VERIFYING THE TRANSLATIONS

There are four types of procedures for verifying translations: multiple-forward translation, back-translation, translation review by bilingual judges, and statistical review. In TIMSS, at least two and usually three of these procedures were used.

- **Multiple-forward translation.** This form of verification was carried out in the individual national centers. As mentioned above, NRCs were asked to obtain multiple independent translations of the instruments, followed by an item by item comparison.
- **Back-translation.** 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. This procedure was not used in TIMSS for a number of reasons. First, it would have exceeded the resources of most national centers. Additionally, the procedure can obscure significant flaws in the translated instrument.³ Finally, “the back translator [may be] able to do a good translation even though the original translation

³ For example, in the question, “What does a carnivore eat?”, the word “carnivore” would read “meat-eater” in many translations, making the questions very much easier. But if in the back-translation “meat-eater” was translated back to “carnivore,” one would not know about the flawed original translation.

was poorly done and resulted in a non-equivalent target language version of the test ” (Hambleton, 1992, p. 14).

- Translation review by bilingual judges. This may be considered as a variation of back-translation; however, unlike that procedure, this focuses on both the target and the source language. This procedure was used in all stages of TIMSS. It was favored because in addition to checking the accuracy of the translations per se, it allowed checking cultural adaptations and comparison of the levels of reading difficulty.
- Statistical review. In both the item pilot and the field trial, NRCs were provided with item statistics for their sampled populations. Anomalies in the results were flagged, so that NRCs could check suspect items for translating or printing errors.

The ICC enlisted a professional translation agency in Vancouver, Canada, to select the personnel for verifying the translated TIMSS achievement instruments. The criteria for translators included:

- Formal credentials as a translator into the target language
- First-language experience in the target language
- Excellent knowledge of English
- Experience living and working in an English-language environment
- Familiarity with the culture associated with the target language.

Because of Canada’s multicultural history, it was usually possible to engage translators who had immigrated to Canada from countries involved in TIMSS and who had experience in both cultures. Most of the translators lived in the Greater Vancouver area; the rest were located in other Canadian cities.

For verification of the main survey translation, each of these “verifiers” was provided with a package containing the following materials.

- A two-page introduction summarizing the TIMSS project, the instruments, and the translation goals, as background information
- A set of the translated instruments (as either assembled booklets or item clusters)
- A set of the international versions of the instruments
- A copy of “Guidelines for Translation and Cultural Adaptation” (an excerpt from the *Survey Operations Manual* (TIMSS, 1994a, 1994b), containing the original instructions for translating the instruments; this allowed the verifier to know what instructions were given to the original translator)
- Instructions for verifying the general layout (checking that the message to students appeared at the beginning of the book, the questions appeared in the correct order, the illustrations were in the right place, all labels were translated, and page breaks were the same as in the international versions)
- Instructions for verifying the message to students (a list of points that the message must have clearly communicated)

- Instructions for item-by-item checking (including the procedures for coding observations to indicate the type and severity of the error)
- An example of a verified translation, including an annotated verifier's report.

After checking the general layout and the message to students, the verifiers compared each item with its international version. If the translated item was judged equivalent to the international version, no observation was made in the verifier's report. If it differed in any way from the original, an observation was made, composed of a *severity* code, a *type* code, and an explanation.

The severity code ranged from 1 (serious error) to 4 (acceptable adaptation).

- 1 – Major Change or Error: This could affect the results. Examples include incorrect ordering of choices in a multiple-choice item; omission of a graph that is essential to a solution; an incorrect translation of text such that the answer is indicated by the question.
- 2 – Minor Change or Error: This should be corrected if possible, but will not affect the results. Examples include spelling errors that do not affect comprehension; misalignment of margins or tabs; incorrect font or font size.
- 3 – Suggestions for Alternative: The translation may be adequate, but the verifier suggests a different wording for the item. The NRC would be asked to review such suggestions and decide whether to make the suggested changes.
- 4 – Acceptable Changes: The verifier identifies changes that are acceptable and appropriate adaptations. This is done to provide information and requires no action from the NRC. An example is where a reference to winter is changed from January to July for the Southern Hemisphere.

Type codes allowed the verifier to use a “shorthand” for indicating the type of adaptation in addition to its severity. Codes A through J were used for text, and K through N for graphics and layout. For example, an appropriate change in vocabulary (coyote to dingo) would be coded as 4-C. An inappropriate change (gravity to weight) would be coded as 1-C. In cases where the verifier was unsure about the coding, a question mark was used in place of a code, and the uncertainty was elaborated in the explanation.

The type codes are:

- A. Spelling
- B. Grammar
- C. Vocabulary
- D. Incorrect number or value
- E. Error in equation or numeric notation
- F. Missing or additional text

-
- G. Change in meaning
 - H. Change in level of reading difficulty
 - I. Tabs, alignment, or text layout
 - J. Other problem with the text
 - K. Labels are missing
 - L. Wrong picture or picture is missing
 - M. Picture has been modified
 - N. Labels have been modified.

The verifiers' reports consisted of an overall statement of the quality of the translation, followed by a list of observations associated with individual items. The reports were sent to the ICC, where they were reviewed and subsequently forwarded to the International Study Center and the appropriate NRC. It became apparent that two features of the coding greatly facilitated the review of the reports. First, for a report with numerous observations, the frequency of each severity code provided a quick indication of which and how many items required immediate attention. This was useful for the NRC, and for the quality control monitor responsible for checking that the report recommendations had been followed. Second, in some cases the observation consisted of an alternative translation without explanation. For English-speaking reviewers at the ICC and International Study Center, the severity and type codes were necessary for understanding the nature of the observation.

Finally, the translation verification reports contributed to understanding the initial analyses of the achievement data. The IEA Data Processing Center in Hamburg, Germany, received each country's data files following administration and data entry. As those files were cleaned, several routines were performed to check for anomalous data. During this process, the translation verification reports were consulted for possible explanations for the anomalies. And, as the International Study Center staff reviewed the item statistics the translation reports were reviewed.

The procedures for verifying translations in the TIMSS study were highly effective. In most cases they confirmed that the national centers had produced high quality translations; in other cases they alerted the centers to flaws in translations in time to make changes. As a serendipitous outcome of the procedures, the careful documentation of acceptable and unacceptable adaptations will be a useful resource for researchers developing guidelines and procedures in subsequent studies.

REFERENCES

Hambleton, R. (1992). *Translating Achievement Tests for Use in Cross-National Studies* (Doc. Ref.: ICC454/NRC127). Paper prepared for the Third International Mathematics and Science Study (TIMSS).

Third International Mathematics and Science Study (TIMSS). (1994). *Survey Operations Manual—Populations 1 and 2* (Doc. Ref.: ICC889/NRC425). Prepared by the IEA Data Processing Center. Chestnut Hill, MA: Boston College.

Third International Mathematics and Science Study (TIMSS). (1994). *Survey Operations Manual—Population 3* (Doc. Ref.: ICC907/NRC440). Prepared by the IEA Data Processing Center. Chestnut Hill, MA: Boston College.