

Mullis, I.V.S., Kelly, D.L., and Haley, K. (1996). "Translation Verification Procedures" in M.O. Martin and I.V.S. Mullis *Third International Mathematics and Science Study: Quality Assurance in Data Collection*. Chestnut Hill, MA: Boston College.

1. TRANSLATION VERIFICATION PROCEDURES.....1

Ina V.S. Mullis, Dana L. Kelly, and Kathleen Haley

1.1	OVERVIEW.....	1-1
1.2	TRANSLATION GUIDELINES.....	1-4
1.3	ADAPTATION REQUESTS.....	1-5
1.4	TRANSLATION VERIFICATION.....	1-6
1.5	VERIFICATION BY THE TIMSS QUALITY CONTROL MONITORS.....	1-9
1.6	VERIFICATION BY THE INTERNATIONAL STUDY CENTER.....	1-9
1.7	STATISTICAL AND CONTENT-RELATED CONSIDERATIONS.....	1-11
1.8	SUMMARY.....	1-12

1. TRANSLATION VERIFICATION PROCEDURES

Ina V.S. Mullis
Dana L. Kelly
Kathleen Haley

1.1 OVERVIEW

The TIMSS instruments were prepared in English and translated into 30 additional languages across the 45 participating countries (see Table 1.1, below, for the list of languages). In addition, it sometimes was necessary to adapt the international, that is, original, versions for cultural purposes, even for the 11 countries that tested in English. To ensure the standardization of the instruments across languages and countries, and thus the comparability of the TIMSS data, explicit guidelines for translation and cultural adaptation were developed and comprehensive verification procedures were implemented. Specifically, the TIMSS instrument translation effort included the development of explicit guidelines for translation and cultural adaptation; translation of the instruments by the national centers in accordance with the guidelines; verification of the quality of the translations and booklet layout by independent translators; verification by quality control monitors that suggested changes to the translations (if any) were made by the national centers; and a series of statistical checks after the testing to detect items that did not perform comparably across countries.

International versions of the data collection instruments were produced centrally and translated by the national centers in accordance with the translation guidelines.¹ The International Study Center provided each national center with paper and electronic versions of the cognitive items, together with fully assembled booklets in paper format. The questionnaires were provided both in electronic and paper versions. Explicit instructions were provided for translation and cultural adaptation of the cognitive items and assembly and layout of the test booklets and questionnaires. In addition, the questionnaires were accompanied by instructions for adapting certain questionnaire items. Each national center was responsible for producing the translated instruments using the translation guidelines provided and for submitting the translated, camera-ready versions of the instruments to the International Study Center for verification.

¹ The International Coordinating Center (ICC) produced the item booklets for the 1993 item pilot and for the 1994 field trial; the International Study Center produced the instruments for the main survey.

Table 1.1
Languages in Which the TIMSS Instruments Were Administered

Language	Countries
Afrikaans	South Africa
Arabic	Kuwait
Bulgarian	Bulgaria
Chinese	Hong Kong
Czech	Czech Republic
Danish	Denmark
Dutch	The Netherlands
English	Australia, Canada, England, Hong Kong, Ireland, New Zealand, Philippines, Scotland, Singapore, South Africa, United States
Farsi	Iran
Flemish	Belgium
French	Belgium, Canada, France, Switzerland
German	Austria, Germany, Switzerland
Greek	Cyprus, Greece
Hebrew	Israel
Hungarian	Hungary
Icelandic	Iceland
Indonesian	Indonesia
Italian	Italy, Switzerland
Japanese	Japan
Korean	Korea
Latvian	Latvia
Lithuanian	Lithuania
Norwegian	Norway
Portuguese	Portugal
Romanian	Romania
Russian Federation	Russian Federation
Slovak	Slovak Republic
Slovene	Slovenia
Spanish	Argentina, Colombia, Mexico, Spain
Swedish	Sweden
Thai	Thailand

1.2 TRANSLATION GUIDELINES

In 1992, the International Coordinating Center (ICC) drafted guidelines for translation and cultural adaptation of the TIMSS tests based on a paper prepared for TIMSS by Ronald K. Hambleton (1992) regarding the translation of achievement tests into multiple languages. These guidelines were developed to ensure that the cognitive items would be translated from the international versions into the target languages without changes in meaning or difficulty; that cultural differences would be kept to a minimum; and that the meaning and content of the questionnaire items would be retained through translation. The goal was to obtain translated instruments of high quality that would provide comparable data across countries and cultures.

The translation guidelines developed for TIMSS and described below are documented in the *Survey Operations Manuals* (TIMSS, 1994a, 1994b). They recommended that each national center engage a minimum of four translators—two mathematics and two science education specialists with fluency in English and the target language. Only two translators were necessary for the translation of the questionnaire items. The mathematics specialists were to produce two independent translations of the mathematics cognitive items, and the two science specialists were to produce two independent translations of the science cognitive items.

In general, the translators' work included the following:

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

The *Survey Operations Manuals* (TIMSS, 1994a, 1994b) also provide guidelines regarding decisions about vocabulary, meaning, layout, and cultural adaptations. These guidelines include examples of acceptable and unacceptable cultural adaptations to the items, such as changes in punctuation, units, proper nouns, common nouns, spelling, verbs (not related to content), and usage. Furthermore, translators were instructed that when modifying the text of an item for cultural adaptation purposes the following were to remain the same as the international version:

- The meaning of the question
- The reading level of the text
- The difficulty of the item
- The likelihood of another possible correct answer for the test item.

Each pair of translators (for mathematics and for science) was to translate the test items independently and then have the two versions compared by a third party. When there were differences between the two versions, the best version of the translation was selected. Any deviations in vocabulary, meaning, or item layout from the international versions were recorded on the Translation Deviation Form. The completed forms were then submitted to the International Study Center together with the translated instruments, and used during the international translation review.

Countries were also encouraged to send their suggestions for adaptations to the International Study Center as they were translating the items. This procedure, described further in section 1.3, allowed the national centers to receive approval of adaptations quickly, which expedited the preparation of the test booklets.

Due to limited resources, some countries were unable to engage more than one translator each for mathematics and science. However, by engaging translators with the appropriate and recommended qualifications and by adhering to the international procedures for translation, countries were still able to produce high-quality translated instruments which were commensurate with the international versions.

1.3 ADAPTATION REQUESTS

While the intention of TIMSS was to devise items that would be interpreted similarly across countries, it was clear that a single version, strictly translated, would not be wholly appropriate in every country. Therefore, a procedure was put in place for requesting approval for specific adaptations to the items. This process was distinct from the translation verification process in that it allowed an NRC to obtain timely approval for particular adaptations, usually within two days. This expedited the production of the test booklets at the national centers and brought any deviations to the attention of the subject area coordinators and the International Study Center prior to the printing and duplication of the booklets at the national centers.

In order to gain approval for an adaptation, the NRC documented the request on the Item Adaptation Request Form and sent it to the International Study Center. This took place either during the national translation and booklet production effort or after submission of testing materials for translation verification. Upon receipt of the request forms at the International Study Center, the adaptations were forwarded to the appropriate subject area coordinator for review. The math or science coordinator considered each request and approved the change, rejected it, or specified additional modifications necessary to make it acceptable. Staff at the International Study Center were responsible for informing the national centers whether the requested adaptations were approved.

Item adaptations were approved if they did not in any way change the substance or intent of the question or answer choices. For example, a change from “weight” to “weight (mass)” was an acceptable clarification for students unaccustomed to the colloquial use of “weight,” while simply changing “weight” to “mass” would be unacceptable as it would

make the units inappropriate. Similarly, requests from a number of countries to replace “congruent” with “same shape and size” were approved, whereas the use of “equal” instead of “congruent” was regarded as too imprecise and therefore was not approved.

1.4 TRANSLATION VERIFICATION

The items administered in the 1993 item pilot were translated by the national centers according to the guidelines described in section 1.2 and submitted to the ICC for verification. The translation verification was conducted after the items had been administered, the results had been analyzed, and items that were potential candidates for the field trial had been identified. Each NRC received a report on the quality of the translation of each of those items and had the opportunity to improve the translation, if necessary, for the field trial.

In the 1994 field trial, a slightly different approach was taken in order to conserve resources. If a country’s item pilot translations had been deemed acceptable by the internationally commissioned translators, a 25% sample of the items was reviewed (following administration) to verify that the same quality existed in the field trial instruments. If a country’s item pilot translations were not acceptable, or if that country did not participate in the pilot study, then all of the items were checked for the quality of the translation (Maxwell, 1996).

In the main survey, procedures required that the translated items and camera-ready copies of the test booklets be verified prior to printing and administration of the tests. Following translation and assembly of the test booklets, countries were required to send the items and the assembled booklets for each population that participated to the International Study Center. Additionally, the Population 2 performance assessment student booklets were verified. By having the translations verified prior to duplication, any errors or deviations found in the international review could be corrected. Due to the tight timeline between the completion of the international versions of the instruments and the main survey administration, some countries were not able to have their translations verified prior to printing. In these cases, efforts were made to ensure that the instruments were nonetheless verified.

Once the translated items, camera-ready test booklets, and completed Translation Deviation Forms were received by the International Study Center, they were forwarded to the ICC in Vancouver, Canada, for review by professional translators.² When NRCs included Item Adaptation Request Forms, the International Study Center forwarded the requests to the subject area coordinators for review. The translation of the items and booklet layout was checked by translators from the same professional translation company that had completed the verification of the instruments in the item pilot and field trial. The agency is based in Vancouver. Beverley Maxwell of the ICC coordinated the international

² Translators reviewed instruments translated into languages other than English; booklets adapted by English-testing countries were reviewed by staff at the ICC.

review. She communicated with the translators regarding procedures, reviewed translators' reports and suggestions, and forwarded the translation verification reports to the national centers and the International Study Center. All those who verified the national translations had formal credentials as translators into the target language, first-language experience in the target language, excellent knowledge of English, experience living and working in an English-language environment, and familiarity with the culture associated with the target language.

Translators were provided with a number of materials to aid them in their understanding of the translation procedures used by the national centers and with instructions to carry out the review of the instruments. These materials are listed below (excerpted from Maxwell, 1996).

- 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 booklets or clusters)
- A set of the international originals
- 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.

For each country, a translator reviewed the overall layout of the instruments, the translation of the student instructions, and the translation of each item. The translator compared each translated item with the international version and documented any adaptations in a translation verification report. The translator assigned to each item that differed from the international version a code for the *type* of deviation and a code for the *severity* of deviation. The translator further provided an explanation of the change and how it could be corrected or improved upon, if necessary. The “type codes” and “severity codes” are described below.

TYPE CODES

The type codes, listed below, indicate what kind of change was made in the translation from the international version to the target language. Codes A through J refer to deviations in the text of an item; K through N refer to deviations in the graphics or layout of an item.

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.

SEVERITY CODES

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

1. **Major Change or Error:** This could affect the results and NRCs were to make corrections. 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 was to be corrected if possible, but would 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 have been adequate, but the verifier suggested a different wording for the item. The NRC was asked to review such suggestions and decide whether to make the suggested changes.
4. **Acceptable Changes:** The verifier identified acceptable changes and appropriate adaptations. This was done to provide information and required no action from the NRC. An example is where a reference to winter was changed from January to July for the Southern Hemisphere.

A code, comprised of the severity code and the type code, was assigned to each item for which the translator noted a deviation from the international version. 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 upon in the explanation (Maxwell, 1996).

The translation verification report 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 by Beverley Maxwell and subsequently forwarded to the International Study Center and the appropriate NRC. The above procedure required between four and six weeks; less time was required for instruments adapted for English-testing countries. Verification reports for all countries were entered into a database at the International Study Center.

1.5 VERIFICATION BY THE TIMSS QUALITY CONTROL MONITORS

When visiting the national center, the quality control monitor checked the final instruments against the translation verification report prepared by the independent reviewer to ensure that the suggestions for corrections and improvements had been followed. The quality control monitor recorded this in his/her report to the International Study Center.

1.6 VERIFICATION BY THE INTERNATIONAL STUDY CENTER

After translation verification, a further quality check of the instruments was made by the International Study Center. All deviations/adaptations coded 1 (major change or error) in the translation verification report were reviewed to determine whether a threat to validity existed. Final printed booklets were inspected to determine whether the error had been corrected. Between the verification conducted by the quality control monitors and that conducted by the International Study Center, it was determined that nearly all corrections to items coded as Type 1 had been made at Population 2. (Checking at Populations 1 and 3 is still in progress.) Table 1.2 presents for each country the number of translation deviations coded as Type 1 that were still present in the final Population 2 test booklets.

Table 1.2
Number of Potential Translation Errors, After Checking Final Test Booklets

Country	Population 1		Population 2		Population 3		
	Math	Science	Math	Science	Literacy	Advanced Math	Physics
Australia	0	0	0	0	0	0	0
Austria	1	0	0	0	0	0	0
Belgium (Fl)	0	0	0	0	–	–	–
Belgium (Fr)	–	–	*	*	–	–	–
Bulgaria	0	0	0	0	0	0	0
Canada (Eng)	0	0	0	0	0	0	0
Canada (Fr)	0	0	0	0	1	0	0
Colombia	–	–	0	0	–	–	–
Cyprus	0	0	0	2	1	1	0
Czech Republic	0	0	0	0	1	0	1
Denmark	–	–	0	0	0	0	0
England	0	0	0	0	–	–	–
France	0	–	0	0	0	0	0
Germany	–	–	*	*	*	*	*
Greece	0	0	0	0	*	*	*
Hong Kong	0	0	0	0	0	1	0
Hungary	0	0	1	0	*	–	–
Iceland	1	0	0	0	*	–	–
Indonesia	5	5	5	2	–	–	–
Iran	*	*	*	*	–	–	–
Ireland	0	0	0	0	–	–	–
Israel	0	0	0	1	*	*	*
Japan	0	1	0	0	–	–	–
Korea	*	*	*	*	–	–	–
Kuwait	1	1	5	1	–	–	–
Latvia	0	0	3	0	–	–	0
Lithuania	–	–	0	0	1	0	–
Mexico	0	0	0	0	1	*	*
Netherlands	0	0	0	0	0	–	–
New Zealand	0	0	0	0	0	0	0
Norway	0	0	0	0	0	–	0
Philippines	–	–	0	0	–	–	–
Portugal	0	0	0	0	–	–	–
Romania	–	–	1	0	–	–	–
Russia	–	–	*	*	*	*	*
Scotland	0	0	0	0	–	–	–
Singapore	0	0	0	0	–	–	–
Slovak Republic	–	–	0	1	–	–	–
Slovenia	0	0	0	0	0	2	0
South Africa (Afr)	–	–	*	*	*	–	–
South Africa (Eng)	–	–	*	*	*	–	–
Spain	–	–	0	0	–	–	–
Sweden	–	–	0	0	0	0	0
Switzerland (Fr)	–	–	0	0	0	0	0
Switzerland (Ger)	–	–	0	0	0	0	0
Switzerland (It)	–	–	0	0	0	0	0
Thailand	0	0	1	1	–	–	–
United States	0	0	0	0	0	0	0

* Final test booklets were unavailable for review.

– Did not participate.

1.7 STATISTICAL AND CONTENT-RELATED CONSIDERATIONS

TIMSS also conducted a set of elaborate statistical checks on the data to further ensure that items were performing comparably across countries. Although only a small number of items were found to be inappropriate for international comparisons, throughout the series of item-checking steps a number of reasons were discovered for differences in items across countries. Most of these were inadvertent changes in the items during the printing process, including omitting an item option or misprinting the graphics associated with an item. However, differences attributable to translation problems were found for an item or two in several countries.

Each country was provided with its item analysis information. Specially produced by the IEA Data Processing Center (DPC), these data contained automatic flags for a number of conditions that can indicate an item may not be performing properly. In addition, the Australian Council for Educational Research (ACER) produced graphical representations of item statistics for each participating country. Two countries deleted one Population 2 item each based on this information. Examples of these materials are displayed in Appendix A.

Prior to the international scaling of the Population 2 achievement data by ACER, the International Study Center conducted a thorough review of the item statistics for all participating countries. The process was empirically based, with data about the items being produced from several sources, including the translation verification process, the IEA DPC, and the item analysis information specially prepared for TIMSS by ACER. The intention was to detect inadvertent errors that were made during the processes of translation, printing, coding, and data entry.

As shown in Figure A.1 in Appendix A, the IEA DPC summarized on a page the item analysis results across countries for each item. The IEA DPC also produced information about the inter-rater agreement for the free-response items. ACER produced across-country graphical representations of item statistics, indices of fit, and item-by-country interactions. Figure A.2 in Appendix A provides an example of this type of item information. ACER screened the item statistics for particular problems, such as positive point-biserials for any non-key options and negative point-biserials for the key (see Figure A.3 in Appendix A). These summaries were particularly useful in highlighting items with potential problems.

In particular, items with the following problems were checked for possible deletion from the international database:

1. Errors were detected in the translation verification process that were not corrected.
2. The data cleaning process revealed more or fewer options than in the original version of the item.
3. The item analysis information showed the item to have a negative biserial.

4. The item-by-country interaction results showed a very large negative interaction for that country.
5. The item-fit statistic indicated the item was not fitting the model.
6. For free-response items, the within-country scoring reliability data showed an agreement of less than 70% for the score level. Also, performance in items with more than one score level was not ordered by score, or correct levels were associated with negative point-biserials.

The statistics and translation verification documentation were used as pointers toward checking actual booklets and contacting NRCs. If a problem could be detected by the International Study Center (such as a negative point-biserial for a correct answer or too few options for the multiple-choice questions), a decision was made to delete the item from the international scaling. However, if there was a question about potential translation or cultural issues, then the NRC was queried, and the International Study Center abided by the decision made by the NRC. In several cases, NRCs consulted mathematics or science experts before making a decision.

Considering that the checking involved approximately 500 items for each of more than 40 countries, very few deviations from the international format were revealed. Appendix B contains a list of the changes made in the international database for Population 2. Twenty-one countries had one or more items (usually only one) deleted as a result of translation, adaptation, or printing deviations. For three countries, options became switched in printing but were corrected in the database.

1.8 SUMMARY

Because the international versions of the TIMSS instruments were prepared in English, translating the materials into the 30 different languages of testing used by the 45 participating countries, and adapting the originals for countries testing in English, was an enormous challenge. Considerable energy, time, and resources were expended to help ensure that the translations yielded comparable test instruments across the countries.

The international English versions of the instruments were produced centrally by the TIMSS International Study Center, and then translated by the national centers in accordance with detailed translation guidelines. These guidelines included information about cultural adaptations, item difficulty, reading level, and layout as well as about retaining the meaning of the items. The specified procedures called for using two independent translators each for mathematics and science and having a third translator compare the versions. The best version was selected when there were differences in the translations. Most countries were able to adopt these procedures although resources were a difficulty in some instances.

The national centers submitted their translated instruments to the TIMSS International Study Center for review by professional translators. The translators completed verification forms noting any errors that needed to be corrected and making suggestions for improvements. These forms were returned to the national centers to make

any necessary changes. Although time constraints made complete checking of camera-ready materials for all countries prior to testing impractical, the printed materials were checked thoroughly by both the TIMSS quality control monitors and the TIMSS International Study Center. Using information from the translation verification process in conjunction with a rigorous review of the item statistics did reveal some mistranslated and misprinted items for some countries. Such items are being deleted from the international database before scaling and analysis. Considering, however, the number of countries, languages, and items involved in the TIMSS testing, very few deviations from the international format were revealed in the final printed instruments. For example, at Population 2, most countries did not have any items deleted as a result of translation or printing problems.

REFERENCES

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

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 Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.

Third International Mathematics and Science Study (TIMSS). (1994a). *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). (1994b). *Survey Operations Manual–Population 3* (Doc. Ref.: ICC 906/NRC439). Prepared by the IEA Data Processing Center. Chestnut Hill, MA: Boston College.