

## Problems in Translating International Educational Assessment Instruments: a literature review

LUELA LIÇI, MSc

PhD candidate, University of Tirana

### Abstract

*The paper presents a literature review regarding problems in translating international educational assessment instruments used in large-scale studies organized by the IEA and OECD. The translation process has faced challenges, issues and problems that have impacted in some aspects the quality of translation. The paper presents the views of researchers on the translation problems faced during translation process. In the focus of this paper are viewpoints on problems regarding to translating tests and especially linked to equivalence and translator quality. The specific purposes of this paper are: 1) to present the viewpoints of researcher about the translation problems; 2) to draw attention to the issues encountered in the translation process during 20-year experience of translating assessment instruments used in the international assessments such as PISA, TIMSS and PIRLS; and 3) to present recommended solutions to the identified problems. Understanding the types and the nature of the translation problems helps test developers, teams of the international educational assessments, translators, and researchers to contribute in improving translation procedure, to ensure the quality of the instruments, and to ensure the effectiveness of their tasks. The literature review is based on theories of translation equivalence, on findings and lesson learnt presented in the different research papers and international assessment reports.*

**Keywords:** international assessment, problem, translation, translation equivalence, translation problem, translation quality, translator

## INTRODUCTION

Translating international achievement tests and questionnaires from the source language to the target language has been a long-standing practice.

For years, assessment instruments have been translated and adapted from one language and culture to another in order to use existing products with different target populations for international test comparisons (Fegert, 1989; Geisinger, 2003). It has been long indicated that the challenge of properly translating and adapting assessment instruments across different languages involves providing equivalent, culturally-appropriate, and fair instruments with meaningful, interpretive, and valid outcomes (Allalouf, 1999; Budgell, 1992; Chang & Myers, 2003; Hambleton, Merenda, & Spielberger, 2005; Misra, Sahoo, & Puhan, 1997; Sireci & Bastari, 1998; Tanzer, 1998).

In the last two decades international comparative assessments, such as PISA, TIMSS, and PIRLS have been administered in different education systems worldwide. Translation quality is one of the most important issues as the assessment instruments are translated into different local language versions. This process has encountered challenges to the validity of the results due to the quality of the translation. As such, test developers and researchers have emphasized high quality test translation and rigorous test translation review procedures as critical to ensuring valid testing. (Hambleton, 2001; Hambleton, Merenda, & Spielberger, 2005; Solano-Flores, Backhoff, & Contreras-Niño, 2009).

The practice and research conducted regarding the translation of international assessment instruments has discovered a number of translation challenges and problems that required and actually ask for solutions.

There is considerable technical evidence that suggests that the quality of test translations varies considerably, and often the translations are not very good, thus reducing the validity of any results produced with the translated tests and questionnaires (Hambleton, 1993). Some researchers have presented in their reports evidence and conclusions that with respect to cross-national studies of achievement, there are many threats to the validity of the interpretations of the findings including threat of an improper translation.

Hambleton (1993) has reviewed issues and methods associated with translating achievement tests and has identified four important problems that arise in translation which must be considered: 1) the selection of translators; 2) identifying the appropriate language for the target version of the test; 3) identifying and minimizing cultural differences; and (4) finding equivalent words or phrases.

Hambleton (1993) has reviewed the problems and in each case has offered suggestions on how the problems might be addressed in practice. The extent to which the four problems occur in practice depends upon many factors including the test format, test content, test difficulty, the particular language and/or cultural groups involved, the expertise of the test developers and translators, and the amount of verbal load in the test.

Brislin (1970) has reported (1) the languages involved can greatly influence the difficulty of the test translation process, so the more similar the structure (e.g., English and French are more similar than English and Chinese) the better the translation, (2) the technical knowledge of the translators is an extremely important factor, and (3) translations tend to be better if translators are given practice and feedback before they begin the task.

In this paper the translation problems are presented through four dimensions: translation equivalence theory and translation problems; theory, analysis and models on translation problems; translator quality problems; recommendations for translation problem elimination. In this way, the literature review helps with a better understanding of the nature of translation equivalence, the nature of translation problems. It also helps to identify translation problems, and to recognize recommendations for eliminating errors.

### **Translation equivalence and translation problems**

Equivalence is said to be the central issue in translation although its relevance, definition and applicability within the field of translation theory have caused controversy, and many different theories of the concept of equivalence have been elaborated. The concept of equivalence has been of particular concern to translation scholars since it has been closely linked with both definitional and practical aspects of translating. Becoming an essential feature of translation theories in the 1960s and 1970s, equivalence was meant to indicate that source text

(henceforth ST) and target text (henceforth TT) share some kind of „sameness“. The question was as to the kind and degree of sameness which gave birth to different kinds of equivalence. The equivalence was conceptualized by the following scholars in the field, namely, Vinay and Darbelnet (1958), Jakobson (1959), Nida and Taber (1969), Catford (1965), House (1997), Newmark (1981), Baker (1992), and Pym (2010). For a better understanding of the translation equivalence problems, a summary of the concept of equivalence would be helpful.

Vinay and Darbelnet in their works (1958, 1995) distinguish between *direct* and *oblique* translation, the former referring to literal translation and the latter to free translation. Moreover, they propose seven procedures, the first three covered by direct translation and the remaining four by oblique translation. These procedures are: borrowing, calque, literal translation, transposition, modulation, equivalence and adaptation. In particular, it is argued that equivalence is viewed as a procedure in which the same situation is replicated as in the original but different wording is used (Vinay and Darbelnet, 1995). Jakobson (1959) elaborates three kinds of translation, that is, *intralingual* (rewording or paraphrasing within one language), *interlingual* (rewording or paraphrasing between two languages), and *intersemiotic* (rewording or paraphrasing between sign systems). It is *interlingual translation* that has been the focus of translation studies. Jakobson's study of equivalence gave new insight to the theoretical analysis of translation since he introduced the notion of 'equivalence in difference'. On the basis of his semiotic approach to language and his aphorism, 'there is no *signatum* (meaning) without *signum*' (sign or verbal code), he suggests three kinds of translation; intralingual (within one language, i.e. rewording or paraphrasing), interlingual (between two languages), and intersemiotic (between sign systems) (Jakobson, 1959). Jakobson (1959) further claims that the translator makes use of synonyms in the case of interlingual translation in order to get the ST message across. This means that in interlingual translations there is no full equivalence between code units. According to this theory, translation is about two equivalent messages in two different codes (Jakobson, 1959). According to Jakobson, from a grammatical point of view, languages may differ from one another to a lesser or greater degree, but this does not mean that a translation cannot be possible. In other words, the translator may face the problem of not finding a translation equivalent. He suggests that when there is

deficiency, terminology may be amplified and qualified by loanwords or loan-translations, neologisms or semantic shifts and by circumlocutions (Jacobson, 1959).

With regard to equivalence, Nida 1964, 1969) explains that there are two basic types of equivalence: (1) *formal equivalence* and (2) *dynamic equivalence*. In particular, Nida argues that in formal equivalence the TT resembles very much the ST in both form and content whereas in dynamic equivalence an effort is made to convey the ST message in the TT as naturally as possible. It could be argued that Nida is in favor of dynamic equivalence since he considers it to be a more effective translation procedure.

Catford (1965) developed what he calls linguistic theory to translation. Catford's main contribution in the field of translation studies lies in the introduction of his idea of *types* and *shifts* of translation. *Shifts* refer to the changes that take place during the translation process. More specifically, Catford describes very broad types of translation according to three criteria. Firstly, *full translation* is contrasted with *partial translation* which differs according to the extent of translation. Secondly, *total translation* differs from *restricted translation* according to the levels of language involved in translation, and, thirdly, Catford distinguishes between *rank-bound translation* and *unbounded translation*, depending on the grammatical or phonological rank at which equivalence is established with regard to translation shifts, Catford (1965) defines them as departures from formal correspondence when translating from the SL to the TL. Moreover, he maintains that there are two main types of translation shifts, that is, *level shifts* (where an SL item at one linguistic level, for example grammar, has a TL equivalent at a different level, for instance lexis) and *category shifts*, which are divided into (a) structure-shifts involving change in grammatical structure, (b) unit-shifts involving changes in rank, (c) class-shifts involving changes in class, and (d) intra-system shifts which occur internally when source and target language systems share the same constitution but a non-corresponding term in the TL is selected when translating.

According to House (1997) a translation model in which the basic requirement for equivalence of ST and TT is that original and translation should match one another in function. This function should be achieved by employing equivalent pragmatic means. The translation is only, therefore, considered to be adequate in quality if it matches the

„textual“ profile and function of the original. House has distinguished between two basic types of translation, namely, *overt translation* and *covert translation*. As the term itself denotes, an overt translation points to a TT that consists of elements that „betray“ that it is a translation. On the other hand, a covert translation is a TT that has the same function with the ST since the translator has made every possible effort to alleviate cultural differences.

Koller (1979) distinguishes five different types of equivalence: (a) *denotative equivalence* involving the extra linguistic content of a text, (b) *connotative equivalence* relating to lexical choices, (c) *text-normative equivalence* relating to text-types, (d) *pragmatic equivalence* involving the receiver of the text or message, and, finally, (e) *formal equivalence* relating to the form and aesthetics of the text (p.186-191). Having identified different types of equivalence, Koller (1979) goes on to argue that a *hierarchy of values* can be preserved in translation only if the translator comes up with a hierarchy of equivalence requirements for the target text.

Newmark's in two works *Approaches to Translation* (1981) and *A Textbook of Translation* (1988) describes a basis for dealing with problems encountered during the translation process. Newmark replaces Nida's terms of formal and dynamic equivalence with *semantic* and *communicative translation* respectively. The major difference between the two types of translation proposed by Newmark is that semantic translation focuses on meaning whereas communicative translation concentrates on effect. Newmark (1981) strongly believes that literal translation is the best approach in both semantic and communicative translation. However, he is careful to note that when there is a conflict between the two forms of translation, then communicative translation should be favoured in order to avoid producing an abnormal, odd-sounding or semantically inaccurate result.

Baker (1992) combines linguistic and communicative approaches to translation. She analyzes equivalence at different levels taking into account the problems that translators may be faced with during the translation process. Baker introduced such notions as equivalence at word level, above word level, and grammatical, textual, and pragmatic equivalence. Baker in her book *In Other Words* (1992) adopted a more neutral approach when she argues that equivalence is a relative notion because it is influenced by a variety of linguistic and

cultural factors. Baker's contribution to the field of translation studies is recognized especially for offering a systematic approach to training translators through the elaboration of specific strategies that can be used to deal with the considerable translation problems.

Pym (2010) elaborated the concept of equivalence by pointing out that there is no such thing as perfect equivalence between languages and it is always *assumed equivalence*. In particular, for Pym (2010) equivalence is a relation of “equal value” between an ST segment and a TT segment and can be established on any linguistic level from form to function. He differentiates *natural* and *directional equivalence*. Natural equivalence exists between languages prior to the act of translating, and, secondly, it is not affected by directionality. On the other hand, theories of directional equivalence give the translator the freedom to choose between several translation strategies which are not dictated by the ST. Although there are usually many ways of translating, the strategies for directional equivalence are reduced into two opposing poles; one adhering to SL norms and the other to TL norms. Perhaps, the most important assumption of directional equivalence is that it involves some kind of asymmetry since when translating one way and creating an equivalent does not imply the creation of the same equivalent when translating another way (Pym, 2010).

### **Theory, analysis, and models on translation problems**

The research in the field of international comparisons has shown that even a slightly inaccurate translation of a word can be enough to affect the differential operation of an item (Ericka, 1998). The procedures used in the translation of tests in international comparisons have evolved in accordance with the developments of knowledge.

A number of researchers (Lörscher, 1991; Hambleton, 2001; Hambleton, Merenda, & Spielberger, 2005; Solano-Flores et al., 2005; 2009; 2013; Arffman, 2007) have identified and analyzed translation problems and developed theories, models, and recommendations of how to overcome translation problems.

Lörscher (1991) has defined translation problems as (linguistic) problems that the translator is faced with when making a translation. A translation problem, more specifically, occurs when the translator

realizes that s/he is unable to transfer adequately a source language text segment into the target language (p. 80).

Arffman (2007) mention that it is important to note that in the definition – based on Lörscher's study focusing exclusively on the translator – translation problems only include those which the translator him/herself considers to be problems; translation problems regarded as such by analysts, for example, on the other hand, are not considered be translation problems.

Lörscher (1991) has defined the translation strategies as procedures which the subjects employ in order to solve translation problems. He has explained that translation strategies are procedures for solving translation problems. They range from the realization of a translational problem to its solution or the realization of its insolubility by a subject at a given moment. They are constituted by those minimal problem-solving steps.

According to Lörscher (1991), translation strategies have their starting-point in the realization of a problem by a subject, and their termination in a (possibly preliminary) solution to the problem or in the subject's realization of the insolubility of the problem at the given point in time. Between the realization of a translation problem and the realization of its solution or insolubility, further verbal and/or mental activities can occur which can be interpreted as being strategy steps or elements of translation strategies. They can be formalized to yield categories of a model for the strategic analysis of the translation process. Such a model was developed on the basis of a corpus of translations made by foreign language students. In the second stage of the project it was applied to translations performed by professional translators. Modifications of the model were hardly necessary for an adequate analysis of professional translation processes although the quality and structure of the translation strategies and their elements, as well as their quantitative distribution differ considerably, at least in part. The model consists of two hierarchical levels. The first and lowest contains those phenomena which can be interpreted to be elements of translation strategies, i.e. the smallest detectable problem-solving steps. The second level captures the manifestations of translation strategies. Translation versions can be located within strategies or can comprise several strategies, and are thus intra- or interstrategic phenomena. The translation process contains both strategic phases,



which are directed towards solving translational problems, and non-strategic phases, which aim at accomplishing tasks.

Guillermo Solano-Flores, Luis Ángel Contreras-Niño, and Eduardo Backhoff (2005, 2009) developed the theory of test translation error (TTTE), which serves as a conceptual framework for more effective translation work and guides the implementation of more practical test translation review procedures.

Solano-Flores et al. (2006) present a conceptual model and methodology for the review of translated tests in the context of such international comparisons as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA). Their conceptual framework is based on five premises: (a) the inevitability of error in translating tests; (b) the dimensions of mistranslation; (c) the relativity of the dimensions of error; (d) the multidimensional nature of translation errors; and (e) the probabilistic nature of the acceptability of an item's translation.

Solano-Flores et al. (2009) examined the translation of PISA test items based on the theory of test translation error (TTTE), which has proven to allow detection of translation errors with unprecedented levels of detail. Translation error (TE) is defined as the lack of equivalence between the original and translated versions of items on multiple translation error dimensions (TEDs) that involve design, language, and content. The theory of test translation error (TTTE) is not only about errors made in test translation, but also about errors in translated tests.

“According to our theory, test translation does not refer exclusively to the action of translating items but also to multiple aspects of the entire process through which translated versions of those items are created. Translation error does not result exclusively from poor translation job (e.g., inaccuracy of a chosen term, word-by-word translation, use of false cognates); it also results from factors that are beyond the translators' translation skills”, wrote Guillermo Solano-Flores et al. (2009).

The theory of test translation error (TTTE) developed by Solano-Flores et al. (2009), postulates the existence of test translation error dimensions (TEDs), grouped in three broad categories: design, language, and content. Each TED comprises several types of TE. While it parallels the systems of dimensions and types of TEs used in other investigations (e.g., Backhoff et al., 2011; Solano-Flores et al., 2009),

the definitions of TEDs and the types of TE they comprise were respectively adapted and included with the intent to meet the needs of this particular translation review project. The theory postulates that test TE is multidimensional.

Finally, the theory postulates that there is a tension between TEDs. Actions intended to avoid TE on a given TED may involve making errors on other TEDs. Accordingly, error-free test translation is impossible; effective test translation minimizes but does not eliminate error. The theory also postulates that while items usually have multiple TEs, most of them are mild and even unnoticeable. Objectionable translated items have many and severe TEs and are likely to pose serious linguistic challenges to examinees who are given the translated version of a test (SolanoFlores et al., 2009).

Evidence from empirical studies (e.g., Solano-Flores et al., 2005; 2009; 2013) suggests that the TTTE can effectively contribute to ensuring more valid testing in the context of international test comparisons. First, the translation error dimensions can be used for reviewing translated test items. Reviewers with different areas of expertise examine translation errors according to the error dimensions relevant to their fields. Independent coding and group discussions are used in the review procedure to finalize their translation review decisions concerning error types. This approach appears to allow the detection of translation errors with high, unprecedented levels of sensitivity that may not be detected by the conventional translation review approach, thus addressing the limitations of procedures that rely on either *back translation* or *double translation* (e.g., Grisay, 2003; 2006). Second, these TTTE-based procedures make it possible to quantify translation errors, for example, by calculating the number of different error types or different error dimensions observed (identified by consensus) in translated items. This has made it possible to correlate translation error and item difficulty in PISA and TIMSS translated items (Solano-Flores et al., 2009; 2013), which in turns allows for making more valid interpretation of students' test scores on translated tests.

According to the theory, translation error (TE) results not only from poor translation, but also from factors that are beyond the translators' skills (e.g., languages encode meaning in different ways). In the discussions, six major factors causing translation problems came up: characteristics of the source instrument; deficiencies in the

competences of the translators; the vague goal of the translation task and deficiencies in the translation guidelines and translation notes; the use of parallel source versions; deficiencies in revision and verification; and time pressure and haste (Solano-Flores et al., 2009)

In translation studies, it has long been known that the source text plays a significant role in how difficult the translation task is and what the quality of the translation will be like. After all, it is the source text from which the translation task starts and which forms the basis of the entire translation task. For example, certain linguistic features are known to be difficult to translate. These include, for instance, unfamiliar (e.g., technical) topics and vocabulary, abstract concepts, broad and vague meanings, dense language (a lot of information packaged in a compact, condensed form; e.g., complex noun phrases, reduced clauses), word order, textual factors, stylistic and aesthetic factors, connotative and associative meanings, non-literal language (e.g., metaphors, puns), phonological and orthographic devices, and conventions (implicit or tacit non-binding regulations of behaviour, based on common knowledge and expectations; Nord, 1991, p. 96) (Baker, 1992; Campbell, 1999; Danks & Griffin, 1997; Hale & Campbell, 2002; Kuhiwczak, 2003; Nord, 1997; Wilss, 1990).

Usually, the difficulties caused by these factors arise from differences between languages. Because of these differences, the translation cannot use, for example, similar sentence structures, word order, metaphors or style as the source text, or the words in the translation do not have the same connotations or aesthetic values as those in the source text. In addition to this, however, translation problems may also stem from comprehension problems. As a result of these problems, the meaning of the source text may be changed, or the translation may end up being unduly literal (Chesterman, 2010; Pym, 2008). Also, the translation easily becomes simpler, or “flatter”, than the original text (Laviosa-Braithwaite, 1998).

Literary texts are typically loaded with stylistic, aesthetic and artistic overtones and multiple and multi-layered (e.g., connotative and associative) meanings, which typically cannot be transferred equivalently across languages (e.g., Danks & Griffin, 1997; Hassan, 2011; Kuhiwczak, 2003; Lefevere, 1992; Neubert & Shreve, 1992; Scarpa, 2002; Bruner, 1986; Rosenblatt, 1994). Literary texts are therefore usually considered the most difficult text type to translate. If, furthermore, the text is a poem, there is the extra difficulty as to

whether or to what extent to maintain its formal features (alliteration, rhyme, etc; Danks & Griffin, 1997), because normally both form and content cannot be preserved at the same time (Bell, 1991).

Special language, or scientific or technical, texts, for their part, focus more on facts and universal and scientific truths and operate with basic and literal meanings (Bruner, 1986; Neubert & Shreve, 1992; Newmark, 2003; Rosenblatt, 1994; Scarpa, 2002). They are therefore generally thought to be easier to translate than, for example, literary texts (Kuhiwczak, 2003; Scarpa, 2002; Wilss, 1990). However, if the translator lacks knowledge of the facts, topics and (technical) terms discussed in the text, comprehension is hindered, which, in turn, easily leads to errors and overly literal and incomprehensible translations (Kim, 2006; see also Danks & Griffin, 1997). Another factor which may add to the difficulty of both understanding and translating special language texts is their lexical (number of content words per clause) and syntactic density (number of words per clause) (cf. Halliday & Martin, 1993).

In test translation, too, poorly designed source instruments have been a significant source of translation and equivalence problems (e.g., Harkness, Edwards, Hansen, Miller & Villar, 2010).

More attention thus needs to be paid to making the source instrument translatable (see also Brislin, 1986, p. 143). This seems to be the case with the question items, in particular (Alderson, 2000). Thus, when formulating the questions, it is good to avoid using words and structures that are hard to find good and natural equivalents for and/or that are hard to understand. This means avoiding, for example, dense language, sentence-initial adverbials, technical terms, stylistic and aesthetic devices, vague meanings (broad, abstract and affective meanings, words with several meanings), idioms, metaphors and puns (Allalouf, 2003; Allalouf, Rapp & Stoller, 2009; Arffman, 2007; Elosua & López Jauregui, 2007; Ercikan, 2002; Gierl & Khaliq, 2001; Grisay, 2004; Solano-Flores, Backhoff & Contreras-Niño, 2009). It also means seeing to it that the sentences used in the items are short and simple: that they contain less than 16 words and only one idea (Brislin, 1986).

When selecting and preparing stimulus materials for international studies of reading literacy, the advice to make them translatable is problematic. This is because it jeopardizes the construct validity of the tests. The tests should assess testees' reading skills in as

diverse and authentic contexts as possible. This requires that the tests cover a wide variety of different types of texts with as diverse language as possible. (OECD, 2009b.)

Also, when designing the instrument, it is good to make use of all the knowledge that exists today on factors that have a negative effect on translatability. In PISA, for example, such data are collected by means of verification reports, and on the basis of the data, instructions are written on how to prepare as translatable source instruments as possible (Dept, Ferrari & Wäyrynen, 2010).

Other methods for improving the translatability of the source instrument include *ex ante*, or “advance translation” (Dorer, 2011; Harkness & Schoua-Glusberg, 1998): Translators and researchers translate the source instrument into one or more languages before it is finalized, during the drafting stage. They comment on the translation process and problems encountered, and this feedback is used when finalizing the source instrument. In PISA, the translation of the French source version from the English version has also served as a kind of *ex ante* translation (Grisay, 2003; OECD, 2009a).

In some other studies, too, the English source instrument has been translated into some other languages, and countries have then chosen which one of these versions they have used as the basis of their translations. For example, the English versions of the instruments used in the IEA Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS) have been translated into Arabic (since 2007), and the instrument in the Literacy Assessment and Monitoring Programme (LAMP) study has been rendered into French and Spanish. In all these studies, the *ex ante* translations have helped to make the source instrument more translatable (Dept, Ferrari & Wäyrynen, 2010). Because of differences between languages, source instruments can never be made fully translatable (especially in reading literacy studies). There is a technique in translation studies that might help to deal with language-specific differences and the issue of untranslatability and to improve the construct validity of the tests: compensation (see also Arffman, 2007).

Compensation means that if, for example, a metaphor in the source text cannot be translated as a metaphor, a metaphor is added somewhere else in the translated text (where it does not exist in the source text) so as to keep the number of metaphors in the translation

the same as it is in the source text (Baker, 1992). The technique, however, has its problems. First, it only works on the level of the text, helping to ensure that the translated text as a whole is equivalent (e.g. stylistically) to the source text. It does not usually work on the level of individual items and cannot be used to ensure that individual metaphors are translated in an equivalent way. Second, using the technique is extremely demanding and requires highly qualified, experienced and innovative translators. Third, for it to be possible for translators to use the technique and to be creative, for example, a considerable amount of time is needed (see e.g., Fontanet, 2005).

Harkness (1998) has written about the problems of translating surveys and questionnaires. She notes that an extensive amount of research is reported in the methodology of survey construction literature, but little of it relates to the uses of surveys in multiple languages and cultures. And as she notes, rarely is there an isomorphism of words across languages. With a rating scale, for example, a translation, word for word, may create smaller or larger psychological gaps between points on the rating scale. For example, with a rating scale anchored by the extremes “allow” and “not allow,” in one of the language translations the extremes became “allow” and “forbid.” But the word “forbid” turned out to be considerably more negative than the extreme “not allow”.

Poorly translated assessment materials can have many consequences. Awkward or improper translations may make the test instruments easier or harder for students in some countries. In one recent international assessment, it was learned through self-report that test translators in one country had simplified the language in the mathematics assessment by one grade level to make it more understandable to students. The reading difficulty from the mathematics items had been removed to place the focus of these items on the assessment of mathematics skills only. The consequence was that the test items were easier in this country than they would have been had the reading difficulty of the test items not been removed. Cross-national comparisons of mathematics achievement for the country involved were no longer meaningful. Also, just plain bad translations may make the test instrument totally invalid. Literal translations are usually problematic (Hambleton, 2002)

## Translator quality problems

The translator has a vital role to play in the entire process of translating.

One of the common problems involves the selection of translators. The task of choosing translators seems straightforward enough. The characteristics of the quality translator are included in all translation guides of PISA, TIMSS and PIRLS. The technical literature suggests that at least three qualifications are necessary.

First, successful translations are carried out by persons who are knowledgeable about the subject matter. Technical knowledge on the part of translators is essential or the meaning of the source material can easily be lost in the translation (Brislin, 1970).

Second, successful translations are carried out by translators who have experiences in both languages. Experienced test translators such as Woodcock (1985) recommend however when translating from the source language to the target language should be dominant in the target language and have experiences in that culture. Otherwise, it is often very difficult to achieve a satisfactory translation. Hambleton (1993) declares that according to Woodcock, "Few persons, for whom the target language has been acquired later, will be as sensitive to the unique patterns of a language that, when present, makes a translation sound natural and not stilted."

Thirdly, test translations are done best by persons who have skills in test development, and know the principles of writing good test items. These skills are essential so that common errors in item writing do not enter during the translation process. For example, a translator not familiar with multiple-choice test item writing could introduce "clang" associations, unusually long correct answers, distractors that have the same meaning, awkward item stems, etc., that reduce the validity of the test items in the target language. Some of the errors could make the test items harder (e.g., awkward item stems) and other errors might make the test items easier (2.g., two or more distractors with the same meaning or value). The result, however, is the same: a non-equivalent test (Hambleton, 1993).

Deficient translator competences linked to *language skills, knowledge of the source language, subject matter knowledge and literary skills, translational knowledge and familiarity with test translation* were also one of the factors identified by the research

studies that has caused translation problems and non-equivalences in international achievement studies.

Behr (2018) introduces the multi-dimensional concept of translation competence. She mentions that translation is a multi-dimensional concept, that translation competence models and approaches are manifold. According to Behr translation competences have received heightened attention over the past two decades, driven in particular through empirical studies with increasingly sophisticated study set-ups Göpferich & Jääskeläinen in 2009.

Professional translators, and translation service providers in particular, are encouraged to follow the international standard ISO 17100 Translation services — Requirements for translation services (2015), which specifies “requirements for all aspects of the translation process directly affecting the quality and delivery of translation services” (p. vi). It lists the competences that a professional translator should have. These are (and are described in the next section in more detail): translation competence; linguistic and textual competence in both the target and the source language; competence in the domain; competence in research, information acquisition, and processing; cultural competence; and technical competence (Behr, 2018)

In practice this means above all that more care is needed when selecting and hiring translators (cf. e.g., Hambleton, 2001; 2005; Harkness, Villar & Edwards 2010). It may even be necessary to test them (for more suggestions for such tests, see e.g., European Social Survey [ESS], 2010).

According to Arffman, (2007), it is important to ensure that all translators have a good command of the target and source languages. If, moreover, the test is translated from two source languages, the decisions of the reconciler rest on a more solid and reliable basis, if s/he is proficient in both languages. Likewise, translators translating literary texts should have experience of literary translation. In practice this means that in reading tests, in particular, different translators may be needed to translate factual texts on different topics and literary texts. However, finding several translators who would have knowledge of a certain subject may sometimes be difficult, especially in a reading test. When this is the case, the problem may be at least partly alleviated if there are subject matter specialists in the translation team with whom the translators can discuss (e.g., ESS, 2010; Harkness, Villar & Edwards, 2010; see also Hambleton, 2001; 2005) – and if so much time



is allotted in the testing schedule to translation that translators have enough time to discuss with the specialists and possibly also to consult reference works (Livbjerg & Mees, 2002).

Care should also be taken that translators are well versed in the principles and strategies of translation. Translation of international achievement test differs from other types of translation, in that it requires knowledge not only of translation but also, for example, of testing, cognitive strategies and response processes (cf. Dept, Ferrari & Wäyrynen, 2010, p. 167). Therefore, in addition to being fully competent translators, translators translating international achievement tests also need to be well trained in test translation (see also ESS, 2010; Hambleton, 1994; 2002, p. 66; Harkness, Villar & Edwards, 2010).

### **Recommendations for translation problem elimination**

Hambleton (1993) presented 14 preliminary suggestions for guidelines. Six of them are related to translation problems.

- When it is anticipated or known that a test will be prepared in one language and translated into others, every effort should be made at the item writing stage to use straightforward directions, item stems, and answer choices. Test items with many details are more difficult to translate. Additional suggestions include the repetition of nouns rather than the use of pronouns, avoidance of metaphors, avoidance of the English passive tense (because it's more difficult to translate), and avoidance of hypothetical phrasings or subjunctive mood (Werner & Campbell, 1970). In the mathematics and science areas, for example, conventions about the use of time, money, and units of length, volume, and weight should be agreed upon at the outset. Conventions should insure test fairness for all examinees. (One convention might be to minimize the number of problems which require units.)
- The most useful design for establishing the equivalence of two versions of a test requires the source language monolinguals taking the source version and target language monolinguals taking the target version. However, the advantages of this design are lost if statistical

techniques are used which require the 24 26 questionable assumption of equal ability groups. Use conditional statistical techniques such as IRT or the Mantel-Haenszel procedure whenever possible.

- Whenever possible, both judgmental and statistical methods should be included in a study to determine the equivalence of an original and a translation of the original version of a test.
- A study of the factorial structures of multiple language versions of a test is valuable in judging the appropriateness of the test translations.
- Test translators should be chosen for their expertise in the source and target languages and their familiarity with the test content and their experiences in both cultures. (Normally, knowledge of both languages will not be sufficient to produce a satisfactory test translation.) The preferable situation is for test translators to be most familiar with the target language and culture. Knowledge of the principles of writing test items is valuable too.
- Whenever possible, and certainly for all large scale test translation projects, multiple judgmental and empirical methods should be used: For example, the process might include (a) training test developers in words, phrases, and concepts to avoid in item writing; (b) evaluation by test translators of the match between the source language and the back-translation of the test; (c) the use of bilingual translators to evaluate the similarity between the source and target language versions of the test; and (d) the collection of data using monolingual examinees taking each version of the test and then subjecting the data to an item bias analysis.

Solano-Flores (2006) proposed recommendations related to: (a) the characteristics of the individuals in charge of translating instruments; (b) the use of review, not simply at the end of the process, but during the process of test translation; (c) the minimum time needed for various translation review iterations to take place; and (d) the need for proper documentation of the entire process of test translation.

Taking everything into account, the study showed that to develop the translation work in international reading literacy studies, more weight needs to be put on linguistics and knowledge about translation than seems to be done today. Moreover, such knowledge is needed not only while actually translating the texts but also while selecting, producing and verifying them. (Affreman 2007)

In order to ensure the translation quality more research is needed in the future on translation problems.

## REFERENCES

1. Alderson, J. C. (2000). *Assessing reading*, (pp.86). Cambridge University Press. doi:10.1017/CBO9780511732935
2. Allalouf, A. (1999). Scoring and equating at the National Institute for Testing and Evaluation (*Research report 296*) Jerusalem: National Institute for Testing and Evaluation.
3. Allalouf, A. (2003). Revising translated differential item functioning items as a tool for improving cross-lingual assessment. *Applied Measurement in Education*, 16, 55–73.
4. Allalouf, A., Rapp, J., & Stoller, R. (2009). Which item types are better suited to the linking of verbal adapted tests. *International Journal of Testing*, 9, 92–107.
5. Arffman, I. (2007). The problem of equivalence in translating texts in international reading literacy studies. A text analytic study of three English and Finnish texts used in the PISA 2000 reading test (*Research Reports 21*). Jyväskylä: University of Jyväskylä, Institute for Educational Research.
6. Backhoff, E., Contreras-Niño, L. A., & Solano-Flores, G. (2011). *The theory of test translation error and the TIMSS and PISA international test comparisons*. Mexico: National Institute for Educational Evaluation [Sp.].
7. Baker, M. (1992). *In Other Words. A Coursebook on Translation*. London: Routledge.
8. Behr, D. Translating questionnaires for cross-national surveys: A description of a genre and its particularities based on the ISO 17100 categorization of translator competences. *The International Journal for Translation & Interpreting Research*. Vol. 10 No.2 (2018).DOI: 10.12807/ti.110202.2018.a02
9. Bell, R. T. (1991). *Translation and translating: Theory and practice*, 6 .London: Longman.
10. Brislin, R. (1986). The wording and translation of research instruments. In W. Lonner & J. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Beverly Hills: Sage.
11. Brislin, R.W. (1970). Back translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1, 185–216
12. Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

13. Budgell, G. R. (1992, May). *Analysis of differential item functioning in translated assessment instruments*. Paper presented at the Annual Conference of the Society of Industrial and Organizational Psychology, Montreal, Canada
14. Campbell, S. (1999). *A cognitive approach to source text: Difficulty in translation*. *Target* 11, 33–63.
15. Catford, J.C. (1965). *A Linguistic Theory of Translation*. London: Oxford University Press.
16. Chang, C. Y., & Myers, J. E. (2003). Cultural adaptation of the Wellness Evaluation of Lifestyle: An assessment challenge. *Measurement & Evaluation in Counseling & Development*, 35(4), 239-250.
17. Chesterman, A. (2010). Why study translation universals? In R. Hartama-Heinonen & P. Kukkonen (Eds.), *Kiasm (Acta Translatologica Helsingiensia)*. Vol. 1; pp. 38–48). Helsinki: University of Helsinki, Department of Finnish, Finno-Ugrian and Scandinavian Studies.
18. Danks, J., & Griffin, J. (1997). Reading and translation. A psycholinguistic perspective. In J. Danks, G. Shreve, S. Fountain & M. McBeath (Eds.), *Cognitive processes in translation and interpreting (Applied Psychology. Individual, Social, and Community Issues)*. Vol. 3; pp. 161–175). Thousand Oaks, CA: Sage.
19. Dept, S., Ferrari, A. & Wäyrynen, L. (2010). Developments in translation verification procedures in three multilingual assessments: A plea for an integrated translation and adaptation monitoring tool. In J. Harkness, M. Braun, B. Edwards, T. Johnson, L. Lyberg, P. Mohler, B.-E. Pennell & T. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 157–173). Wiley: Hoboken, NJ.
20. Dorer, B. (2011). Advance translation in the 5th round of the European Social Survey (ESS). *FORS Working Paper Series*, paper 2011–4. Lausanne: FORS.
21. Elosua, P., & López-Jauregui, A. (2007). Potential sources of differential item functioning in the adaptation of tests. *International Journal of Testing*, 7 (1), 39–52.
22. Ercikan, K. Disentangling Sources of Differential Item Functioning in Multilanguage Assessments. *International Journal of Testing*, 2(3&4), 199–215.
23. European Social Survey. (2010). *ESS round 5 translation guidelines*. Mannheim, European Social Survey GESIS.
24. Fegert, J. M. (1989). Bias factors in the translation of questionnaires and classification systems in international comparative child and adolescent psychiatric research. *European journal of Child & Adolescent Psychiatry: Acta Paedopsychiatrica*, 52(4), 279-286.
25. Fontanet, M. (2005). Temps de créativité en traduction. *Meta*, Volume 50, Issue 2, p. 432–447.
26. Geisinger, K. F. (2003). Testing and assessment in cross-cultural psychology. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Assessment psychology*, Vol. 10, pp. 95–117). John Wiley & Sons Inc.
27. Gierl, M. J., & Khaliq, S. (2001). Identifying sources of differential item and bundle functioning on translated achievement tests: A confirmatory analysis. *Journal of Educational Measurement*, 38 (2), 164–187.

28. Göpferich, S. & Jääskeläinen, R. Process research into the development of translation competence: where are we, and where do we need to go? Across Languages and Cultures 10 (2), pp. 169–191 (2009) DOI: 10.1556/Acr.10.2009.2.1 1585-1923
29. Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing*, 20(2), 225–240
30. Grisay, A. (2004). PISA 2000: Differences in item difficulty between English, French and German countries. Unpublished manuscript.
31. Hale, S., & Campbell, S. (2002). The interaction between text difficulty and translation accuracy. *Babel*, 48, 14–33.
32. Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. London: Falmer.
33. Hambleton, R. (2001). The next generation of ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*, 17, 164–172.
34. Hambleton, R. (2002). Adapting achievement tests into multiple languages for international assessments. In A. Porter & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 58–79). Washington: National Academy Press.
35. Hambleton, R. K. (1993). Translating achievement tests for use in cross-national studies. *European Journal of Psychological Assessment*, 9, 54-65.
36. Hambleton, R. K., Merenda, P., & Spielberger, C. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Hillsdale, NJ: Lawrence S. Erlbaum Publishers.
37. Harkness, J., Edwards, B., Hansen, S. E., Miller, D. and Villar, A. (2010), Designing Questionnaires for Multipopulation Research. In Harkness, J. et al (eds.), *Survey Methods in Multinational, Multiregional and Multicultural Contexts*. John Wiley & Sons, Hoboken, NJ, pp 33-57.
38. Harkness, J., Villar, A., & Edwards, B. (2010). Translation, adaptation and design. In J. Harkness, M. Braun, B. Edwards, T. Johnson, L. Lyberg, P. Mohler, B.-E. Pennell & T. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 117–140).
39. Harkness, J.A. and Schoua-Glusberg, A. (1998) Questionnaires in Translation. *ZUMA-Nachrichten Spezial*, 3, 87-127.
40. Hassan, B. A. (2011). *Literary translation: Aspects of pragmatic meaning*. Cambridge: Cambridge Scholars.
41. House, J. (1997). *Translation Quality Assessment: A Model Revisited*. Tübingen: Narr.
42. Jakobson, R. (1959/2000). On linguistics aspects of translation. In Venuti, L. (ed.) (2000), *The Translation Studies Reader*. London and New York: Routledge, 113-118.
43. Kim, R. (2006). Use of extra linguistic knowledge in translation. *Meta*, 51, 284-303.
44. Koller, W. (1979). Einführung in die Übersetzungswissenschaft. Heidelberg: Quelle and Meyer.
45. Kuhlwezak, P. (2003). The troubled identity of literary translation. In G. Anderman & M. Rogers (Eds.). *Translation today: Trends and perspectives* (pp. 112–124). Clevedon: Multilingual Matters.

46. Laviosa-Braithwaite, S. (1998). Universals of translation. In M. Baker (Ed.), *Routledge encyclopedia of translation studies* (pp. 288–291). London: Routledge.
47. Lefevere, A. (1992). *Translating literature. Practice and theory in a comparative literature context*. New York: The Modern Language Association of America
48. Livbjerg, I., & Mees, I. (2002). Problem-solving at different points in the translation process: Quantitative and qualitative data. In G. Hansen (Ed.), *Empirical translation studies: Process and product (Copenhagen Studies in Language 27)*, pp. 147–190). Copenhagen: Copenhagen Business School, Faculty of Modern Languages.
49. Lörscher, W. (1991). Translation performance, translation process, and translation strategies. A psycholinguistic investigation (*Language and Performance No. 4*). Tübingen: Narr.
50. Misra, G., Sahoo, F. M., & Puhan, B. N. (1997). Cultural Bias in testing: India. *European Review of Applied Psychology*, 47(4), 309-317.
51. Neubert, A., & Shreve, G. (1992). *Translation as text*. Kent, OH: Kent State University Press
52. Newmark, P. (1981). *Approaches to Translation*. Oxford and New York: Pergamon Press.
53. Newmark, P. (2003). No global communication without translation. In G. Anderman & M. Rogers (Eds.), *Translation today: Trends and perspectives* (pp. 55–67). Clevedon: Multilingual Matters.
54. Nida, E. (1964). *Towards a Science of Translating*. Leiden: E.J. Brill.
55. Nida, E. (1969). *Science of translation*. *Language* 45, 483-498.
56. Nida, E. and Taber, C.R. (1969). *The Theory and Practice of Translation*. Leiden: E.J. Brill.
57. Nord, C. (1991). *Text analysis in translation. Theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam: Rodopi.
58. Nord, C. (1997). A functional typology of translations. In A. Trosborg (Ed.), *Text typology and translation* (pp. 43–66). Amsterdam: Benjamins.
59. OECD. (2009a). *PISA 2006 technical report*. Paris: Author. Retrieved, June, 28, 2012, <http://www.oecd.org/dataoecd/0/47/42025182.pdf>.
60. OECD. (2009b). *PISA 2009 assessment framework: Key competencies in reading, mathematics and science*. Paris: Author. Retrieved, June, 28, 2012, <http://www.oecd.org/dataoecd/47/23/41943106.pdf>.
61. Pym, A. (2008). On Toury's law of how translators translate. In A. Pym, M. Schlesinger & D. Simeoni (Eds.), *Beyond descriptive translation studies: Investigations in homage to Gideon Toury* (pp. 311–328). Amsterdam: Benjamins.
62. Pym, A. (2010). *Exploring Translation Theories* (pp. 324). London and New York: Routledge.
63. Rosenblatt, L. (1994). The transactional theory of reading and writing. In R. Ruddell, M. Ruddell & H. Singer (Eds.), *Theoretical models and processes of reading* (4th ed., pp. 1057– 1092). Newark, DE: International Reading Association.
64. Scarpa, F. (2002). Closer and closer apart? Specialized translation in a cognitive perspective. In A. Riccardi (Ed.) *Translation studies. Perspectives on an*

- emerging discipline* (pp. 133–149). Cambridge, MA: Cambridge University Press.
65. Sireci, G., & Bastari, B. (1998, August). *Evaluating construct equivalence across adapted tests*. Paper presented at the Annual Meeting of the American Psychological Association, San Francisco, CA
66. Solano-Flores, G., Backhoff, E., & Contreras-Niño, L. (2009). Theory of test translation error. *International Journal of Testing*, 9, 78–91.
67. Solano-Flores, G., Backhoff, E., & Contreras-Niño, L. A. (2006). *Methodology for evaluating the quality of test translations in international test comparisons: The case of Mexico, TIMSS-1995*. [Sp.]. Mexico: National Institute for Educational Evaluation (INEE).
68. Solano-Flores, G., Contreras-Niño, L. A., & Backhoff, E. (2005, April 12–14). *The Mexican translation of TIMSS-95: Test translation lessons from a post-mortem study*. Paper presented at the 2005 annual meeting of the National Council on Measurement in Education. Montreal, Quebec, Canada.
69. Solano-Flores, G., Contreras-Niño, L. A., & Backhoff, E. (2006). Test translation and adaptation: Lessons learned and recommendations for countries participating in TIMSS, PISA, and other international comparisons. *REDIE: Electronic Journal of Educational Research*, 8(2). [Sp.] <http://redie.uabc.mx/vol8no2/contents-solano2.html>
70. Solano-Flores, G., Contreras-Niño, L.A., & Backhoff, E. (2013). The measurement of translation error in PISA-2006 items: An application of the theory of test translation error. In M. Prenzel, M. Kobarg, K. Schöps, & S. Rönnebeck (Eds.), *Research on the PISA research conference 2009* (pp. 71-85). Heidelberg: Springer Verl
71. Tanzer, N. K. (1998). Developing tests for use in multiple languages and cultures: A plea for simultaneous development. In R. Hambleton, P. Merenda, & C. D. Spielberger (Eds.), *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Hillsdale, NJ: Lawrence Erlbaum.
72. Vinay, J. P., and Darbelnet, J. (1995). *Comparative Stylistics of French and English: A Methodology for Translation*. North America, Amsterdam and Philadelphia: John Benjamins
73. Vinay, J.P. and Darbelnet, J. (1958). *Stylistique Comparée du Français et de l'Anglais: Méthode de Traduction*. Paris: Didier. (Transl. and ed. by Sager, J.C. and Hamel, M.J. (1995) as *Comparative Stylistics of French and English: A Methodology for Translation*. Amsterdam and Philadelphia: John Benjamins.)
74. Werner, O., & Campbell, D. T. (1970). Translation, wording through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A handbook of method in cultural anthropology*, (pp. 398–420). New York: Natural History Press.
75. Wilss, W. (1990). Cognitive aspects of the translation process. *Language & Communication* 10, (1), 19–36.
76. Woodcock, R. W. (1985). Woodcock Language Proficiency Battery. Spanish Form: Technical Summary (Assessment Service Bulletin, Number 9). Allen, TX: DLM.