The translator's perspective on translation quality control processes for international large-scale assessment studies

Britta Upsing
Deutsches Institut für Internationale Pädagogische Forschung, Germany
upsing@dipf.de

Marc Rittberger
Deutsches Institut für Internationale Pädagogische Forschung, Germany
rittberger@dipf.de

Abstract: International Large-Scale Assessment studies (iLSAs) like PISA (the Programme for International Student Assessment) and PIAAC (the Programme for the International Assessment of Adult Competencies) employ stringent quality control procedures for the translation of their test items. The translations are crucial: a test item should not become more or less difficult because of its translation, given that test results are used to assess and compare the competency levels of different populations across countries. This article discusses how PISA and PIAAC translation procedures have evolved from earlier translation quality processes, and the underlying assumptions about translation that have shaped this evolution. We then report on findings from a qualitative interview study with translators, reviewers, and translation managers who have been involved in PISA or PIAAC translation processes. The objective of the interview study is to analyse the quality control procedures from the perspective of translation players. How do translators prepare for and perform translations for iLSA studies, and how does this process compare with other translation assignments they receive? This comparison will give an understanding of whether translators, in general, believe the iLSA translation process and its quality control procedures provide adequate guidance for performing these translations. We finish by proposing recommendations for future iLSA translation processes.

Keywords: international large-scale assessment studies, test translation, questionnaire translation, qualitative interview study

1. Introduction

In 2006 a sudden shift in public opinion seemed to have occurred in Vietnam: one-third of respondents to an international survey favoured military rule when the previous survey had indicated 99 percent of the population was in favor (Kurzman, 2014). As Kurzman, a professor of sociology, argued in a Washington Post article, this shift had nothing to do with sociological or demographic changes in Vietnam: instead, he thinks that the 2006 survey had switched the wording of its Vietnamese translations. The corresponding source questions, which were in English, had not been changed.

The validity of cross-national surveys or tests may be at stake when the translated versions of questionnaire or test items contain errors or (slight) deviations. The example also posits that valid comparisons between countries (Vietnam and the United States, for instance) are only possible if the test instruments measure the same constructs across languages and countries.

For international Large-Scale Assessment studies (iLSAs), stringent translation processes have been set up to avoid mistranslation. The hope is that these processes will result in good translation output as well as improved cross-country comparability of test results.

Two prominent examples of iLSAs that use such translation procedures are PISA (the Programme for International Student Assessment) and PIAAC (the Programme for the International Assessment of Adult Competencies), both commissioned by the Organisation for Economic Co-operation and Development (OECD). PIAAC and PISA tests are used for "monitoring the achievement level in a particular population, for comparing assessed (sub)populations", and may therefore "form the basis for developing and/or revising educational policies" (Upsing, Gissler, Goldhammer, Rölke, & Ferrari, 2011, p. 45). These tests are used to compare the literacy levels of adolescents or adults, for example, across multiple countries. Questionnaires within these studies provide socio-economic background data of test takers and thus possible explanations for particular performance patterns in the test (see Figure 1 for an example of a test item).

Figure 1: Sample test item from the PIAAC literacy test (the stimulus text must be read to answer the question) (Source: Organisation for Economic Co-operation and Development (OECD), 2013a, p. 1)

The OECD, which assigns the task of designing a PISA or PIACC study to an international consortium, is responsible for the iLSA overall. The international consortium creates the PISA or PIAAC test and questionnaire items and the accompanying documents, manages the overall study, and establishes the quality control procedures and standards. Participating countries set up their own national centres in charge of implementing the study in their own country according to the standards set by the consortium. They are also responsible for creating the translation (Adams, 2002, pp. 15–18; Kirsch & Thorn, 2013, pp. 9–11). PISA studies take place every three years and the PIAAC study has only been through one survey round so far.

Before the same test can be administered in different countries and languages, both the accompanying questionnaire and the competency test items require translation. The main goal of the translation process for test items is that "a person of the same ability will have the same probability of answering any assessment item successfully independent of his or her linguistic or cultural background" (Thorn, 2009, p. 9). Therefore, when a test is translated, the test should not become easier or harder to respond to because of its translation.

In this article, the main focus will be on the translation quality control procedures that were developed for iLSA studies, and in particular those that were developed for PISA and PIAAC. For these studies, procedures were set up to support translators, check their work, and ensure optimal cross-language comparability of translations. This was deemed necessary because most translators are not familiar with test item translation and may not understand which factors might unduly impact the translated version of an item.

Despite these efforts, iLSA translations have been criticized. On the one hand, the quality of the translations themselves or the translation process has been questioned (Arffman, 2012a, 2012b; Dolin, 2007; Ercikan, 1998; Karg, 2005; Wuttke, 2007). On the other hand, unexpected statistical differences between language versions of tests have been pointed out (Eivers, 2010, p. 102; Grisay, Gonzalez, & Monseur, 2009, p. 80). So doubts remain about what stringent translation processes can achieve and how they should be set up. This article will analyse the latter question by taking into account the translator's perspective as an additional aspect. The focus here is to understand how translators work, and if the conditions they encounter when translating for iLSA are beneficial for them.

To find out whether these quality control procedures actually support translators, translation managers, and reviewers, and to explore how the quality control procedures function, the following points are discussed in this article:

- The evolution of translation procedures for iLSA studies.
- The underlying assumptions of the translation process.
- The results of a qualitative interview study including interviews with translators, translation managers, and reviewers.

The qualitative study presented herein will both provide a broad understanding of how translators prepare for and perform translations for iLSA studies and compare iLSA translations to other translation assignments. This will give a fundamental understanding of whether translators, in general, believe the iLSA translation process and its quality control procedures provide adequate guidance for performing these translations.

In sum: the goal of this article is to understand how translators can excel when asked to translate test items for iLSA. The results can be used to give recommendations for future translation processes for iLSA studies.

2. Translation quality control processes for iLSA

Translation quality has been identified as vital to the validity of study results, as has gradually been recognized in the field of testing and survey methodology (see McKay et al., 1996, p. 93). Therefore, the choice of translation and quality control procedures is vital as well.

2.1 First developments: Back translation

Back translation was one of the first quality control procedures developed for survey and test translation. It is based on procedures described by Brislin (1970). During back translation, the source version of a test is translated into the target language by one translator, then that version is back translated into the source language by a second translator. The two source language versions of the test (the original source version and the back-translated source version) are then compared and inconsistencies or errors are identified (Harkness & Schoua-Glusberg, 1998, p. 111). The expectation with back translations is that "if the adaptation has been effective, the back-adapted version of the test should look very much like the original" (Hambleton & Patsula, 1999, p. 6).

Back translation may appear to be an obvious choice of quality control procedures for (monolingual?) test developers: when items are developed in the source language, the extensive review and pre-testing process involves a large investment of time and effort before test finalization and before translation. According to Behr (2009, pp. 63–64), back translation may be popular because – via the back-translated version – it seemingly gives access to (and control of) the content of a translation to test developers with no command of the corresponding target language. Brislin (1970, p. 213) warns against over-reliance on back translation as it does not detect grammatical or orthographical errors in the target text. In addition, back translation gives no indication of the comprehensibility of the target text. In other words, a literal translation may result in a "good" back translation while at the same time be incomprehensible to a target language recipient. Furthermore, empirical studies by Behr (2017) and Ozolins (2009) demonstrate that back translation produces a high number of false-positive errors and does not detect errors that may be revealed by a thorough review of the target text itself. Back translation is also criticized for its "drive for equivalence [which may result in an] underlying positivism which treats language as the neutral transmission of messages" (Chidlow, Plakoyiannaki, & Welch, 2014, p. 566). The underlying assumptions of back translation itself are considered out-of-date (Colina, Marrone, Ingram, & Sánchez, 2016, p. 9):

> The use of BT [back translation] as a quality control method for translation in crosscultural research highlights a lay view of translation, based on the notions prevalent in the 1970s, such as meaning as an objective reality and equivalence (at the conceptual or semantic level) as the ultimate measure of translation quality. These notions remain mostly unchallenged in crosscultural research [...] even into the first decade of the 21st century. (Colina et al., 2016, p. 9)

The back translation procedure still enjoys widespread popularity today, for example for the translation of surveys in health research (also see Colina et al., 2016). According to Grisay (2002, p. 59), back translation was used in iLSA translations until the early 1990s.

2.2 Moving ahead: Forward translation

Some of the underlying assumptions of back translation have been perpetuated into today's translation quality control procedures. About 25 years ago, the first systematic approach regarding translation procedures was initiated within the context of the *Third International Mathematics and Science Study* (TIMSS), which measures student competencies in mathematics and science. In 1993, Hambleton (1993), a psychometrician, was asked to evaluate how to create valid cross-cultural tests and to make recommendations for, among other processes, the translation process for assessment items. The translation process of TIMSS items was based on these recommendations (Maxwell, 1996, p. 1) and includes the following measures or steps (Maxwell, 1996, pp. 7–8):

- Fulfilling a list of minimum requirements for competencies of translators and proofreaders
- Compliance with guidance documents, including general and item-specific guidelines for translators and proofreaders
- Following a translation process that employs "multiple forward translations" (described below)
- Verifying by reviewing the translation text, comparing it to the source text, checking compliance with guidance documents, and providing suggestions for corrections
- Allowing for acceptance or rejection of corrections by the translator
- Checking the document to which corrections were implemented

The "multiple forward translations" step entails the following tasks: First, two translators – independently from each other – translate the source text into the same target language. Second, personnel at the national centre read both translations and consolidate them into one text (this is later called the "reconciler" role). For TIMSS, no information is given on whether this step was done in a team or by only one person. This twice-performed *forward translation* procedure was developed to replace *back translation*.

The assumption underlying the approach of multiple forward translations, as provided by the TIMSS developers who suggested the above measures, was that most items would be translated "the same" by two independent translators (Maxwell, 1996, p. 3). Furthermore, the developers offered a solution for when the translations differed: "the differences would be discussed, and the best translation of the item selected for the test instrument" (Maxwell, 1996, p. 3). Despite these foreseen discussions over translation differences, when combining two sets of text, a reconciler faces a challenging and often error-prone task, especially given that independent translations of single items are frequently not identical. The assumption that two translators will, for the most part, translate items equally, is an unreasonable one. The helpfulness of having two interpretations of a source text may be negated by the extra effort needed to reconcile the combination of the two versions.

The first PISA study in 2000 used quality control procedures based on the TIMSS study (Grisay, 2002, pp. 58–59), which included the step "multiple forward translations" (here called a "double translation"). For PISA 2000, reconciliation was supposed to be done "by a third person" (Grisay, 2002, p. 58), not by a team. Translators of the 2000 PISA study could choose to base their (entire) translation on either an English or French version of the items (the consortium created the French translations from the English text before the national centres started their work) (Grisay, 2002, p. 57). Because text was translated from either an English or French source, three processes were possible when the test items were translated:

1) a double translation from one source language (English)
2) a double translation from both languages (English and French) or
3) a single translation from English (with or without crosschecks with the French version).

Grisay (2002, pp. 67–69), as a member of the PISA 2000 consortium, used the PISA 2000 data to evaluate which of the three approaches was most successful in avoiding translation errors. Her results indicated the most successful approaches were double translating from both English and French or single translating from English with a crosscheck against the French version. Double translation from English and single translation from English without crosschecks against the French version, were less successful processes, according to Grisay. Therefore, double translation may not be as beneficent in providing a successful translation for result comparability as previously thought.

2.3 Application of the translation process to iLSAs
The translation procedures in iLSA commissioned by the OECD have evolved over the years and are built on experiences from previous studies (OECD, 2002, 2005, 2009, 2012, 2013b, 2014, 2017). The following elements are typically included, and will be described in more detail below (further improvements have been made but will not be discussed in this article):

1) A general guideline document is provided that explains the peculiarities of test translation.
2) The source text is supplemented by item-specific guidelines.
3) A monitoring sheet documents the source text, any item-specific guidelines, and records comments and changes to the translation.

4) The translation process utilizes at least the following steps: Double translation, reconciliation, verification, discussion, and final check.

The *general guideline document* (1) serves as an introduction to test and survey translation for all personnel involved in the process (survey experts in national centres, translators, reviewers). It provides a list of common translation errors, and outlines possible adaptations (for example whether names can be localized or not) (see PISA Consortium, 2010, for an example). The source language test items are supplemented with *item-specific guidelines* (2), which also refer back to the possible adaptations outlined in the general guideline document. Item-specific guidelines consist of

> […] advice covering recommended, allowed or proscribed adaptations, literal or synonymous matches to be maintained, other psychometric characteristics to be considered (e.g. relative length or other patterns in multiple choice responses), desirable register of terms to be maintained, emphasis to be echoed, tips for the translation of difficult or idiomatic terms, etc. (OECD, 2014, p. 95)

In practice, this means that the source text contains comments on individual text elements. These comments explain difficult English expressions or idioms, require translators (if possible) to use the same translated expression in the question (or item) as in the stimulus (i.e., the material the test person works with to answer the question), or ask translators to use a local name instead of the English name.

For instance, in PISA 2009 the following item-specific guidelines were given for the translation of certain source text elements (OECD, 2012, p. 89):

> Fly by Night -> Even if a published version of this book exists in your language, translate 'Fly by Night' as something meaning 'Flying at night'. Do not use the translation in your language of the title of 'Vol de nuit' by Antoine de St Exupéry
>
> Mosca, Clent, Frances Harding -> Do not adapt
>
> Sam, Stephanie -> Adapt to common names in your languages
>
> Loose ends 'tied up' -> Use an idiom that connotes resolving details that were previously unresolved / unexplained
>
> Blurb -> Definition of 'blurb': a short piece of writing that praises and promotes something, especially a paragraph on the cover of a book

These guidelines are supposed to ensure that translators across languages find comparable solutions for difficult translation decisions by drawing the their attention to possible translation problems (Ferrari, Wäyrynen, Behr, & Zabal, 2013, p. 10).

Since PISA 2003, this item-specific information has been integrated into a monitoring sheet (3), which also includes the source text (See Figure 2).

There is one monitoring sheet for each test item. It contains the source language text and its corresponding guidelines. It is used as a reference document during the translation process. The translation itself is either made directly in this document or in a Word document (for tests administered on paper) or in Extensible Markup Language (XML: For tests administered on a computer).

VERIFICATION FOLLOW-UP FORM
COMPUTER-BASED

PIAAC
FIELD TRIAL 2009
Country: PT   UNIT: Election Results   PIAAC ID: C302BC02   ALL ID: COREQ2S1
Target language: pt

PLEASE INSERT NEW LINES, IF NEEDED, TO DOCUMENT ADDITIONAL ISSUES

LOCATION	ENGLISH SOURCE	PROPOSED TARGET VERSION	CONSORTIUM RECOMMENDATION	NPM COMMENT	V INTE
stimulus	Nationwide Manufacturing Company Union Council ELECTION RESULTS		Note: 'Union' is to be understood as trade union, i.e. an organization representing workers		
stimulus	Posting Date: June 22, 2000		Eliminate ', 2000' versus ALL version		
stimulus	The election of a new member of the Union Council for election group 3, at the Carver plant took place on June 21, 2000.		The name 'Carver' may be changed. Note: 'plant' means here 'factory'. Eliminate ', 2000' versus ALL version.		
stimulus	The results of the election were as follows:				
stimulus	Candidates Number of votes A. Greer 120 votes H.A. Holliday 80 votes G. F. Reynolds 29 votes		Names of people may be changed. Keep the three numbers aligned over each other.		
stimulus	Consequently Mr. A. Greer was formally elected as a member of the Union Council for Nationwide Manufacturing Company.		If name 'A. Greer' is changed, change it here too		
stimulus	In accordance with article 16, paragraph 1 of the Union Council bylaws, any interested party may lodge a complaint with the council within one week after publication of these results.				

Figure 2: Sample monitoring sheet with item-specific guidelines (here called "consortium recommendations") (Source: Ferrari et al., 2013, p. 3)

At minimum, the *translation process* (4) contains the same workflow steps as in the TIMSS study. Despite Grisay's findings, double translation has become a requirement in the PISA studies after 2000, and it was also recommended for the PIAAC study. The steps double translation and reconciliation are done by the national centre and their work is checked centrally by the consortium to ensure that similar translation decisions are made across language versions. Furthermore, the verification step aids the consortium in checking whether the national centres created a translation that is in line with consortium requirements (as specified in the general and item-specific guidelines). The consortium's corrections, comments, and suggestions are checked by the national centre. The consortium's final check is used to ensure that all important corrections have been implemented by the country. This process is also documented in the monitoring sheet.

NPM COMMENT	VERIFIER INTERVENTION	SEVERITY CODE	VERIFIER COMMENT	DI
	Missing Info	2	'and places' missing from translation.	
	OK			

Figure 3: The monitoring sheets contains columns for comments by the national centre (NPM), the consortium (verifier), discussion, and final check (last two steps not included in figure) (Source: Ferrari et al., 2013, p. 13)

These monitoring sheets therefore also enable communication between the consortium and national centres and between the different players involved in the translation process.

2.4 Evaluation of current procedures
Some of the measures that were first introduced for TIMSS and then further evolved in OECD studies correspond with recent translation quality control procedures, such as those formulated in the ISO 17100[1]. The

---

[1] The data from DIN EN 17100:2016-05 has been reproduced with the permission of DIN

recommendations in TIMSS, PISA, and PIAAC specify that translators should have native command of the target language as well as subject-matter knowledge. The ISO standard is more detailed though, since it further differentiates between translation competence, linguistic and textual competence, competence in research, cultural, technical, and domain competence (ISO 17100, 2015, p. 11). Furthermore, in helping to identify eligible translators, translators are required to meet one of the following criteria:

    a)   a recognized graduate qualification in translation from an institution of higher education;

    b)   a recognized graduate qualification in any other field from an institution of higher education plus two years of full-time professional experience in translating;

    c)   five years of full-time professional experience in translating. (ISO 17100, 2015, p. 6)

Revision (or verification) of the target text by a second translator is one of the key elements of the ISO standard, and this step is also included in the PIAAC and PISA recommendations. By introducing recommended translator qualifications and the step "verification" into iLSA, significant progress has been made over earlier approaches (e.g., back translation).

Some key elements of the quality control procedures are criticized, however. Arffman (2007, pp. 274–275) states that the general and item-specific guidelines emphasize lexical and formal similarities between source and target texts and prescribe too many translation decisions. She suggests that instead of giving detailed item-specific guidelines, the general guidelines document should explain what the exact translation goal is and how to attain it (Arffman, 2013, p. 5).

The results of her exploratory interview study indicate that double translation with reconciliation by a third translator is considered difficult and inefficient (Arffman, 2012a, pp. 58–62). In surveys like the ESS (European Social Survey), the double translation approach is used, but with a recommended reconciliation step completed by a team of translators and survey experts instead of by an individual (European Social Survey [ESS], 2016). Similarly, Nida, one of the founders of modern translation theory, advocated for team discussions of Bible translations, another field in which – as in survey and test translation – combined expertise from different fields is preferred for the translation (Nida & Taber, 1969, pp. 174–188).

In an attempt to remove another pitfall of test translation processes, Bolaños-Medina and González-Ruíz (2012) support collaboration between linguistic and test experts, which prevents non-translators from seeing translation as a "one way word-by-word-substitution procedure" (Bolaños-Medina & González-Ruíz, 2012, p. 725). This substitution procedure is an assumption originally underlying both the back translation and double translation approaches.

The research done for the present study is supposed to further shed light on test translation and its evaluation, and will be introduced in the next section.

3. Method

In 2016, 20 in-depth qualitative interviews were conducted with translators, project managers, and reviewers with the goal of learning how translators compare translating for iLSA with other translation jobs they receive – and ultimately how these players perceive their work (also see Upsing, 2017).

The average length of an interview was one hour. The interviews were recorded and transcribed. The analysis of the interviews was done using qualitative content analysis (see Kuckartz, 2014 or Mayring, 2010). All interviews were anonymized.

3.1 Interview participants

Interviewees consisted of individuals holding the following roles in iLSAs: national project manager, general project manager, translator, verifier/ reviewer, and survey expert. National project managers are survey experts who work for the national centre and organize the translation process in their respective country. General project managers work for the consortium and provide information and tools to ensure that all other participants in the translation process can perform their work. Translators are employed by national project managers to create the first version of the translation of the source text. Verifiers/reviewers are employed by the consortium and responsible for checking the text created by translators.

Of the 20 interviewees, 13 had experience with roles other than their "main role"; 12 interviewees identified themselves as professional translators, and four (B01, B02, B15, B16) had no experience with iLSA, but were included due to their experience as professional translators within large-scale translation processes.

All interviewees with experience with iLSA have either worked for PISA and/or PIAAC.

Table 1: Interviewee characteristics

	Main role	Other roles
B01	Translator	Reviewer, Project Manager
B02	Translator	Reviewer, Project Manager
B03	Project Manager (national)	Survey expert
B04	Project Manager	
B05	Project Manager (national)	Survey expert
B06	Project Manager (national)	Survey expert
B07	Project Manager	
B08	Project Manager	Verifier/Reviewer
B09	Project Manager	Verifier/Reviewer
B10	Project Manager	Verifier/Reviewer
B11	Verifier/Reviewer	Translator
B12	Verifier/Reviewer	Translator
B13	Translator	
B14	Translator	
B15	Translator	
B16	Translator	
B17	Project Manager	Translator
B18	Verifier/Reviewer	Translator
B19	Verifier/Reviewer	Translator
B20	Translator	

3.2 Interview questions

The semi-structured qualitative interviews addressed three main topics: Information provided for a translation job; process involvement (feedback, meetings, double translation); and software use and attitude towards technology. Interviewees were asked to compare and contrast the different experiences they had encountered in their work life, and to reflect on advantages and disadvantages of these experiences. This article will focus on the first two topics and on the role of translators. Here, the main questions that were addressed in the interviews with translators were the following:

Information and translation decisions:
- When you worked for your last big translation project: What information or help did you receive to be able to fulfil your task?
- Did this setup differ from the setup in other translation projects you have worked on?
- What kind of information do you like to receive?
- What is too much / too little information for you?

Item-specific guidelines:
- Which aspects do you like / do you not like about item-specific guidelines?
- What impact do you think these guidelines have on your translation decisions?

Feedback:
- What kind of feedback do you receive for your translation work?
- Does this setup differ from the feedback you receive for iLSA translation?

Group reconciliations, meetings, and trainings:
- With whom do you exchange information / discuss (translation) problems in a translation project?
- Does this setup differ from information exchange in iLSA projects?

4. Results

The results presented pertain to four distinct parts of the translation process: Information and translation decisions, item-specific guidelines, feedback, and group reconciliations and meetings.

4.1 On information and translation decisions

All interviewed translators reported that the following two questions precede their actual translation process: 1) What purpose is served by the translation? and 2) How is the translation going to be used? The answers to each of these questions are important because the purpose of the translation, as well as the text type, have an impact on how the translator translates, and on how he or she tackles translation difficulties or makes translation decisions. A simple example helps to illustrate this point: The interviewed translators reported varied approaches to their routine translation work regarding the use of synonyms in their translations (for example, for stylistic reasons or for consistency).

Table 2: Synonyms and consistency[2] in translations according to interviewees

	"I never use synonyms"	"I know translators who use synonyms (when they shouldn't)"	"I use synonyms for stylistic reasons"	Should depend on text type / translation purpose
B01	X			
B02			X	
B03		X		
B07		X		
B13	X	X		X
B14				X
B15	X			X
B17		X	X	X

The answers ranged from "I never use synonyms" (B01:135[3]), to explaining that "experienced translators" do not use synonyms when the source text is consistent (B13:112), to not seeing a problem with using "synonyms for stylistic reasons" (B02:146). B15 explained that for her work she never uses synonyms in a translation, but that for literary text the approach might be different (B15:90). B17 also reported that she understands the tendency of translators to make the target text more interesting by using synonyms (B17:140). B14 thought that this decision depends on the text type. The answers that the translators gave are shaped by their previous translation experiences: Technical or legal translators favour consistency whereas translators who work in areas like marketing see no problem with text variety. B13 described how previous experiences and professional specialization on certain text types and translation tasks can subconsciously be transferred onto unknown or new tasks. She related an account of a technical translation on which she worked with another translator who usually translates marketing material and literary texts, and favoured using synonyms since in his opinion their use made for a better text. Ultimately she felt it necessary to go through the entire text and make sure it was consistent. She thought that the "work one is used to as a translator" shapes the way one translates (B13:104). She explained that when a translator wants to write an interesting and entertaining text, he or she should only do so when this is in line with the translation assignment and the purpose of the translation. Still, whether or not such a decision is made seems to depend on a translator's previous experience (this estimation is shared by the translation managers: B03:262, B07:128, B17:144).

In summary, according to interviewees, decisions regarding consistency should ideally be in line with the translation assignment in question. When asked about the kind of information they would receive with a translation job, the interviewed translators indicated that they are not usually given explicit information about its end purpose. Instead, they receive translation memories, parallel text, texts that show the larger context of their work or no information at all. Interviewees also explained that they do not necessarily need explicit information from their client to understand what is expected of them. If they do not receive explicit information, then the text type or previous experiences with a client help them to understand the (implicit) purpose of their translation. B01, B11, B12, B13, B14, and B17 all described situations in which they did not receive any additional information or reference material

---

[2] When translating literacy test items, the question of consistency between words in a question and its corresponding stimulus is critical.
[3] References point to the quoted line in the transcribed interview.

apart from the source text, but in which it was still clear to them what was expected of them. B13 explained that for her it is obvious that a manual will always be published with the product. B14, who regularly translates scientific books or articles, already knows the target market and that these translations are intended for publication.

Four of the twelve interviewed translators were employees (B02, B15, B16, B17). These translators reported that they have had years to learn about their company's expectations regarding their work. Eight translators were freelancers (B01, B11, B12, B13, B14, B17, B19, B20). All but one of these translators reported that they have a small client base which makes up most of their work. The one translator who had a wider client base translates books and scientific articles, which for her means that the scope and purpose of her work is well-defined but originates from a larger number of individuals. All interviewed translators reported that working for a small circle of main clients has the advantage of enabling them to anticipate what typifies the texts and expectations of their clients (also see B01 or B12):

> B11:92: [...] I always work for one particular thing. And I'm the only one doing it. [...] And so I get, I know the context. [...] because I have already translated a lot of things for them. //mhm// I tend to keep things, so I can always go back and look [...].

Thus – according to the interviewed translators – this familiarity could be one of the reasons why it is often not problematic for them when the intended use and aim of a translation is unknown: most of them can infer this information from their previous experience.

4.2 On item-specific guidelines

In cases where no prior experience with a text type existed, or the client was unknown, the interviewed translators indicated that the lack of information and guidance was problematic. Interviewees B01 and B15 described missing information as preventing them from understanding the aim of the translation, which made them place less confidence in their translation decisions.

Only one interviewed translator reported that she had translated for more than two iLSA studies. Therefore, it is possible that translating test items is an uncommon assignment for translators. The PISA/PIAAC consortium has tried to make up for this lack of experience by providing general and item-specific guidelines. Interviewees have only encountered such general or item-specific guidelines when translating (or verifying) test or questionnaire items. Therefore, this type of information seems rather unusual.

From the translators interviewed, those who used item-specific guidelines for *verifying* (B11, B12, B18, B19) saw these as positive and would not want to modify them; the guidelines served them as a check-list for comparing the source and target texts. Translators who worked with item-specific guidelines for *translating*, on the other hand, showed a more mixed evaluation of them. Interviewee B14 explained that she understands that the item-specific guidelines help to create equivalent target language versions across languages (B14:94), whereas Interviewee B13 pointed out that she often asked herself what translators are needed for if every translation decision is prescribed (B13:94). She also explained that she felt like these guidelines cut off her creativity (B13:126). The guidelines gave her the feeling that she was just supposed to render the semantic meaning of the English source text in the target language, even though the wording would sound strange to a target language recipient (B13:212). She also wondered why the "translation needs to be so literal". This feeling was expressed by other translators as well. For example, Interviewee B20 also interpreted the guidelines as quite restrictive on her creativity (B20:118), but did perceive them as helpful since in her

opinion they reduce ambiguity. She specified that, when a guideline called for using exactly the same word again, she would "not have to think about using a different word" (B20:106). Three of the general translation managers interviewed (B03, B04, and B17) voiced their fear that translators might over-interpret the item-specific guidelines. As B03 explained, it is important that translators "keep thinking" and not rely on using the same word again in the target text as in the source text. For example, the word 'work' might be rendered differently in different target languages depending on the context (B03:258), a point that is also stressed by Interviewee B04 (B04:97).

There seems to be the danger of interpreting item-specific guidelines as a list of requirements instead of a list of suggestions. Translators who lack experience with test translation may feel obliged to fulfil guidelines that they perceive as required, even if they do not think they make sense. Generally, instead of relying too much on item-specific guidelines, it might be better to teach translators about similarities and differences of test translation with other text types, so that translators feel more secure about using their own translation competence.

4.3 On feedback

Generally, the interviewed translators receive varied feedback from clients. Four interviewed translators (B01, B11, B15, and B20) had been confronted with translation jobs for which no feedback was given, possibly because no quality control procedures were in place that required that it be given (see for example B20:123 or B01:24). Interviewee B19 reported that sometimes feedback is provided by non-qualified personnel who happen to identify non-errors. All interviewed translators (and general project managers) agreed that feedback is important, and one translator had the opinion that "good [translation] agencies give feedback" (B17:146). According to her, feedback is especially helpful if a translator continues to work with a client and can identify potential problems and fix them (B17:146). Interviewee B02, a manager and translator, regularly organizes her own quality control procedures for translations. She first assigns a translation job to a professional translator and then reviews the translation herself. Afterwards she assigns a review to a subject-matter expert. In the end, all feedback is processed by the original translator and the final product is checked again by Interviewee B02.

In the case of iLSA, Interviewee B20 explained that translators are not necessarily informed about corrections and changes in their translations. These processes are organized by the national centres and not subject to the consortium's control. This behaviour is problematic for two reasons. Firstly, translators have no chance to learn from their mistakes. Secondly, the translator's input might be relevant. Interviewee B06, a psychometrician working for a national centre, does provide feedback to her translators after verification by the consortium. She explained that after discussing feedback with the original translators, "90 percent of the cases [where she suggested a correction] the translator convinced me that her choice was the best" (B06:62).

As explained in one of the previous sections, it is up to the national centres to implement the verification feedback from the consortium. Based on the interviews, translators who also work as verifiers showed mixed feelings regarding this feedback process. Three of the four verifiers who were interviewed believed no problems existed with the process. Interviewee B11, however, has seen that errors she identified in an earlier survey were repeated and not corrected in the next round (B11:140). Interviewee B09 thinks it is important that explanations be given in situations where suggested changes (i.e. the feedback) are not accepted. The reason given was that it is otherwise not clear whether the lack of feedback incorporation was intentional or accidental (B09:06).

4.4 On group reconciliations, meetings, and training sessions

Some translators received feedback regarding their translation because their national centre modified the recommended translation procedure: instead of reconciliation by a third translator, they were part in a team reconciliation meeting. Six out of 20 interviewees have been involved as translators or survey experts in one or more of these meetings (B03, B04, B13, B14, B17, B19); all six gave a positive evaluation. For the translators interviewed, receiving direct feedback on their translations, exchanging views and working together on equal footing with experts of different backgrounds in order to create a translation is a rare experience and a welcome change to daily routines (B13, B14, and B19). B13 explained that translators may also feel valued for their linguistic contributions when they bring in their expertise in a reconciliation meeting with experts from heterogeneous backgrounds. Project manager B03 pointed out a downside to the process: it is time-consuming. All interviewees agreed that discussions can be difficult (because of different discussion styles). According to the translators, it takes a sensitive discussion leader to treat all group members equally, and to understand when a discussion must end and a resolution be made. But fundamentally, when translators are involved in group review meetings their feedback will not be lost, as might otherwise happen (see B06:62 from previous section).

All four interviewed verifiers have participated in training sessions with other verifiers, which they think are positive. Interviewee B18 explains that a "common language" (B18:64) is developed in these meetings, which is a view shared by Interviewee B11. She described a training session: "It's so much more happening […] when people are together. And not behind a screen" (B11:136).

Two translators (B15, B17) were also involved in question-and-answer sessions for translators working on the same text, which was an experience they both evaluated as positive.

Interviewees B13 and B14 reported that the daily routine of a freelance translator does not involve colleagues and work exchanges, which could positively influence the evaluation of meetings and collaborative work for those translators.

5. Discussion and conclusions

Given the exploratory and qualitative nature of the interview study, the commonalities and main themes should be considered as starting points for discussion, not as final conclusions relating to practices common to iLSA translations. All interviewees have extensive experience in their profession. However, the interviewees in this study cannot be assumed to be representative of all translators, verifiers or translation managers in iLSA. Further studies are needed to understand whether the conclusions drawn from these interviews also stand when a representative sample of iLSA translators is interviewed.

The interviews indicated that in iLSA, translators may not receive sufficient information throughout the process to make confident translation decisions. This observation is in line with the findings of Arffman (2013). Translators who are only involved in the translation step (and do not receive extensive training, further feedback or participate in later steps) simply produce text and must rely on the information they receive in the form of general or item-specific guidelines. For these translators, over-reliance on guidelines can be a pitfall (by understanding them as requirements versus recommendations). Even though item-specific guidelines seem to be helpful as a verifier's checklist, these guidelines may have a detrimental effect on the result. There are findings in translation process research to indicate that

> […] translators seem to subordinate local decisions to global ones, they do not always aim at an optimal result but at a text product which is adequate and sufficient for a particular communicative situation; and are ready to use their world knowledge and inferences about the text in general, and text type in particular, in order to make decisions. (Bolaños-Medina & González-Ruíz, 2012, p. 732)

Professional translators may thus be prevented from using their translation expertise in making translation decisions when relying heavily on (narrow) item-specific guidelines, which may, in the end, be contrary to the intent of these guidelines.

As shown above, item-specific guidelines for PISA or PIAAC are intended to help translators make comparable translation decisions across languages and to alert them to translation problems. Still, item-specific guidelines were initially developed for TIMSS, where the assumption was stated that two translators would, for the most part, translate items equally (the underlying premise being that translators are mere transcoders). Consequently, translators are – in this mindset – "likely to be dismissed as accessory participants in the process of adaptation, merely in charge of a superficial transfer of words between languages" (Bolaños-Medina & González-Ruíz, 2012, p. 728). Item-specific guidelines like the ones which explain English idioms or stress consistency support the impression that "psychologists [e.g. the developers of iLSA studies] are somewhat sceptical about translators and their work" (Bolaños-Medina & González-Ruíz, 2012, p. 728).

This mistrust might result from unsatisfactory experiences with previous translations. Still, by (unintentionally) narrowing the focus of (good) professional translators, the potential of translators as linguistic experts might get lost.

The following recommendations may be drawn to build trust and to strengthen the translator's role in the process:

First, translators selected for iLSA translation should meet the eligibility criteria as set by ISO 17100 (which may be difficult for some languages) to make sure they possess sufficient translation competence.

Secondly, the consortium should require national centres to train their translators on tests and the translation of tests. If translators feel sufficiently competent and fully understand the peculiarities of test translation, they may be more confident in using their translation expertise for solving translation problems. At the same time, the consortium should apply caution when writing item-specific guidelines for translators (if used at all) as they are easily misunderstood.

Involving translators in reconciliation meetings may be another means for boosting their confidence, and for including their feedback in later translation steps as well. In the long term, this may also help survey experts to better value translators' expertise and contribution. The consortium may also advise national centres to keep translators notified of the results of possible internal review or verification processes.

At the same time, the merits of reconciliation by a third translator remain questionable. As seen in the interview study, discussions between translators and reviewers may change the reviewer's first corrections. Furthermore, it is not evident why reconciliation by one person was recommended in the first place.

Meetings were seen as positive by all interviewees who had participated in them. Because of the generally positive effect on the translation process, national centres may be advised to use meetings as an introduction to test translation, and encourage participating countries to hold reconciliation meetings in practice. As an additional step, the consortium may consider

setting up (online) translators' and / or verifiers' meetings to discuss the source text and translation problems.

Generally, facilitating exchange with and between translators, and understanding them as experts in the process – a role also supported by functionalist theory (see Nord, 1993) – may help them improve at fulfilling their task. In summary: the role of translators needs to be strengthened because they are the key element of the iLSA translation process.

References

Adams, R. (2002). The Programme for International Student Assessment: An overview. In Organisation for Economic Co-operation and Development (OECD) (Ed.), *PISA 2000 technical report* (pp. 15–20). Paris. Retrieved from http://www.oecd.org/dataoecd/53/19/33688233.pdf

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* (Dissertation). University of Jyväskyla, Jyväskyla. Retrieved from https://ktl.jyu.fi/julkaisut/julkaisuluettelo/julkaisut/2007/t021

Arffman, I. (2012a). Translating international achievement tests: Translators' view. Reports. *Finnish Institute for Educational Research Reports: Vol. 44*. Jyväskylä: University of Jyväskylä. Retrieved from https://ktl.jyu.fi/julkaisut/ julkaisuluettelo/julkaisut/2012/g044

Arffman, I. (2012b). Unwanted literal translation: An underdiscussed problem in international achievement studies. *Education Research International, 2*, 1–13. https://doi.org/10.1155/2012/503824

Arffman, I. (2013). Problems and issues in translating international educational achievement tests. *Educational Measurement: Issues and Practice, 32*(2), 2–14. https://doi.org/10.1111/emip.12007

Behr, D. (2009). Translationswissenschaft und international vergleichende Umfrageforschung. Qualitätssicherung bei Fragebögenübersetzungen als Gegenstand einer Prozessanalyse. *GESIS-Schriftenreihe: Vol. 2*. Bonn: GESIS.

Behr, D. (2017). Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology, 20*(6), 573–584. https://doi.org/10.1080/13645579.2016.1252188

Bolaños-Medina, A., & González-Ruíz, V. (2012). Deconstructing the translation of psychological tests. *Meta: Journal des traducteurs / Translators' Journal, 57*(3), 715–739. https://doi.org/10.7202/1017088ar

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185-216. https://doi.org/10.1177/135910457000100301

Chidlow, A., Plakoyiannaki, E., & Welch, C. (2014). Translation in cross-language international business research: Beyond equivalence. *Journal of International Business Studies, 45*(5), 562–582. https://doi.org/10.1057/jibs.2013.67

Colina, S., Marrone, N., Ingram, M., & Sánchez, D. (2016). Translation quality assessment in health research: A functionalist alternative to back-translation. *Evaluation & the Health Professions*. Advance online publication. https://doi.org/10.1177/0163278716648191

Dolin, J. (2007). PISA – An example of the use and misuse of large-scale comparative tests. In S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *Schulpädagogik und Pädagogische Psychologie: Vol. 6. PISA zufolge PISA – PISA According to PISA: Hält PISA, was es verspricht? – Does PISA Keep What It Promises*? (pp. 93–125). Wien: LIT Verlag.

Eivers, E. (2010). PISA: Issues in implementation and interpretation. *Irish Journal of Education, 38*, 94-118. Retrieved from http://www.erc.ie/documents/vol38chp5.pdf

Ercikan, K. (1998). Translation effects in international assessments. *International Journal of Educational Research, 29*(6), 543–553. https://doi.org/10.1016/S0883-0355(98)00047-0

European Social Survey (ESS). (2016). *ESS round 8 translation guidelines*. Retrieved from http://www.europeansocialsurvey.org/docs/round8/methods/ESS8_translation_guidelines.pdf

Ferrari, A., Wäyrynen, L., Behr, D., & Zabal, A. (2013). Translation, adaptation, and verification of test and survey materials. In Organisation for Economic Co-operation and Development (OECD) (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)*. (Pre-publication copy). Retrieved from https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf

Grisay, A. (2002). Translation and cultural appropriateness of the test and survey material. In Organisation for Economic Co-operation and Development (OECD) (Ed.), *PISA 2000 technical report* (pp. 57–70). Paris. Retrieved from http://www.oecd.org/dataoecd/53/19/33688233.pdf

Grisay, A., Gonzalez, E., & Monseur, C. (2009). Equivalence of item difficulties across national versions of the PIRLS and PISA reading assessment. In M. von Davier & D. Hastedt (Eds.), *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessment* (Vol. 2, pp. 63–83). Retrieved from http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02_Chapter_03.pdf

Hambleton, R. K. (1993). Translating achievement tests for use in cross-national studies. *European Journal of Psychological Assessment, 9*(1), 57–68. Retrieved from http://files.eric.ed.gov/fulltext/ED358128.pdf

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology, 1*(1), 1–13. Retrieved from http://www.jattjournal.com/index.php/atp/article/view/48345/39215

Harkness, J. A., & Schoua-Glusberg, A. (1998). Questionnaires in translation. In J. A. Harkness (Ed.), *ZUMA-Nachrichten Spezial: Vol. 3. Cross-Cultural Survey Equivalence* (Vol. 3, pp. 87–127). Mannheim: Zentrum für Umfragen, Methoden und Analysen (ZUMA). Retrieved from http://nbn-resolving.de/urn:nbn:de:0168-ssoar-49733-1

ISO 17100. (2015). *Translation services - Requirements for translation services*. Berlin: Beuth.

Karg, I. (2005). *Mythos PISA. Vermeintliche Vergleichbarkeit und die Wirklichkeit eines Vergleichs*. Göttingen: V-&-R-Unipress.

Kirsch, I. S., & Thorn, W. (2013). The Programme for International Assessment of Adult Competencies: An overview. In Organisation for Economic Co-operation and Development (OECD) (Ed.), *Technical report of the Survey of Adult Skills (PIAAC)* (pp. 1-20). (Pre-publication copy). Retrieved from https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf

Kuckartz, U. (2014). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung (2. Aufl)*. Weinheim, Bergstr: Beltz Juventa.

Kurzman, C. (2014, September 2). World values lost in translation. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/monkeycage/wp/2014/09/02/world-values-lost-in-translation/?utm_term=.f3bbe26e5b57

Maxwell, B. (1996). Translation and cultural adaptation of the survey instruments. In M. O. Martin & D. L. Kelly (Eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report: Volume I: Design and Development*. Chestnut Hill, MA: TIMSS International Study Center, Boston College. Retrieved from http://timssandpirls.bc.edu/timss1995i/TIMSSPDF/TRCHP8.PDF

Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* (11th ed.). Studium Pädagogik. Weinheim: Beltz Juventa.

McKay, R. B., Breslow, M. J., Sangster, R. L., Gabbard, S. M., Reynolds, R. W., Nakamoto, J. M., & Tarnai, J. (1996). Translating survey questionnaires: Lessons learned. *New Directions for Evaluation, 70,* 93–104. https://doi.org/10.1002/ev.1037

Nida, E. A., & Taber, C. R. (1969). *The theory and practice of translation*. Leiden: Brill.

Nord, C. (1993). *Einführung in das funktionale Übersetzen. Am Beispiel von Titeln und Überschriften*. Tübingen: Francke.

Organisation for Economic Co-operation and Development (OECD) (Ed.). (2002). *PISA 2000 technical report*. Paris. Retrieved from http://www.oecd.org/dataoecd/53/19/33688233.pdf

Organisation for Economic Co-operation and Development (OECD). (2005). *PISA 2003 technical report. Programme for International Student Assessment*. Retrieved from http://www.oecd-ilibrary.org/docserver/download/9805011e.pdf?expires=1506674254&id=id&accname=guest&checksum=89C22C3A66C870305AF7DCCF137D9928

Organisation for Economic Co-operation and Development (OECD). (2009). *PISA 2006 technical report. Programme for International Student Assessment*. Retrieved from https://www.oecd.org/pisa/data/42025182.pdf

Organisation for Economic Co-operation and Development (OECD). (2012). *PISA 2009 technical report. Programme for International Student Assessment*. Retrieved from https://www.oecd.org/pisa/pisaproducts/50036771.pdf

Organisation for Economic Co-operation and Development (OECD). (2013a). *PIAAC sample items literacy*. Retrieved from http://www.oecd.org/site/piaac/Literacy%20Sample%20Items.pdf

Organisation for Economic Co-operation and Development (OECD) (Ed.). (2013b). *Technical report of the Survey of Adult Skills (PIAAC)*. (Pre-publication copy). Retrieved from https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf

Organisation for Economic Co-operation and Development (OECD). (2014). *PISA 2012 technical report. Programme for International Student Assessment*. Retrieved from http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf

Organisation for Economic Co-operation and Development (OECD). (2017). *PISA 2015. Technical report. Programme for International Student Assessment*. Retrieved from http://www.oecd.org/pisa/data/2015-technical-report

Ozolins, U. (2009). Back translation as a means of giving translators a voice. *Translation & Interpreting, 1*(2), 1–13. Retrieved from http://trans-int.org/index.php/transint/article/view/38/55

PISA Consortium. (2010). *Translation and adaptation guidelines for PISA 2012*. Retrieved from https://www.oecd.org/pisa/pisaproducts/49273486.pdf

Thorn, W. (2009). *International adult literacy and basic skills surveys in the OECD region*. OECD Publishing. Advance online publication. https://doi.org/10.1787/221351213600

Upsing, B. (2017). *Lokalisierung in internationalen Bildungsvergleichsstudien aus informationswissenschaftlicher Sicht: Herausforderungen, Übersetzungen, Prozesse* (Dissertation). Stiftung Universität Hildesheim, Hildesheim. Retrieved from http://d-nb.info/1145140017/34

Upsing, B., Gissler, G., Goldhammer, F., Rölke, H., & Ferrari, A. (2011). Localisation in international large-scale assessments of competencies: Challenges and solutions. *Localisation Focus, 10*(1), 44–57. Retrieved from http://www.localisation.ie/sites/default/files/publications/Vol10_1UpsingGissleretAl.pdf

Wuttke, J. (2007). Uncertainties and bias in PISA. In S. T. Hopmann, G. Brinek, & M. Retzl (Eds.), *Schulpädagogik und Pädagogische Psychologie: Vol. 6. PISA zufolge PISA – PISA According to PISA: Hält PISA, was es verspricht? – Does PISA Keep What It Promises?* (pp. 241–263). Wien: LIT Verlag.