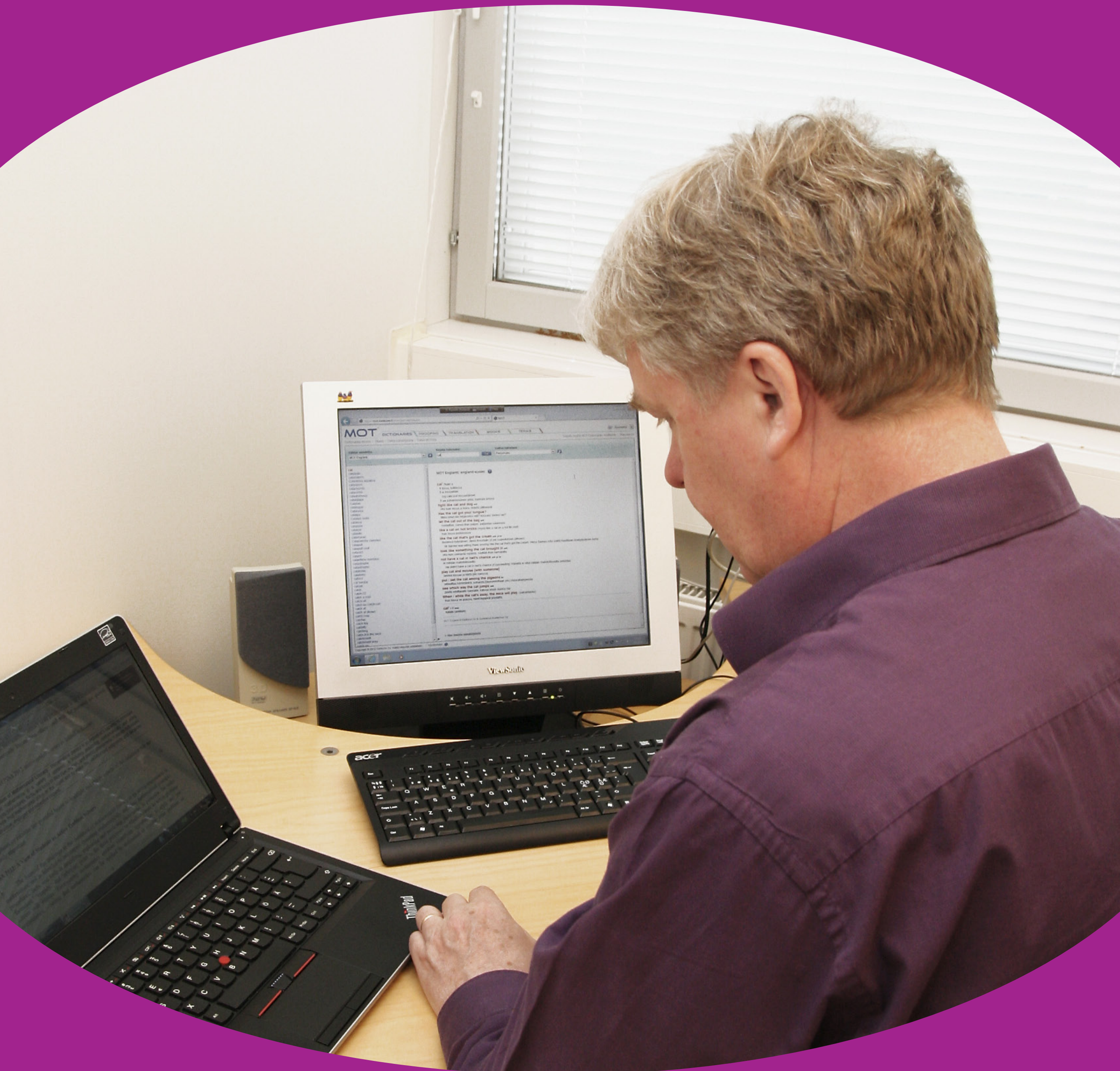
UNIVERSITY OF JYVÄSKYLÄ
FINNISH INSTITUTE FOR
EDUCATIONAL RESEARCH

Translating International Achievement Tests

TRANSLATORS' VIEW



INGA ARFFMAN

Translating International Achievement Tests:
Translators' View

Finnish Institute for Educational Research
Reports 44

Translating International Achievement Tests: Translators' View

Inga Arffman

UNIVERSITY OF JYVÄSKYLÄ
FINNISH INSTITUTE FOR
EDUCATIONAL RESEARCH

THIS PUBLICATION
CAN BE OBTAINED FROM:
Finnish Institute for Educational Research
Customer services
University of Jyväskylä
P.O. Box 35
FI-40014 Jyväskylä, Finland
Phone +358 40 805 4276
Fax +358 14 617 418
E-mail: ier-customerservices@jyu.fi
www.ier-publications.fi/

Cover and graphic design: Martti Minkkinen
Layout: Kaija Mannström

ISSN-L 1456-5153
ISSN 2243-139X (pdf)
ISBN 978-951-39-4811-5

Printed by University Printing House
Jyväskylä 2012

Contents

Arffman, I. 2012
TRANSLATING INTERNATIONAL ACHIEVEMENT TESTS: TRANSLATORS' VIEW

University of Jyväskylä
Finnish Institute for Educational Research
Reports 44

ISSN-L 1456-5153
ISSN 2243-139X (pdf)
ISBN 978-951-39-4811-5

Abstract

In international achievement studies, a common test is typically used which is translated into the languages of the participating countries. For the test to be valid, all translations and different-language versions of the test need to be equivalent to each other. Rigorous translation procedures and practices have been developed to ensure that this would be the case. This study explored how translators feel about these procedures and practices, what problems they think there have been in them. The ultimate purpose was to help to develop the procedures and practices and in this way to increase the equivalence of the test versions and the validity of the studies. An action research study was carried out: Semi-structured face-to-face and email discussions were conducted with the five translators participating in translating the PISA (Programme for International Student Assessment) 2009 reading test into Finnish. The discussions were analyzed by means of content analysis, the analysis drawing on findings from translation studies and test translation. The results show that while the translation procedures in international achievement studies have developed over the years, there are still deficiencies in them which may jeopardize equivalence. Problems have been caused by 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. Suggestions are given on how to develop the procedures and practices.

Keywords: Equivalence, validity, international achievement studies, translation, translators

Arffman, I. 2012
KÄÄNTÄJIEN NÄKEMYKSIÄ KANSAINVÄLISTEN ARVIOINTITUTKIMUSTEN KÄÄNTÄMISEN ONGELMISTA

Jyväskylän yliopisto
Koulutuksen tutkimuslaitos
Tutkimusselosteita 44

ISSN-L 1456-5153
ISSN 2243-139X (pdf)
ISBN 978-951-39-4811-5

Tiivistelmä

Kansainväliset arviointitutkimukset edellyttävät laajaa käännöstyötä, koska kaikki koemateriaalit on käännettävä kunkin osallistujamaan kielille. Kaikkien erikielisten koeversioiden on lisäksi oltava keskenään vertailukelpoisia, jotta tutkimustulokset olisivat valideja. Arviointitutkimuksissa onkin kehitetty nimenomaisia käännösprosesseja ja -käytänteitä varmistamaan koeversioiden vertailukelpoisuus ja tutkimustulosten validius. Tässä tutkimuksessa tarkasteltiin kääntäjien kokemuksia näistä prosesseista ja käytänteistä ja niissä ilmenneistä ongelmista. Tavoitteena oli kehittää kansainvälisten arviointitutkimusten käännösprosesseja ja -käytänteitä ja lisätä näin näiden tutkimusten tulosten validiutta. Tutkimus suoritettiin toimintatutkimuksena. Aineisto kerättiin keskusteluina viiden suomalaiskääntäjän kanssa, jotka käänsivät PISA (Programme for International Student Assessment) 2009 -lukukokeen materiaalit suomen kielelle. Keskustelut analysoitiin sisällönanalyysin avulla. Analyysien tulokset osoittivat, että vaikka kansainvälisten arviointitutkimusten käännösprosessit ja -käytänteet ovat vuosien kuluessa kehittyneet, niissä on edelleen ongelmia, jotka saattavat vaarantaa tutkimustulosten validiuden. Ongelmia ovat aiheuttaneet muun muassa alkuperäinen koeversio, epäpätevät kääntäjät, käännöstavoitteen epämääräisyys sekä epäsopivat käännösohjeet, kahden erikielisen alkuperäisversion käyttö, käännösten puutteellinen tarkistus ja verifiointi sekä ajan puute. Tulosten pohjalta esitetään ehdotuksia kansainvälisten arviointitutkimusten käännösprosessien ja -käytänteiden kehittämiseksi.

Asiasanat: Vertailukelpoisuus, validiteetti, kansainväliset arviointitutkimukset, kääntäminen, kääntäjät

Preface

As long as international achievement tests have been administered, concerns have been voiced about the comparability of the different-language versions and translations of these tests. To answer these concerns, those conducting the studies have referred, among other things, to their standardized translation procedures and practices and claimed that thanks to these, they have been able to overcome most of the potential translation problems and ensure equivalent translations. At the same time, however, observations and experiences suggest that the procedures and practices have often not worked as they should have and that the quality of the translated versions has not been as high it should have been.

How, then, have the translation procedures and practices really worked? Have they been easy to follow? Have they helped to ensure high-quality and equivalent translations? Or have there been problems in them? Surprisingly little research has been done on the procedures and practices. This book is an attempt to fill this research gap.

To find answers to the above questions, translators were addressed and asked about their experiences and views when translating the tests and following the procedures and practices. The translators voicing their views were those rendering the PISA 2009 reading test into Finnish. My most sincere thanks go to all these translators. This study would not have been possible without their kind cooperation and help. I also humbly thank the Academy of Finland for supporting me with a grant (Grant No. 206176).

I dedicate this book to all devoted test translators, who play a significant role in ensuring the validity of international tests. I hope this book will help to improve the procedures and practices followed when translating these tests and the quality and equivalence of the different-language versions of these tests.

Jyväskylä, June 2012

Inga Arffman

CHAPTER 1

Introduction

1.1 Background

Recent years have witnessed a huge interest in international achievement studies, the results of these studies being increasingly used, for example, in educational decision-making. Studies have been organized, for example, by the International Association for the Evaluation of Educational Achievement (IEA), the Educational Testing Service (ETS), Statistics Canada (STATCAN), the Organization for Economic Co-operation and Development (OECD), the United Nations Educational, Scientific and Cultural Organization (UNESCO), and the Southern Africa Consortium for Monitoring Education Quality (SACMEQ). In all these studies, a common test has been used which has been translated or adapted[1] into the languages of the participating countries.

When translating international achievement tests, it is important to ensure that the versions (i.e., the question items and the texts and materials, or stimuli, accompanying them) are equivalent, or comparable, to each other – that they measure the same construct and are equally easy to answer. For this to be the case, the mental effort required by testees to respond to the test items needs to remain the same across languages. No version must

[1] In this book, the term 'translation' refers to the process of reproducing a text originating in one language and culture (i.e., a source text) for use in another language and culture (i.e., a target text). The term involves the whole process of translation, from the reading and comprehension of the source text to the revision and finalizing of the target text. Also, it covers all kinds of between-language meaning transfer, from literal translation to adaptation, or the making of major changes to the target version so as to make it more suitable for the target population (e.g., changes in currency or measurement units). The term 'adaptation', thus, refers to a special subtype of translation.

place a heavier cognitive load (Sweller, 1988; see also e.g., Rueda, 2011, pp. 93-4), or consume more of the limited processing capacity of testees' working, or short-term, memory compared to the other versions. This, in turn, requires, among other things, that all versions be equally easy to understand. If this is not the case and some items or stimuli are harder to understand than others, more working memory is needed to decode and make meaning of them – comprehension of the question items being the first phase in the survey response process (Tourangeau, Rips & Rasinski, 2000, p. 8). As a result, less memory is left for actually responding to the items (i.e., retrieving the relevant information, forming the judgment, and editing the answer; e.g., ibid., p. 8). Readers of versions that are harder to understand are thus at a disadvantage, which, in turn, jeopardizes the validity of inferences made on the basis of the test.

Thus, if, for example, the source version uses a literal match to link an item to its stimulus text (i.e., exactly the same word or expression in both), a literal match should also be used in all target versions, because other types of matches (e.g., synonyms) require the testee to do more inferencing and thus consume more memory capacity (cf. Kirsch, 2001; Mosenthal & Kirsch, 1991). Or if there are no explicit markings (e.g., conjunctions) to signal a link or meaning relationship (e.g., cause-and-effect) in the source version, no explicit signals should be used in the target versions either, because these would reduce the amount of inferencing and cognitive processing required of the testee (cf. Kemper, 1983). Or if natural, idiomatic and authentic language is employed in the source version, similar language should also be used in all target versions, because unnatural, odd and cumbersome language cannot be processed in as large chunks and as automatically and effortlessly as natural and idiomatic language (cf. Kintsch & van Dijk, 1978; van Dijk & Kintsch, 1983).

Clearly, translating international achievement tests is a most responsible and demanding task. Rigorous translation procedures and practices have therefore been developed to ensure high-quality and equivalent translations. However, research on translations used in these studies shows that the quality of the translations has often not been as good as it should have been (e.g., Arffman, 2007; Bechger et al., 1998; Blum, Goldstein & Guérin-Pace, 2001; Bonnet, 2002; Ercikan & Koh, 2005; Karg, 2005). This, in turn, suggests that there have been deficiencies and problems in the procedures and practices followed when making the translations. Therefore, to be able to ensure equivalent translations, it is important to find the problems and ways to solve them. However, extremely little research exists on the procedures and practices followed when translating international achievement tests.

1.2 Purpose of the Study

This study looks at the procedures and practices followed when translating international achievement tests. The aim is to examine how translators feel about these procedures and practices, what problems they think there have been in them. The ultimate goal is to help to develop the procedures and practices and in this way to increase the equivalence of translations used in international achievement studies and the validity of interpretations made on the basis of these studies.

Listening to the translators and to their views was considered important, because translators are the ones who actually translate the tests. They are the ones who actually follow the translation procedures and practices and are affected by them. They are the ones who make the final decisions on how to translate – and they are the ones on whom the quality of the translations is, in the end, dependent. They are the experts in the translation process, as emphasized in translation studies (e.g., Vermeer, 1989).

This study is a sub-study of a larger action research study. Action research is the process by which practitioners study their own practice scientifically in order to evaluate it and to solve their practical problems (Corey, 1953; McKernan, 2000, p. 5). The practitioners in this study are the translators rendering the OECD PISA (Programme for International Student Assessment) 2009 materials into Finnish. Strictly speaking, the discussion in this study thus only applies to the translation of the PISA 2009 test as implemented in Finland. However, since translation procedures in other international achievement studies and other countries are often in many respects similar to those followed in PISA and in Finland, the discussion may be expected to apply largely also to these other studies and countries.

Theoretically, the study is mainly grounded in translation studies, whose principles should guide all translation work. However, it seems that in test translation the principles have often been ignored (see e.g., Harkness, Villar & Edwards, 2010, p. 118).

1.3 Outline of the Book

Following this introductory chapter, Chapter 2 briefly describes the major context of the study: the PISA programme, the composition of PISA tests, and the translation and verification procedure followed in PISA. Chapter 3 depicts the method that was used in this study: face-to-face and email discussions conducted with the five translators translating the PISA 2009 materials into Finnish. Chapter 4 discusses the findings of the discussions, the problems encountered by the translators while translating international achievement tests, relating them to previous studies of translation problems in translation studies and test translation and providing suggestions on how to solve them. Chapter 5 summarizes the problems and recommendations and provides suggestions for further research in the field.

CHAPTER 2

PISA Translation Procedure

PISA is a regular, ongoing assessment administered to 15-year-olds. The first PISA assessment was organized in 2000, and thereafter assessments have taken place every three years. During every assessment cycle, a closer look is taken at one of the three major domains of PISA: reading literacy, mathematical literacy, or scientific literacy. In the first PISA assessment, in 2000, the focus was on reading literacy. In PISA 2009, too, reading literacy was the major domain. (For more information on PISA, see e.g., http://www.pisa.oecd.org).

2.1 PISA Tests

PISA tests consist of units which contain authentic, or real-life, stimulus material and a number of question items related to it. The stimulus materials may take the form, for example, of text, a table, a graph or a diagram. In reading literacy tests, moreover, the texts represent different text types (description, narration, exposition, argumentation, instruction; for the definitions of the text types, see e.g. OECD, 2010b, p. 42) and genres (e.g., advertisements, text books, short stories). The items, again, are either multiple-choice or constructed response items. (For examples of PISA texts and question items, see Appendix 1 and OECD, 2009c.) Traditionally, the tests have been provided on paper. In more recent years, however, also optional computer-based components have been provided. In PISA 2009, 19 countries and economies took part in an electronic reading assessment. However, Finland was not among these countries. In PISA 2012, computer-based assessments were also administered in problem solving (where it was the only option; no paper-and-pencil test was arranged) and mathematics (where it was an additional option). In PISA 2012, Finland took part in the computer-based problem-solving test.

2.2 Development of the Source Versions

Some of the texts and items included in the PISA instruments are submitted by the participating countries, and the rest are chosen by the test developers. All the proposed materials are translated into English and exposed to extensive reviews by test developers, subject experts and translators, and to pre-pilots in some participating countries. Each country also reviews the items from the point of view of cultural relevance, interest, sensibility, and translatability. In more recent PISA studies, moreover, some of the items have been subjected to cognitive interviews (interviews with testees to obtain data on their response processes, e.g., how they understand items and why they answer as they do) and think aloud protocols (or TAPs, where subjects think aloud while performing a task). (OECD, 2009a, pp. 31–5; 2012.) In PISA 2009, new texts and items were only developed for reading literacy. For mathematics and science, new materials were not needed, because when assessing minor domains, only link items (i.e., items retained from earlier PISA rounds) are used.

All the stimulus materials and items are translated into French. The method is the same as for the national versions: double translation followed by reconciliation and verification (the method is described in more detail in the following section, 2.3). The translation takes place while the English source instrument is still being developed, and the experiences gained in translating the French source version are used to make further modifications to the source instrument (e.g., Grisay, 2003; OECD, 2009a, pp. 35, 86–7; 2012). The finalized English and French versions are then regarded as the source versions, on the basis of which the participating countries make their translations. In PISA 2009, all the materials that were translated in Finland (and most other countries) were reading literacy materials (new countries also had to translate e.g., older mathematics and science materials). This study therefore primarily deals with the translation of reading literacy tests.

2.3 The Recommended PISA Translation Procedure and the Procedure Followed in Finland

The test materials are translated into all the languages of instruction used in the schools sampled in the participating countries (see e.g., http://www.oecd.org/dataoecd/15/47/48580428.pdf). In practice this means that, for example, in Finland the test has to be translated not only into Finnish but also into Swedish, which is the other of the two national languages spoken in the country (albeit by only some 5–6% of the population). However, these Swedish materials are not translated in Finland but in Sweden, from where they are borrowed, and only adapted for use in the Finnish context[2] (e.g., Swedish *knonor* are changed to euros). This study focuses on the translation procedures and practices followed in Finland and thus mainly discusses the translation of Finland's Finnish materials.

The recommended PISA translation and verification procedure (as the procedure is called in its entirety) is *double translation from English and French, followed by national reconciliation and international verification* (Grisay, 2003): Two translators produce two independent target versions, one from English and the other from French; the two target versions are merged into one national version by a reconciler, who also checks that the resulting version is correct and natural; and the reconciled version and its equivalence to the source version is verified by an international verifier (see Table 1; see also OECD, 2007; 2010c).

Many countries, however, have not followed the recommended procedure. One of these countries has been Finland. Finland did follow the procedure when translating for the International Adult Literacy Survey (IALS; conducted in 1994–1998), where the practice was used for the first time. However, in PISA, Finland's procedure has usually been slightly different. In PISA 2009, the procedure was as shown in Table 1: For every unit, one national translator produced one Finnish version, on the basis of the English source version; these first Finnish drafts were compared carefully with the English and French source versions by a first reconciler, who also checked the semantic accuracy and linguistic quality of the Finnish versions; these versions were compared briefly with the English source versions by a second reconciler, who, however, mainly concentrated on checking that the Finnish version was correct and in natural language; and these versions were verified by an international verifier. The Finnish translation procedure in PISA 2009 might thus be termed *single translation from English, followed by two national revisions, one with extensive cross-checks against the French source version, and an international verification.*

Finland followed the same procedure also in PISA 2003. However, in the first PISA study, PISA 2000, the procedure was slightly different, in that less use was made of the French source versions: During Step 2, the first Finnish versions were typically only checked against the English source versions; and only some occasional and minor crosschecks were made against the French versions. In PISA 2006, the assessment was implemented by another Finnish institute and the recommended procedure was followed.

---

2 Borrowing a verified version from another country where the same language is spoken and only adapting it for use in the country in question is the recommended procedure in PISA when preparing test materials for minority languages (e.g., OECD, 2009a, p. 90; OECD, 2012). In Finland, the adaptation of the Swedish materials has normally been carried out by one (or two) Swedish-speaking Finn(s). However, the adaptation of the PISA 2012 problem-solving materials was performed by a Finn who had lived in both Finland and Sweden.

Table 1. *The recommended PISA translation and verification procedure and the procedure followed in Finland in PISA 2009*

Step	Recommended PISA procedure	Finland's procedure in PISA 2009
1	For every unit, two national translators produce two independent target language versions, one of them based on the English and the other on the French source version.	For every unit, one national translator produces one Finnish version, on the basis of the English source version.
2	The two independent versions are reconciled into one national version by a third translator (reconciler).	The first Finnish drafts are compared and checked carefully with the English and French source versions, revised linguistically and reviewed for appropriateness of content and terminology by another translator (first reconciler), who also acts as a domain expert. The translator makes corrections and suggestions for improvement in the version.
3	The reconciled versions are reviewed for appropriateness of content and terminology by a domain expert. The expert discusses with the reconciler, suggesting possible edits to the versions. The reconciler then decides on the "next-final" national versions.	The revised versions are compared briefly with the English source versions, revised linguistically and reviewed for appropriateness of content and terminology by a third translator (second reconciler), who mainly acts as a domain expert. The translator briefly discusses some of the most problematic cases with the first reconciler and then decides on the "next-final" Finnish versions.
4	The "next-final" national versions are verified by a fourth, independent translator (verifier) from the International Project Centre. The verifier makes his or her suggestions for correction and improvement in the versions, some of them obligatory and some optional.	The "next-final" Finnish versions are verified by a fourth, independent translator, verifier, from the International Project Centre, who makes his or her suggestions for correction and improvement in the versions.
5	The national translators decide on the final versions and have them compiled into test booklets.	The first reconciler decides on the final Finnish versions and has them compiled into test booklets.
6	The verifier checks that the obligatory corrections have been made, and the booklets are checked optically for layout errors at the International Project Centre.	The verifier checks that the obligatory corrections have been made, and the booklets are checked optically for layout errors at the International Project Centre.

The main differences between the recommended PISA procedure and the modified procedure followed in Finland have been the following: First, in Finland during Step 1 only one translation has been produced which has then been revised and reworked by other translators during the following phases, whereas PISA recommends that two parallel, or comparable, translations be made (one from English and the other from French) and reconciled into one translation. Second, in Finland somewhat more emphasis has been put on the English than on the French source versions, even though the recommendation is to make equal use of both source versions. Third, in Finland two steps (Steps 2 and 3) have concentrated more or less directly on the target versions and their linguistic refining, whereas in the recommended procedure there is only one such step (Step 3), which, moreover, only involves terminological reviewing.

In practice, most of the translation and adaptation work has been done on screen by overwriting source language text, replacing it with target language text (OECD, 2007). Since PISA 2009, however, the computer-based components have been translated on-line, by using a computer-aided translation tool (for more information on this translation system, see e.g., OECD, 2010a). In Finland, this tool was not used in PISA 2009, because Finland only took part in the paper-and-pencil tests. This study therefore mainly focuses on procedures and practices followed when translating printed tests in non-electronic environments.

2.4 Translators, Translation Guidelines and Translator Training

The materials are translated by translation teams, consisting of translators (national translators making the first drafts), reconcilers (national translators merging the two first drafts into one) and verifiers (independent international translators checking the equivalence of the source and target versions). If needed, the teams may also consult subject matter specialists (see Table 1).

The requirements for the translators (henceforth used as a generic term for all members of the translation team – translators, reconcilers, and verifiers – unless otherwise specified; cf. Footnote 1) vary somewhat according to their roles. However, usually they are expected to have a perfect command of the target language (native speakers) and one of the source languages (English or French). In addition, reconcilers are said to benefit from the command of also the other source language. Especially when translating mathematics and science units, the translators are to be acquainted with the subjects covered. This, as pointed out in the translation guidelines, may make it necessary to use different translators to translate the various tests. (OECD, 2007). The requirements for the verifiers are even higher. For example, they are *required* to have a sufficient command of also the other source language (Grisay, 2002, p. 62).

To help translators in the translation work, PISA provides them with written translation guidelines (see e.g., OECD, 2007, 2010c). These include directions on the layout and presentation of the translations, on how to maintain the difficulty level of the vocabulary and syntax of the stimulus material unchanged, and on how to translate the question items. At times, the translators are also provided with more item-specific translation notes (for examples, see Appendix 1). Translation notes may be used, for example, to ask the translator to imitate the stylistic characteristics (e.g., irony) of the source text, to point out where the translator should make a national adaptation (e.g., change the currency), or to indicate where the translator is to remain strictly true to the original (e.g., not using the same term in a question item as in the stimulus). Finally, in reading literacy tests, translators are told for each item whether answering it calls for forming a general understanding, retrieving information, developing an interpretation, reflecting on the content of the text, or reflecting on the form of the text (see Appendix 1). The purpose of this item-specific information is to help to prevent translators from modifying the nature of the questions and the strategies required to answer them.

Based on the translation guidelines, countries are encouraged to offer training to their national translators. Verifiers are trained by the International Centre.

CHAPTER 3

Research Design

3.1 Action Research

This study, in addition to being a part of a larger action research study, is itself also an action research study, aimed at developing translation work in international achievement studies and increasing the equivalence of the different-language test versions used in these studies. In action research, the researcher often takes an involved role as a participant in the process. In this study, too, I worked not only as the researcher, but also as one of the translators translating the PISA 2009 materials into Finnish.

3.2 Participants

In Finland, altogether five translators took part in rendering the PISA 2009 materials into Finnish. The translating took place in December 2007 to January 2008. Four of the translators were national translators, and one, the verifier, was an independent translator appointed by the International Project Centre. The position of the verifier differs somewhat from that of the other translators, in that s/he also evaluates the quality of the translations. However, contrary to what is usually the case with translation evaluations, the verifier's evaluations are not only final, external appraisings (with no real impact on the translations). His or her comments and suggestions are actually used to improve the translations. Therefore, the verifier is regarded here as part of the translation process and as one of the translators.

All the five translators were native speakers of Finnish. Three were professional translators, and of the three, two were official translators graduated from Finnish translator training schools; one with six years and the other with two to three years experience in translating and proofreading. The latter translator had also taken part in the special training provided for translators or verifiers in PISA and had been involved in translating three international achievement tests. The former had had no such training and experience. The translator did get acquainted with the written PISA translation guidelines, which were sent to him/her well before the translation work started. Yet, when s/he was offered the opportunity to get in-service training for the translation work, s/he felt s/he did not need it, because s/he had already got all the necessary information by reading the guidelines. In addition, s/he was confident that his/her training and experience as a translator would help him/her to succeed also in this translation task. The third professional translator was a BA in applied linguistics, English and Finnish, with, however, no university training in translation. S/he had 15 years experience in translating and proofreading. In addition, s/he had taken part in the training provided for translators or verifiers in PISA and had translated for several (at least seven) international achievement studies.

The two remaining translators were linguists and researchers in reading literacy, with, moreover, ample experience in international achievement studies. One of the translators was a mother tongue teacher, with studies in Finnish, literature and educational science. The translator had had no translator training. However, s/he had some experience in proofreading Finnish texts, and s/he had also acted as a reviser in one international reading literacy study. I myself have studied English, French, educational science, applied linguistics and translation. Primarily I work as a researcher, my main interest being translation and equivalence of international achievement tests. However, I have also translated and proofread from and into English, French and Swedish, and I have revised for four international achievement studies.

The division of labour between us was the following: The translator with ample experience in test translation but no university training in translation and the translator with university training in translation but no training or experience in test translation rendered the units from English into Finnish (Step 1 in Table 1). At first, only the more experienced translator was hired to do this. However, when it became evident that s/he would not be able to translate all the 35 units in time, the other translator was hired, too. This translator had been contacted beforehand, and s/he had promised to do part of the translating, if needed. In the end, the translator who had been hired first translated 24 of the 35 units and the other translator 11 units. For all the other translation steps, there was only one translator: I was the first reconciler and checked all Finnish units (Step 2); the researcher who, in addition, was a mother tongue teacher worked as the second reconciler, going through all units (Step 3); and the professional translator with training and experience in both translation in general and test translation verified all units (Step 4). I also checked all units after they had been verified and decided on their final versions (Step 5), and the verifier made sure that all obligatory corrections had been made to the final versions (Step 6).

3.3 Data Collection

To collect data from the translators, I conducted semi-structured face-to-face and email discussions with them. Initially, I meant to have all discussions face-to-face, which I thought was the most purposeful way of collecting information-rich, in-depth and "thick" data (Patton, 2002, p. 40) from participants working as peers and colleagues in the same venture (cf. Bohm, Factor & Garret, 1991). However, with two of the translators this was not possible. This was because both were staying abroad for a longer period of time right after the translation work was completed, and yet I felt that it was necessary to hear about their experiences as soon as possible in order to ensure that the translation work was still fresh in their minds. Besides, one of the translators explicitly mentioned that s/he preferred commenting via email. Thus, I decided to have face-to face discussions with those translators with whom it was possible and email discussions with those with whom it was not possible. This would enable me to collect as thick and rich data as possible (because when speaking people tend to be much "thicker" than when writing). Also, to ensure that the collection of data would be as consistent as possible, I decided to use the same outline in all discussions (see Appendix 2) and to keep the written communication as conversational as possible (see below).

In the face-to-face discussions, I acted not only as the researcher and interviewer, but also as one of the translators having taken part in the translation work and sharing her experiences and views. This posed a special challenge to the discussions, in that my comments might easily have obtrusively affected the other translators, led the course of the discussion and biased the results. To meet this challenge, I took precautions both before and during the discussions: Before the discussions, I sent the written outline of the discussion to the translators via email so that they were able to prepare for the discussions; and during the discussions, I made special effort not to make leading questions, not to be the first to voice my opinions, and not to speak too much. One of the discussions lasted one hour 40 minutes and the other 2 hours 10 minutes. Both discussions were audiorecorded and transcripted verbatim.

The email discussions were conducted so that the whole outline of the discussion was sent to the two translators at a time, but the translators were free to decide whether to provide all their comments at a time or whether to comment in smaller batches. Both provided all comments at a time. It was also agreed that the discussions could continue via email, each of us having the opportunity to make further questions or comments. However, this opportunity was only used once, when I made a clarifying question to one of the translators. All discussions, whether face-to-face or email, were conducted in February to March 2008.

3.4 Data Analysis

All collected data were analyzed qualitatively, by means of content analysis: The data were read and reread, all references (both direct and indirect) to translation problems and difficulties were located, marked and categorized into seven main factors causing translation problems (along with their subcategories). The analysis and categorizing of the data were based on and guided by previous findings from translation studies and test translation.

CHAPTER 4

Translation Problems

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. These are discussed in more depth in the following. For each factor, the section first describes what has been discovered in previous research in translation studies and when translating international achievement tests, then presents findings of this study, and finally discusses how the findings of this study relate to those of previous studies and what can be done to tackle the problems.

4.1 Source Instrument

4.1.1 Previous Research

Translation studies. 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 behavior, based on common knowledge and expectations; Nord, 1991, p. 96) (Baker, 1992; Campbell, 1999; Danks & Griffin, 1997, pp. 168–70; Hale & Campbell, 2002; Kuhiwczak, 2003, p. 117; Nord, 1997; Wilss, 1990, p. 27).

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, p. 42; Pym, 2008, p. 324). Also, the translation easily becomes simpler, or "flatter", than the original text (Laviosa-Braithwaite, 1998, pp. 288–9).

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, p. 118; Lefevere, 1992; Neubert & Shreve, 1992; Scarpa, 2002; see also 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, p. 170), because normally both form and content cannot be preserved at the same time (Bell, 1991, p. 6).

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, p. 59; Rosenblatt, 1994; Scarpa, 2002). They are therefore generally thought to be easier to translate than, for example, literary texts (Kuhiwczak, 2003, p. 118; 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).

International achievement studies. 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). Words or structures in the source instrument, such as long sentences, broad and vague meanings (e.g., words with several meanings), metaphors, connotations, and differences in word order, may not have been easy to translate into other languages: They may have been hard to find equivalents for, or they may have been misunderstood. This, in turn, may have resulted in changes in meaning or nuances and, ultimately, in items functioning differentially. (Brislin, 1986, pp. 144–9; cf. e.g., Allalouf, 2003; Ercikan, 2002; Gierl & Khaliq, 2001; Grisay, 2004; see also Elosua & López-Jaúregui, 2007; Solano-Flores, Backhoff & Contreras-Niño, 2009).

The above findings have mainly been gained when translating question items. More recently, however, research has also been done on the stimulus materials. In these materials, largely the same linguistic features have been found to be problematic as when translating question items. For example, when Karg (2005) examined German PISA 2000 and 2003 field trial and main study materials (for the materials, see e.g., OECD, 2009c), comparing them to the English and French source versions and to some national (e.g., Finnish) versions, she found that problems were caused by technical and foreign terms, vague meanings and words with several meanings, connotations, and figurative language. Similar results were also obtained by Arffman (2007), who compared three Finnish and English texts used in the PISA 2000 reading test (main study): an expository text[3] (article in a youth magazine), a literary text (short story), and a non-continuous table. In addition to these, however, when translating stimulus materials, also some extra features have been found to be problematic. For example, both Arffman (2002; in a similar study as the one conducted in 2007, except that the texts were used in the PISA 2000 field trial reading literacy test but no longer in the main study) and Karg (2005) found that differences in conventions made it impossible to translate a warranty card in an equivalent way into Finnish and German, respectively. Problems have also been caused by dense language, textual factors, style and unfamiliar content (Arffman, 2007; Karg, 2005). Finally, in the study by Arffman (2007), the literary and non-continuous texts were found to be more difficult to translate than the expository text, the literary text mainly because of its strong stylistic and aesthetic emphases and the non-continuous text because of its density and textual and word order problems.

4.1.2 Findings of This Study

Problematic linguistic features. The translators in this study also mentioned the source instrument as one of the factors that made their translating and attaining equivalence difficult. Also, the features they mentioned as problematic were largely the same that have already

[3] Expository texts aim at informing, describing, explaining, or defining something to the reader. Examples of expository texts include textbooks, scientific articles, directions, guides, and newspapers. Expository texts are the most frequent text type.

been found problematic in previous studies. However, they also brought up some new (or more specific) features. The features included, for example, dense language (e.g., verbless clauses, prepositional phrases), word order (sentence-initial adverbials and intrasentential supplements), translating nouns as nouns (not as e.g., verbs), using the same type of reference or reference chains (e.g., pronouns, synonyms), technical terminology, broad and abstract meanings, words with several meanings, affective meanings, idioms and metaphors (e.g., puns), and style and conventions. From among these, by far the most problems were caused by dense language, word order, technical terminology, style and conventions.

Question items. All in all, the translators felt that the above features caused more problems when translating the question items than when translating the stimulus texts and that, as a result, translating the question items was more difficult than translating the stimulus texts.

> I: *What came as a surprise to me was that often it was these questions in particular which were [difficult to translate].*

One of the problems when translating the question items was dense language. The problem was that the source questions were often so complex and contained so much information packed into one sentence that it was impossible to find fluent Finnish translations for them.

> I: *You can't really do much about it [the difficulty of translating the questions]. You have to have the given elements [elements in the source questions] in the [translated] questions … all of them….*
> Translator A: *Yes, there is this certain convolutedness about these questions.*

Also, the word order in the questions was often strange to Finnish. For example, questions starting with a fronting frame ("*On the basis of the article*, what do you think…") are frequently difficult, if not impossible, to translate into Finnish in a comparable way. Typically, the word order has to be changed, because in Finnish, questions normally start with the main question (e.g., "What do you think…"), which is then followed by all secondary information (including the information contained in the fronting frame, which thus has to be embedded in the remaining part of the interrogative clause). This, in turn, means that English and Finnish questions often differ, for example, in information structure and emphasis. This, again, may be reflected in how students respond to the questions.

> Translator C: *The problem with them [the questions] was how to make them as nice and light in Finnish. Because we haven't got these structures ... which they have in English so that you can put something in front and separate it with a comma, such as "Thinking about this and that".... The English formulation seems quite easy to put up with. At first they give the frame, and then comes the ques-*

tion. But we have to embed the information contained in the frame in the interrogative clause, which makes it much heavier….
I: *And it often even becomes hard to understand. It doesn't flow smoothly.*

Still another factor that caused problems when translating the question items had to do with conventions. The problem was that conventions are culture-specific and that therefore, when translating, conventions in the source text often cannot be transferred to the target text but need to be replaced with conventions in the target culture (e.g., Nord, 1997, p. 54). In practice, however, when translating the question items, the translators were often expected to follow the conventions of the source instrument. For example, at times, items in the source instrument consisted of two or more sentences in which exactly the same key words, phrases or structures were repeated. However, the conventions concerning repetition vary greatly from language to language (e.g., Baker, 1992, p. 210). In Finnish, for example, repetition is usually considered poor language and naïve. Therefore, Finnish language users are normally taught to avoid it and use synonymic expressions instead. However, in PISA, the translation guidelines and/or translation notes often require that reference chains in the source instrument be retained also in the translations and that if, for example, repetition is used in the source items, it should also be used in the target items. Therefore, the translators often felt that they had to use repetitions, even though they knew that these were against Finnish text conventions and made the Finnish items unnatural, unidiomatic, clumsy – and non-equivalent to the corresponding items in languages where repetition is better tolerated and more used.

I: *When talking about the questions, there is this repetition…. English has this writer-responsible system. They say everything explicitly. But for Finns, that sounds naïve. But since the rules [translation instructions] are as they are, what can you do? I can't remember which text it was, but I thought that this does not sound good. Once again I thought that normally I wouldn't say like this….*
Translator C: *In a way, you have to compromise authenticity.*

One of the main reasons for the above problems, as conjectured by the translators, was that not enough attention had been paid to the formulation and translatability of the question items.

Translator A: *If the [stimulus] texts themselves have been internationalized and simplified…, so perhaps the same hasn't been done when formulating the questions…. They know how to ask these things in English. But they haven't thought how to translate them into other languages.*

As a result, the translators felt that the Finnish questions often ended up being unnatural and cumbersome and more difficult to understand than the questions in the source versions.

> I: *When you have to translate them [question items] into Finnish, they sound much more complicated…. Surely, it shouldn't be like that – that it becomes more difficult, the question itself.*
> Translator A: *That's right. Because they are not measuring whether they [students] understand the question. Or at least that shouldn't be the goal.*
> I: *The [Finnish] questions couldn't always be made, well, neat. And often I thought that I wouldn't ask like this.*
> Translator A: *Exactly. Because in Finnish, the question would be asked in a much simpler, clearer and more accurate way…. Several times I felt that, well …. this question could be much more accurate. And the Finnish way of asking … would be clearer.*

Stimulus texts. As concerns the stimulus texts, the translators felt, somewhat surprisingly, that the scientific texts were more difficult to translate than the literary texts.

> I: *Have you noticed that the texts that we have mentioned here [as being among the most difficult texts to translate] have been mainly expository texts? And in a way it was a surprise to me that as a whole these literary texts didn't cause so many problems....*
> Translator C: *Yes, it was especially these expository texts…*
> I: *… those that are of the more scientific type.*

The main source of difficulty in the scientific texts was dense language. The scientific source texts often used formal, "academic" style, of which dense language is an important part. However, since the way dense structures are formed and used differs greatly across languages, it was frequently impossible to translate these structures in a comparable way into Finnish. At times, this resulted in the Finnish translations becoming clumsy and awkward.

> Translator C: *[Talking about the difficulty of translating scientific texts.] The more formal the genre and register, the more difficult it becomes.*
> I: *And the more dense the language is in the English text. Especially as it is possible to use [in English], for example, these verbless clauses. And that doesn't work in Finnish. And then it often leads to these quite incomprehensible [wordings].*

Another source of problems when translating the scientific texts was their specific, technical vocabulary. The terms were often difficult to understand or find exact and full equivalents for. This frequently forced the translators to do extra research or consult subject matter experts, for example. All this, of course, takes time. Also, it may lead to faulty or inaccurate translations.

> Translator A: *These terms and things … well, they had to be googled and searched in encyclopedias…. And I guess I also scanned dictionaries every now and then…. [Discussing a technical text.] There were some problems here with the terms….*
> I: *Yes, I even phoned to the Department of Biology so as to….*
> Translator A: *I think I also made some research on it…. I tried to ponder and checked and read a little about genes….*

I: *But this is the kind of detective work we have done here.*
Translator A: *Exactly. And even though you may find in some dictionary that it is a kind of "gene expression", then you still need to see whether it is the correct expression in this context.*

When translating the literary texts, again, the main problem was that because of the "context", or specifics of test translation (e.g., the strict requirements for equivalence in difficulty and the tight timeline), it was often not possible to maintain the stylistic and aesthetic flavor of the source text. This, in turn, meant that the Finnish texts could not be fully equivalent to their source texts and that the reading experience of Finnish readers was not similar to that of the source text readers.

I: *What often annoyed me was that when translating the literary texts, you would have liked to get more "touch" in them. But in this context, it is often impossible.*
Translator A: *Yes, in this context.... Of course you should be faithful to the source text and material so that you don't change it into something else. So that if it is a literary text, it would still at least remind of a literary text.*

All in all, however, the translators felt that the literary texts did not pose as many problems as they thought they would. The reason for this, as speculated by Translator A, was that, unlike the first PISA assessment (PISA 2000), this time the literary texts had been "simplified" so that the most problematic and untranslatable features and texts had been omitted.

Translator A: *But there weren't many such texts [highly literary texts] this time. I feel there have been more of these stylistically extreme texts before.... There have been texts that have been more literary, even more poetic.*
I: *Yes, poems are a case apart. Especially if they contain a lot of figures of speech. They can't be transferred across languages. But it seems that these have been avoided intentionally.*
Translator A: *And that's because it has been found out that they cause problems. I guess these texts are slightly simplified versions. That the worst, or not the worst, but perhaps the richest figures of speech have been cut down. And idioms and things like that.*

4.1.3 Discussion

The findings of this study confirm findings of previous research, showing that the source instrument is often one of the sources of translation problems and non-equivalent translations in international achievement studies. In practice, the problem is usually that the source instrument lacks translatability: Because of differences between languages, features in the source instrument often have no full equivalents in the target language. Also, technical terms, for instance, may be difficult to understand and may therefore require that the translator spend extra time to find their meanings. The terms may even be misunderstood and mistranslated. In this study, most problems were caused by dense language, differences in word order, technical terminology, stylistic factors and conventions. Question items were more problematic to translate than stimulus texts. Scientific texts, in turn, were more problematic than literary texts, mainly because the latter seemed to have been "simplified".

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 (cf. Alderson, 2000, p. 86). 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 (see also Allalouf, 2003; Allalouf, Rapp & Stoller, 2009; Arffman, 2007; Elosua & López-Jaúregui, 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 (see also Brislin, 1986, p. 144). This rules out not only long sentences, but even shorter sentences containing a lot of information packed in a dense form. Instead of using items consisting of one long and convoluted sentence, it is better to use items consisting of several short and simple sentences.

Generally speaking, the advice on translatability also largely applies when preparing stimulus materials in achievement studies assessing knowledge and skills in subjects other than reading literacy. It is good to make also the stimulus materials as translatable as possible, by not using, for example, vague meanings and long, dense and convoluted sentences. Technical terms cannot probably be fully avoided. However, when terms are used whose meanings may not be fully clear and unambiguous (because they have several meanings or because their meanings in the context are so specific, for example), it is good to ensure that they are correctly understood, by, for example, defining and clarifying them by means of translation notes. Clarifications such as these have already been used, for example, in PISA. However, it seems that they could be used even more. Also, it is important to remember that technical terms at times do not have full equivalents in other languages and that this may lead to non-equivalences in difficulty. For example, in Finnish, technical terms frequently need to be translated as more familiar and everyday words which are also easier to understand than the technical terms (Arffman, 2007).

However, 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. (E.g., Mullis, Martin, Kennedy, Trong & Sainsbury, 2009; OECD, 2009b.) The texts and language, moreover, need to be authentic, or conventional, in all the cultures involved. If only easily translatable texts and language are included in the tests, they will not be able to provide a true picture of the reading skills required of today's readers. For example, if no poems, metaphors and warranty cards (or other highly conventional and culture-specific text genres) are included in the tests, they cannot say much about the skills needed when reading these types of texts and language (see also Mullis, Martin, Kennedy, Trong & Sainsbury, 2009, p. 21; cf. Harkness, Edwards, Hansen, Miller & Villar, 2010, p. 46; and the different levels of translatability of different item types; Allalouf, Rapp & Stoller, 2009, p. 105). Thus, the more the texts in international reading literacy studies are made translatable, the narrower the skills and the construct of reading literacy they will be able to measure. This suggests that international reading literacy tests can never be made fully translatable and that they will continue to pose translation and equivalence problems also in future studies, more so than tests in other, more translatable subjects.

There are steps that can be taken to make the source instrument as translatable as possible. For example, it is good to make a better use of multicultural teams when designing the instrument and involve translators even more in the designing process than is done today (see e.g., Harkness, Edwards, Hansen, Miller & Villar, 2010). 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, pp. 35, 86–7, 2012). 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, p. 166.) However, the advance translation technique could be made even more beneficial, if it was complemented by cognitive laboratories, where testees proficient in the different source languages were asked to complete the different versions of the test and comment on them. (Advance translations followed by cognitive interviews were used in the OECD Programme for the International Assessment of Adult Competencies, or PIAAC. However, they were only used in three countries, France, Greece and Korea; for one instrument, the Job Requirements Approach instrument; and to eliminate cultural bias, not linguistic bias; see OECD, 2009 April.)

Nevertheless, it is evident that because of differences between languages, source instruments can never be made fully translatable (especially in reading literacy studies). Yet, 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, pp. 247–8). 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 is 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, p. 78). 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, p. 444). Yet, experience shows that time is usually missing. The great challenges related to the technique may be the reason why it has thus far not been used in international achievement studies.

4.2 Translators

4.2.1 Previous Research

Translation studies. Translation and equivalence problems may also be due to the translators lacking translator competences. The competences needed, of course, vary somewhat according to the goal of the translation task. However, the minimal requirement for any translator to be able to translate is good linguistic skills: To fully understand the source text, the translator needs mastery of the source language; and to produce a high-quality and natural target language text, s/he needs mastery of the target language (see e.g., Shreve, 1997, p. 122). In addition, however, the translator also needs subject matter knowledge, especially when translating special language texts (Ericsson & Kintsch, 1995; Kim, 2006; Process in the Acquisition of Translation Competence and Evaluation [PACTE], 2005, p. 610). Literary translation, again, requires artistic, or literary, skills (Wechsler, 1998; cf. Hassan, 2011). Finally, to be able to choose the right translation strategies, the best way to translate each text, and to solve translation problems, the translator needs translational and strategic knowledge – knowledge of the principles and theory of translation (PACTE, 2005, p. 610): S/he needs to know, for example, how the goal of the translation task and readers of the translation affect the way the text is to be translated.

Lack of any of these competences – language skills, subject matter knowledge, literary skills, and translational knowledge – easily lead to deficient translations, characterized by errors and excessively literal translations (Nord, 1997, p. 62). This is because the deficiencies "force" the translator to focus on micro level (e.g., word-level) factors and to translate literally and word-for-word (e.g., Dansk & Griffin, 1997, p. 171; Jensen, 2000, p. 166), instead of concentrating on larger and more important entities, such as textual, stylistic and pragmatic[4] factors (e.g., Englund Dimitrova, 2005a, pp. 14–5; Lörscher, 2005, p. 606). The deficiencies are also an obstacle when revising, because they easily "blind" the reviser to translation errors and problems (cf. Hayes, Flower, Schriver, Stratman & Carey, 1987, p. 233).

International achievement studies. All the above competences are also required when translating international achievement tests: language skills, subject matter knowledge, literary skills (when translating reading literacy tests), and translational knowledge. In addition to these, however, translators translating achievement tests also need knowledge of cognitive tests and the specifics of test translation. In this section, only competences of translators making the first drafts and reconcilers will be discussed. Competences of verifiers will be dealt with in section 4.5.

No studies have thus far been conducted on translators translating international achievement studies. However, reports and analyses on translations made in these studies suggest that deficient translator competences have caused translation problems also in these studies (Hambleton, 2002, 2005). For example, Arffman (2002, 2007) found semantic and grammatical errors in Finnish PISA 2000 materials and concluded that these were largely due to the translators having had deficient source and target language skills. Also, when translating materials for the IALS, Finland made the first drafts from both English and French. However, the reconciler had no knowledge of French, and s/he commented that this was a problem especially when the two target versions were very different, because

[4] Pragmatics is "the study of how utterances have meanings in situations" (Leech, 1983, p. X). It has to do with how people produce and comprehend meanings in real situations; what words and sentences really mean when uttered, their intended meanings. For example, the pragmatic meaning of "it's cold in here" might be "Would you, please, shut the window" or "Would you, please, turn up the heat."

s/he had no way of knowing what really was said in the French source versions and to what extent the Finnish version made from it resembled it (P. Linnakylä, personal communication, November 14, 2008). Likewise, when Karg (2005) examined German PISA 2000 and 2003 materials, she detected several mistranslations and interpreted these as suggesting that the translators had lacked subject matter knowledge. She also found several points in the translations where she felt the translators had not taken sufficiently into consideration stylistic factors and German readers – suggesting that they lacked translational knowledge. Along the same lines, when Arffman (2007) found excessively literal renderings and cases of interference (undue influence by the source language) in Finnish PISA 2000 translations, she concluded that these were largely due to most of the Finnish translators not having had training in translation theory and strategies. Finally, experience and reports suggest that translators in international achievement studies have often lacked familiarity with cognitive tests and test translation (Hambleton, 2005, p. 25), which, however, is a precondition for them to be able to ensure equivalent test versions.

4.2.2 Findings of This Study

Deficient translator competences were also one of the factors that the translators in this study mentioned as causing translation problems and non-equivalences in international achievement studies.

Language skills. The first competence where the translators felt there have been deficiencies was knowledge of the source language(s). As suggested by Translator C, this was the case in the first PISA study (PISA 2000), in particular, where Finland only made its translations from English and where knowledge of French was not required. However, Translator C suggests that at least some of the translators also lacked a good command of English. This time, however, s/he felt that the translators had better language skills, not only in English but, even more specifically, in French. According to the translator, the better language skills – especially the command of French of the first reconciler – in turn, helped to improve the quality of the translations.

> Translator C: *It is absolutely necessary to have … [in the Finnish translation procedure] a person who really has a good command of the source language. In this case you [the first reconciler] have a good command of English. Because even though I use English quite a lot at work, my language skills are not nearly enough to be able to distinguish between nuances.... I feel that this is a part which was perhaps missing when we were translating PISA 2000 [materials]. I also feel that this has improved the quality of these translations considerably. Besides, you even made the effort to have a look at those French texts, which ... often offered a better basis for our Finnish translations.*

Subject matter knowledge and literary skills. Other competences where the translators felt there may have been deficiencies were subject matter knowledge and literary skills. Both competences are needed when translating reading literacy tests: knowledge of a great number of subjects – much more so than when translating other tests – because texts in reading literacy tests can deal with almost any field and topic (e.g., technology, biology, medicine, history, economics); and literary skills, because reading literary texts is an important part of all reading literacy tests. Extremely broad knowledge and expertise is thus required of translators translating reading tests. However, the translators in this study felt that it is not reasonable to expect that any one translator could have all that knowledge and expertise. Rather, for instance, in Finland, translators usually specialize in only one field. Also, for example, Translator A admitted that s/he was much more used to translating expository texts than literary texts. The lack of subject matter knowledge or literary skills, in turn, complicates the translation work (e.g., compelling the translator to spend extra time consulting e.g., reference works and subject matter experts) and has a negative impact on the quality of the translated texts (e.g., mistranslations, unduly literal translations, aesthetically and stylistically deficient translations).

> Translator C: *There is still this issue concerning the translation of these different [PISA] domains: That in mathematics you remain in the field of mathematics. And in science, too, the topics have to do with science. But in reading literacy texts, they can vary from one extreme to another. And yet, you can't be an expert in every field…. It can be well seen from [these two scientific texts] that it is so laborious and difficult simply because of this. [Relates how she had to verify a term in a scientific text by consulting an expert.] So I think this is one of the things that add to the challenge of translating reading literacy texts.*
> I: *That's true. You should have expert knowledge of so many topics.*
> Translator C: *Yes. You should be a biologist and what have you, to be able to handle just these few texts.*

> Translator A: *There were texts of so many different types and styles there…. [Discussing an expository text.] Quite ordinary stuff, something I translate every day…. [Discussing a literary text.] A very literary text, representing French literature, which I am not really familiar with.*

Translational knowledge and familiarity with test translation. Still other competences where the translators felt there may have been deficiencies were translational knowledge and familiarity with test translation. For example, one of the two translators making the first Finnish drafts appeared to focus on close formal equivalence to the source text and translated rather literally. The other translator, again, put more emphasis on idiomatic and fluent target language, translating much more freely, sometimes even at the expense of the translation instructions and equivalence in difficulty.

> Translator C: *It could be seen from the first versions .... that they were slightly different translators. It was easy to see that one [the translator with academic translator training but no experience in test translation] translated more freely and the other [the translator who had attended the training provided for translators in PISA but had no academic translator training] was more concerned with equivalence. [The latter] was more faithful to the source text. On the other hand, [the former]'s translations, when you think of them as Finnish texts, were somehow better.*
> I: *Exactly. But maybe even too free.*
> Translator C: *That's right, sometimes even too free. Because there may have been something in these reference chains [which according to the translation instructions had to be kept unchanged, e.g., pronouns were to be translated as pronouns] which made me think that, well, wasn't there a pronoun here in the source text [whereas the translator had translated it as a noun].*

The differences seemed to be related to how much translational knowledge the translators had and how familiar they were with the specifics of test translation. The translator translating more faithfully had attended the special training provided for translators in PISA (in 1999, when the emphasis was strongly on close one-to-one correspondence between the source and target version) and had several years experience in test translation. S/he therefore seemed to be very well acquainted with the specifics of test translation (e.g., whether or not there can be a synonymous match between the stimulus text and a question item). However, s/he had had no academic translator training. Therefore, s/he seemed to be less versed in the general principles and theory of translation (e.g., different types of equivalence, translation strategies and techniques, translation units). On the other hand, the translator translating more freely had received university education in translation and therefore had a good knowledge of the principles and theory of translation. However, s/he had had no training and experience in test translation and was therefore not well acquainted with it.

4.2.3 Discussion

The findings of this study agree with those of previous studies, showing that there have been deficiencies in the competences of translators translating international achievement tests and that these have caused problems when translating the tests and making them equivalent. Deficiencies have been found in language skills, subject matter knowledge, literary skills, translational knowledge, and familiarity with test translation. The deficiencies appear to have been most striking during the first cycles. However, the study suggests that competences may have also been lacking during more recent studies, which may have complicated the translation task and jeopardized equivalence. More decisive steps are therefore needed to ensure that translators in international achievement studies are fully competent.

In practice this means above all that more care is needed when selecting and hiring the translators (cf. e.g., Hambleton, 2001, p. 166; 2005, pp. 24–5; 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, pp. 12–5). For example, 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. For example, if the two target versions are very different from each other, s/he can verify that they are acceptable translations. In addition, s/he will have more alternatives and hints as to how to formulate the target versions.

Also, translators translating factual texts should have knowledge of the subjects concerned. This should be the case not only, for example, in mathematics and science tests, but also in reading tests. 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, p. 166; 2005, pp. 24–5) – 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. This typically requires that they are professional translators with academic translator training.

However, as shown by the in places excessively free and non-equivalent translations of the translator who produced the smaller number of the first Finnish versions, not even being a fully competent, academically trained professional translator is enough. 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).

To ensure that countries know what kind of translators to select and hire, it is good to partly revise the requirements for translators (e.g., reconcilers and verifiers working on texts translated from two source languages should be proficient in both languages; translators of literary texts need to have experience of literary translation).

4.3 Translation Goal, Translation Guidelines and Translation Notes

4.3.1 Previous Research

Translation studies. The translation goal and translation instructions play a significant role in deciding how difficult a text is to translate (Reiss & Vermeer, 1984; Nord, 2006) – the goal, because it governs the entire translation work and determines how a text is to be translated (Reiss & Vermeer, 1984; Vermeer, 1989); and the instructions, because they are the means by which translators are typically informed about the goal and because they often also provide directions on how (e.g., how literally or freely) to translate in order to reach the goal (e.g., Hatim & Mason, 1997; Vermeer, 1989).

Translation goals differ in how difficult they are to translate and attain. For example, when the goal is to make an idiomatic and natural target language text, translation is much more difficult than when the goal is to make a literal, or word-for-word, rendering. This is because when translating, it is normally the literal rendering that first – and almost automatically – comes to mind (Englund Dimitrova, 2005a; Tirkkonen-Condit, 2005, pp. 407-8.). However, when aiming at idiomatic translation, the translator has to distance himself or herself from the literal translation and search for more idiomatic renderings. All this requires extra cognitive effort and time. (Englund Dimitrova, 2005a.)

In addition, translation problems easily arise, if the goal is vague, elusive and difficult to grasp or if the instructions (are vague and) do not state clearly and unequivocally what the goal is and how the translator is to translate to attain it. In each case, the translator is left uncertain as to how to translate, and the uncertainty, in turn, easily tempts him or her into "playing safe" and resorting to the default translation strategy – literal translation (Pym, 2008, p. 324; see also Chesterman, 2010, p. 42; Jensen, 2009; Tirkkonen-Condit 2005, 407–8; Vinay & Darbelnet, 2004).

International achievement studies. When translating international achievement tests, the goal is to make all the translated and different-language versions of the test equally easy to respond to. Most achievement studies have also provided translators with translation guidelines to inform them about the goal and to help them to attain it (no written guidelines, nevertheless, seem to have been provided e.g., in the Latin American Laboratory for Assessment of the Quality of Education, or LLECE; however, translator training has been offered). However, the type and amount of the guidelines has varied considerably between the studies. For example, in PISA, translators have been provided with both general translation guidelines and item-specific translation notes The general translation guidelines have contained a great number of detailed linguistic (e.g., lexical and syntactic) translation instructions, intended to help translators to better understand the translation goal and how to translate in order to reach the goal (e.g., OECD, 2007). Since 2009, moreover, PISA has prepared customized guidelines for translation into Arabic and Chinese. The translation notes, for their part, have concentrated on item-specific translation problems. However, for example, in studies conducted by the IEA (e.g., PIRLS, TIMSS), often only general translation guidelines have been provided. The guidelines, moreover, have contained only a very few specific linguistic instructions. Also, more weight has been put in the guidelines on idiomatic target language. (Johansone & Malak, 2008; see also Dept, Ferrari & Wäyrynen, 2010, p. 164.)

Studies concerning or even touching on the translation goal, translation guidelines or translation notes (or translation annotations; see Behr & Scholz, 2011, pp. 159–60) and how they have affected translators and translating in international achievement studies have been extremely rare. However, the few studies that do exist and the experiences gained when translating the tests suggest that the goal and notes but especially the guidelines may have been a significant source of translation problems in these studies.

For example, Pan, Kleiner & Bouic (2007; see also Kleiner, Pan & Bouic, 2009) conducted a study on the use of translation instructions and annotations in questionnaire translation and found that both had an impact on translation. However, the impact varied (in a yet unknown way) across languages and was sometimes advantageous and sometimes disadvantageous.

As regards the translation guidelines, findings suggest that, for example, those used in PISA (e.g., OECD, 2007) have often not been able to give a clear and unequivocal picture of the translation goal and that this may have been because of the numerous detailed linguistic translation instructions included in them. For instance, when analyzing Finnish PISA 2000 field trial materials, Arffman (2002), found several cases of unduly literal renderings in them and interpreted these as being largely due to the translation guidelines having been vague and contradictory: On one hand, the guidelines said that the goal was to make translations that would be equally difficult to respond to as their source texts and that this required that the translations use natural and idiomatic languages. Mainly, however, the guidelines consisted of detailed instructions on how to remain lexically and syntactically as close to the source version as possible, thus, in practice, fostering literal translation. Similar observations were also made when analyzing Finnish PISA 2000 main study translations (Arffman, 2007).

There are also findings suggesting that the specific linguistic translation instructions – and the fact that they have focused so strongly on close linguistic resemblance to the English and French source versions and have been written mainly from the point of view of English and French – may have been one of the reasons why the quality of certain non-Indo-European (e.g., Middle East and Asian) test versions has not been as good as it should have been (Grisay, de Jong, Gebhardt, Berezner & Halleux-Monseur, 2007, p. 265).

On the other hand, problems appear to have arisen also when the guidelines have contained only a very few linguistic instructions. For example, when the TIMSS 2011 materials were translated into Finnish, the Finnish translators felt that the translation guidelines did not contain enough instructions so as to be able to help them in the translation work. They therefore asked for (and were granted) the opportunity to use the PISA guidelines.

Sometimes, in addition, problems seem to have been caused by the requirement for the target versions to be natural and authentic. Typically, the result has been unduly free, explicit, transparent and/or straightforward translations (see e.g., Hambleton, 2001, p. 166). For instance, when analyzing Finnish PISA 2000 materials, Arffman (2007) found that, in an attempt to make the Finnish versions natural, Finnish translators sometimes improved and explicated the versions by adding grammatical words or by using more concrete expressions. Usually, the reason appears to have been that the translators did not realize how much even seemingly small linguistic changes may affect item difficulty. Also, in the study by Pan, Kleiner & Bouic (2007), some of the wordings in the translated questionnaires ended up being extremely far from those in the source versions. This, again, seems to have been because the instructions were so vague and so strongly encouraged authenticity (e.g., "Feel free to make whatever changes necessary to accomplish this [that questions are phrased in a culturally appropriate way]"; ibid., p. 7; see also Behr & Scholz, 2011, p. 175).

Apart from there having been problems when using the general translation guidelines, findings in questionnaire translation show that there have also been problems when using translation notes or annotations (Behr & Scholz, 2011). One of the main problems seems to have been that it has not always been easy to decide when to use the annotations and when not to use them. Sometimes too much appears to have been expected of the annotations. For example, the source instrument may have contained an idiomatic expression, and an annotation may have been used to define and clarify the meaning of the expression and to tell translators to use a comparable expression also in the target version. However, idiomatic expressions are strongly language-specific and very seldom have full equivalents across languages. An annotation cannot guarantee equivalence when differences between languages make equivalence impossible.

4.3.2 Findings of This Study

In this study, too, the translation goal, translation guidelines and translation notes were mentioned among the factors that have caused translation problems in international achievement studies.

Translation goal. The translation goal was felt to be vague. This, as pointed out by Translator A, was because especially at the time of translating, the translators have no way of accurately assessing and determining the difficulty of the source and target versions. They have no way of knowing – being certain – how difficult the source and target texts really are.

> Translator A: *We are on shaky ground here, because actually we should know how a 15-year-old English student understands these, when we don't know how Finnish students understand them.... So it is in fact just an illusion and speculation as to what might be the case. At least I don't have the capacity to assess how an English student really understands it.*

> Translator C: *It is always a matter of judgment … this equivalence. I mean, whether this is equivalent or not. No data are available on it until after the field trial. And even then, it doesn't necessarily show. You just have to rely on your own judgment, as it were.*

General translation guidelines. The translation guidelines, for their part, were felt to put too much emphasis on faithfulness to the source text. This was felt to be the case during the earliest PISA studies, in particular, where the requirement for close formal resemblance caused problems especially when translating into languages whose lexical and syntactic structures differed most from those of the source languages (e.g., Russian). The translators felt that today the situation was somewhat better, in that in the guidelines slightly more attention was paid to natural target language. This, in turn, made it easier for them to distance themselves from the source version and to translate more idiomatically.

> Translator A: *These PISA guidelines have developed since the first round, when they had to be read like the Bible…. We were given these "Bibles" as it were: "This is what you will have to follow to the letter." And at that stage, there was not much room for negotiating. That "would it perhaps be possible in our language…?" "No, no, no, no." There were all, Russians and all, astonished, because "in our language we don't have something like this".*

> Translator C: *What was something new and most welcome there [in the PISA 2009 translation guidelines] was that, on one hand, there was this emphasis on equivalence, as always, but now they also highlighted this fluency and naturalness…. On the basis of the earlier translation guidelines … you could say that translating has been quite literal, word-for-word. But now they give a somewhat different picture…. I feel that all in all this time we had much more latitude than before. And this contributed to greater fluency.*

However, it seems that despite the improvements, the guidelines have still not managed to provide translators with a clear picture of how to translate. For example, the translators in this study appear to have felt that the emphasis in the guidelines is still strongly on faithfulness and close formal correspondence to the source text. This, in turn, has forced them to translate literally and accept renderings which they have known have not been the best and most Finnish choices and which they would not have accepted in other contexts. At the same time, the fluency and naturalness of the Finnish versions have had to be compromised.

> I: *These translation instructions, as they are today [containing a lot of detailed instructions], … inevitably lead to too word-for-word a translationese [artificial language used in translations] style….*
> Translator C: *I felt that, in the end, what you can't compromise is equivalence [faithfulness to the source version]. That's what you had to prioritize. And this often led to there being only two poor alternatives. If we didn't need to care about equivalence, we would have good choices. It is this equivalence that you simply dare not compromise.... What you sometimes have to compromise is the fluency of the Finnish text.*

> Translator A: *I often thought that if I was doing this directly in Finnish, it would be slightly different. But since you have to bear in mind all these requirements...*

In a similar vein, it seems that the guidelines have not managed to make it clear to translators how freely they can translate. For example, it seems that the translators in this study would have liked to translate more freely, but due to a lack of clear instructions, did not dare to do this.

> I: *So, how freely can you translate?... Would it be possible to translate more freely? Where is the limit? It is awfully difficult to know in this context.*

On the other hand, as shown by the overly free and explicit translations of one of the two translators making the first Finnish drafts (see section 4.2.2), sometimes the translators even translated too freely.

Translation notes. The item-specific translation notes also caused some problems. For example, in the source instruments, slightly different words or expressions (i.e., not the same words and expressions, but e.g., synonyms) were often used in the stimulus text and a question item to refer to the same phenomenon, and a translation note was frequently used to tell translators that they were to use comparable words or expressions (e.g., synonyms) also in the target version. Or the same syntactic structure was employed in several or all the alternatives of a multiple-choice item, and a translation note told translators to use a comparable structure also in the target item. However, often there were no such synonyms or structures in Finnish. In cases such as these, the translators thus had to spend extra time to try to find, for example, expressions or structures which did not exist. Or they may have felt that because of the translation note, they had to accept artificial and clumsy expressions or structures. In both cases, the target version would have ended up being non-equivalent with the source version.

> Translator D: *This [there being a translation note telling that a certain term or structure cannot or has to be used both in the stimulus text and in the question item], I guess, was often what was the most difficult thing. When an expression is bound to a certain word or structure or synonym, it may no longer be natural (or even possible) in Finnish.*

Another problem with the translation notes was that they were not used consistently. For example, usually when the source instrument used different expressions in the stimulus text and a question item, a translation note was employed to tell translators that different expressions had to be used also in the target versions. A similar direction had already been given in the general translation guidelines (OECD, 2007, p. 18), so the note mainly served as a reminder. However, sometimes reminder notes were not used. This was confusing and left the translators wondering whether in these few cases the direction did not apply. The inconsistency in the use of the translation notes thus complicated the work of the translators. In addition, it may even have resulted in them not following a direction when it should have been followed – and, consequently, in the versions being non-equivalent.

> I: *Then there is this lack of translation notes. I think they didn't always say if the same term or a different term had to be used in the stimulus text and the question item. And sometimes they said. This was a little confusing. This should be consistent…. For the sake of logic, it [the translation note] should be there. Otherwise you may start to hesitate that in this case it [following the general translation instruction] is not needed.*
> Translator A: *That's right. Usually you'd better play it safe…. So that you can rely on it. It would also make working much easier, because then you would not need to wonder whether a translation note is really missing here or whether it is missing on purpose.*

4.3.3 Discussion

The findings of this study coincide with those of previous studies, indicating that the translation goal, translation guidelines and translation notes have been among the factors causing translation and equivalence problems in international achievement studies. With regard to the translation goal, this study suggests that it is, by necessity, always vague, because the translators have no way of knowing how difficult the source and target texts and items are. The goal is thus exceptionally demanding, which, in turn, means that high requirements are set for the translation guidelines, for example.

The guidelines, however, have differed between achievement studies. In PISA, the guiding principle when preparing the guidelines seems to have been to make the goal as clear and easy to grasp as possible. To this end, numerous detailed linguistic translation instructions have been included in the guidelines. However, as suggested by previous studies and confirmed by this study, this has resulted in the guidelines putting great emphasis on faithfulness to the source text and on literal translation (see also Arffman, 2002, 2007). This, in turn, has been a problem especially when translating into languages furthest away from English and French (cf. Grisay et al., 2007). The guidelines have also failed to make it clear how freely the tests can be translated. In recent years, improvements have been made to the guidelines, so that, for example, slightly more attention is today paid to natural and idiomatic language than before. Also, customized guidelines have been prepared for translation into certain non-Indo-European languages. However, it seems that the right balance between detailed linguistic translation instructions and the requirement for natural and authentic target language has not yet been found (cf. Kleiner, Pan & Bouic, 2009).

Finally, as concerns the translation notes, the findings of this study agree with previous findings, suggesting, however, that when translating international achievement tests, the most severe problem may be that the notes are sometimes expected to guarantee equivalence when it cannot be guaranteed, because of differences between languages and cultures (see also Behr & Scholz, 2011, p. 171). Also, this study confirms that it is not always easy to decide when to use the notes (ibid., pp. 172–3). In this study, for example, notes were sometimes missing, when they would have been needed.

Adjustments thus seem to be needed in the translation guidelines and notes. For example, it seems that in the guidelines some linguistic instructions are needed so as to make it clear what the translation goal is and how translators can assess the difficulty of the items and texts and make the items and texts as equivalent in difficulty as possible. However, should the number of the instructions perhaps be smaller than it is in PISA today, as suggested, for example, by Dept, Ferrari and Wäyrynen (2010, p. 165)?

Research in translation studies shows that translators have a universal tendency to translate literally (Toury, 1995, p. 275). Therefore, whenever detailed linguistic instructions (emphasizing faithfulness to the source version) are given, it is important to make sure that they do not lure translators into translating overly literally. To this end, it is good to underline in the guidelines, that one of the main prerequisites for equivalence in difficulty is that the target versions are in natural target language (see e.g., the translation guidelines for ESS; 2010, pp. 23–5, 28; cf. Jeanrie & Bertrand, 1999, p. 279). At the same time, strong warnings are needed against unduly literal translations (ESS, 2010, pp. 37–8), accompanied by examples of such translations and how they affect reading and comprehension. When the target language is very far from the languages from the point of view of which the guidelines have mainly been written (cf. Behr & Scholz, 2011, p. 175), it may even be necessary to prepare separate, customized instructions (as in PISA).

However, in the guidelines, warnings are also needed against overly free and straightforward translations (see also Hambleton, 2001, p. 166; Harkness, 2003, p. 46). It is necessary to emphasize that test translation differs from other types of translation, in that it is less free. For example, improvement, explicitation and simplification – all, again, universal tendencies among translators (Laviosa-Braithwaite, 1998; Séguinot, 1989) – are not allowed. Reminders are also needed that even apparently insignificant linguistic modifications may sometimes improve, explicate and simplify texts and bring about unwanted changes in difficulty. (Arffman, 2007.) All these adjustments in the translation guidelines also need to be reflected in the training provided for translators.

To improve the use of translation notes, one of the keys is to remember that the notes cannot remove differences between languages: The notes cannot guarantee equivalence, when languages do not have equivalent ways of expressing meanings. In cases such as these, the only cure is to make changes to the source instrument (see also Behr & Scholz, 2011, p. 172), to render it more translatable (e.g., Brislin, 1986, pp. 143–9; see also section 4.1.3). This may even mean avoiding certain types of tasks, such as those relying strongly on language form (e.g., similar syntactic structures in the alternatives of a multiple-choice item) or meaning (e.g., synonyms, idioms; see also Allalouf, Rapp & Stoller, 2009).

On the other hand, when translation notes are used, they need to be used consistently. Attention also needs to be paid to their wording. For example, if a note says that different expressions should be used in the stimulus text and a question item, the wording must be cautious and acknowledge that finding different expressions, and thus following the direction in the note, may not be possible in every language. Translators should not feel forced to accept artificial and clumsy renderings, when the problem is, in fact, in the source instrument and its untranslatability.

To be able to write translation notes that really help translators to produce equivalent test versions and to be able to avoid notes that may even be harmful, knowledge of translation and factors causing translation problems is needed (see also Behr & Scholz, 2011, p. 173). In practice this necessitates involving translators in designing the test. The translators, moreover, should represent various cultures (see e.g., Harkness, Edwards, Hansen, Miller & Villar, 2010). Advance translation can also help to see whether changes need to be made to the source instrument or whether translation notes could help (cf. Dorer, 2011).

All in all, however, more research is needed on the translation guidelines and translation notes used in international achievement studies. For example, what is the ideal number of linguistic translation instructions to be included in the guidelines (cf. Dept, Ferrari & Wäyrynen, 2010, p. 165)? How can they be presented in such a way that they do not drown out the need for natural target language? How do translation notes that cannot be followed (because of language-specific differences) affect translators (cf. Behr & Scholz, 2011, p. 175)? Do translators disregard them, because they view them as impossible to follow (thus advocating natural target language)? Or do they try to follow them as far as possible, thereby making the target versions at times artificial and clumsy?

4.4 Parallel Source Versions

4.4.1 Previous Research

Translation studies. The availability and use of reference materials can also have an impact on how difficult the translation task is and what the quality of the ensuing translation is. Reference materials (e.g., dictionaries, encyclopedias, parallel texts) are typically used to facilitate translating and to make better-quality translations (Nord, 1991). Making use of more than one different-language source texts can serve the same purpose. The different words, structures and nuances used in the various texts provide extra clues that help the translator to better understand the source material and to convey its meaning more accurately. They also help the translator to see that typically there is not just one but several acceptable ways for expressing one and the same idea. This, in turn, can encourage the translator not to translate literally but idiomatically.

However, using several source texts also involves problems. Firstly, it takes time. When several source texts are used, it means that ideas and/or extracts from all the texts need to be compared to each other and merged, or reconciled, into one text. Moreover, it cannot be assumed that the resulting text would automatically be correct, coherent, harmonious and in good language. Rather, a careful revision and harmonization of the texts is also needed. All these tasks, if performed properly, are extremely time-consuming. Secondly, if the source texts are in different languages, the one doing the merging, the reconciler, should be proficient in all the languages so as to be able to make the most of the texts. If s/he is not, his or her decisions will rest on a shakier ground. Thirdly, the different source texts may increase the risk of non-equivalent translations. This risk, of course, is the greater, the more different, or non-equivalent, the source texts are. Since non-equivalences are typically larger between languages than within them – full equivalence hardly ever existing between languages (e.g., Chesterman, 1997; Pym, 1995) – the risk is especially great, when the source texts are in different languages.

International achievement studies. In most international achievement studies, countries make their translations on the basis of only one, English, source version. In the IALS, however, a few countries (e.g., Finland) based their translations on two different-language (English and French) source versions. Since then, double translations from two source versions have only been made in PISA. In addition to this, however, some countries have used the two source versions in slightly more unorthodox ways. For example, Finland has typically translated the materials from English and then only checked them against the French versions. On top of this, some studies have provided the instrument in two or three languages (e.g., TIMSS and PIRLS in Arabic; and LAMP in French and Spanish), but each country has based its translations on only one of these versions. In the following, the focus will be on the procedure recommended in PISA, double translation from two different-language source versions. At times, however, the discussion also applies to studies where countries choose on which of the several parallel source versions they base their translations.

Only some scattered findings exist on how double translation from two different-language source versions has affected translation and the quality of the translated test. All in all, however, the findings seem to be in line with those gained in translation studies, suggesting that the procedure has both its pros and cons and that it can both help to prepare higher-quality and more equivalent translations and increase translation difficulty and the risk of non-equivalence (cf. Grisay, 2003). For example, when analyzing Finnish PISA 2000 field trial and main study translations, which were made almost exclusively on the basis of the English versions, Arffman (2002, 2007) discovered that words that had several meanings were sometimes mistranslated and that the Finnish versions at times contained overly literal and clumsy translations. S/he concluded that the mistranslations and unduly literal translations could largely have been avoided, if more extensive use had been made of the French versions when translating the materials. Also, verifiers in the PISA 2000 field trial reported that those translations that had been rendered from both source versions contained fewer mistranslations, fewer unduly literal translations and fewer flawed items than those that had been translated from only one source version (Grisay, 2002). Moreover, when Grisay (2002, 2003) calculated the proportion of flawed items in these versions, she found that those translations that had been rendered by using both source versions – especially by double translation from the two languages, but also by using one of the source versions for double translation and the other for extensive cross-checks – contained significantly fewer flawed items than those translations that had been rendered from only one source version. (As for these latter findings, however, it should be noted that the calculations were made on the basis of reports sent by the participating countries and that these reports were not always quite accurate. For example, Finland was classified in Grisay's study among countries having double-translated its materials from English with extensive cross-checks against French. In reality, however, the materials were single-translated from English with very few cross-checks. Nonetheless, Finland was one of the countries where the quality of the translations was deemed to be very high.) In PISA 2003, too, in a similar comparison, double translation from two languages was found to be the best method (OECD, 2005).

Another significant advantage of preparing two or more parallel source versions has been that the translation of one or some of the source versions has served as advance translation (e.g., Dorer, 2011; Grisay, 2002, p. 60; Harkness, Edwards, Hansen, Miller & Villar, 2010, p. 41): It has helped to spot and correct errors and ambiguities in the source versions and to anticipate translation problems, which, in turn, has helped not only to make the source instrument more translatable, but also to revise the translation guidelines and to provide helpful translation notes (Dept, Ferrari & Wäyrynen, 2010, p. 166; Grisay, 2002, p. 60; cf. Harkness, Villar & Edwards, 2010, p. 131). All this makes translation easier and contributes to higher-quality translations (cf. Dorer, 2011, p. 24).

However, there have also been problems with the practice of making two translations from two source versions. This is suggested, for example, by the fact that several countries have not followed this procedure. For example, in PISA 2000, only 6 countries reported having made double-translations from two languages, whereas 19 countries said they applied some other method (e.g., translation from only one language) and 9 countries commented having used versions adapted from either of the source languages (Grisay, 2002). In PISA 2003, the recommended procedure was used in 15 countries, an alternative method in 16 countries, and adapted versions (from one of the source languages or from another common, e.g., German, version) in 24 countries (OECD, 2005). In PISA 2006, the recommended procedure was followed in 16 countries, an alternative method in 30 countries, and 41 countries used adapted versions (OECD, 2009a, p. 91). And in PISA 2009, the recommended procedure was followed in 27 countries, an alternative method in 30 countries, and 44 countries used adapted versions (OECD, 2012). Again, the true number of countries not having followed the recommended translation procedure may have been even higher. Remember, for example, that at least Finland was misclassified in PISA 2000 and 2003. Also, *PISA 2003 Technical Report* (OECD, 2005, p. 76) mentions that even though no country reported applying, for instance, the single-translation method, the verifiers felt that the method had been used. One national version even seemed to be a "carelessly reviewed computer translation" (p. 80).

One problem with using the two source versions in the recommended way, and a possible reason for the reluctance of countries to follow the recommended procedure, has concerned resources. For countries to be able to make efficient use of the two source versions, they should be able to hire translators who can translate from English, translators who can translate from French and, more specifically, reconcilers and verifiers who can translate from both languages. However, in many countries finding such persons – at reasonable cost – has not been possible (B. Halleux-Monseur, personal communication, January 24, 2008).

However, the most serious problem whenever there are more than one source versions is that they can never be fully equivalent to each other. When, for example, Grisay (2004) and Karg (2005) made linguistic comparisons between English and French PISA 2000 materials, they found that the versions differed in key vocabulary, precision of terms, and the link between the stimulus texts and items. Similar results were also obtained for the IALS instrument (see e.g., Blum, Goldstein & Guérin-Pace, 2001). Moreover, when Grisay (2002, 2003) compared the length (number of words in the stimulus texts) and linguistic complexity (measured by means of readability formulas) of the English and French stimulus texts used in the PISA 2000 field trial, she discovered that the French texts tended to be significantly longer and that this, in turn, (modestly) increased the difficulty of some French items. Also, when she compared the psychometric quality of the national versions adapted from the English source versions and those adapted from the French versions, she found that even though there were no significant differences between the groups in the number of flawed items, a few individual items functioned differentially across the groups. Similar results were also obtained in PISA 2003 (OECD, 2005). Finally, when Grisay and Monseur (2007) compared the proportion of DIF (differential item functioning) items within (e.g., in all various English versions) and between language groups (e.g., between all English vs. French versions) in the PISA 2000 main study, they discovered that it was considerably higher between groups. Their conclusion was that whenever tests are translated across languages – from English into French, for example – part of equivalence is always lost. Lack of equivalence between the source versions is, of course, a serious problem, because non-equivalent source versions almost inevitably lead to non-equivalent translations.

4.4.2 Findings of This Study

The translators in this study also mentioned both advantages and disadvantages when using the two parallel source versions.

Level of freedom and words with several meanings. Among the advantages of using parallel source versions, the translators mentioned that the versions helped them to decide on the degree of freedom allowed in the translations. This, in turn, made it possible for the translators to translate more freely and to avoid excessively literal translations. Also, using both the source versions made it easier for them to find the exact, intended meanings of words with several meanings.

> Translator C: *I feel it was an extremely valuable thing to make use of those French texts, because they helped to see the level of freedom.... But I wouldn't have dared to do that [translate more freely], if we, sort of, hadn't had the French translation bearing the status of an official source version. There were often so great differences between the English source text and then the one which was based on the French text that, well, on my own I wouldn't have dared to make so great deviations.*
> I: *Yes, that's one of the great values that those French texts have…. And another aspect where they, in my mind, help … they helped to clarify meanings. For example, if there was a word or a structure somewhere that may have had more than one meaning, the French text then gave a clue that this is the meaning that is looked for here.*

In contrast, not using both the source versions was felt to have a negative effect on translation and equivalence. This, as suggested by Translator C, was seen in that during the first

PISA round, when the French versions were not employed (except in a very few cases), the quality of the Finnish translations was not as good as s/he thought it was now.

> Translator C: *[In PISA 2000], when we didn't have the support of the French texts, they were certainly, well, more cautious translations. Now we have perhaps been able to make them more fluent, since we had a look at those French texts.*

Increased workload and time pressure of the reconciler. Among the problems when using the two source versions, the translators mentioned that the first reconciler had to perform so many time-consuming tasks that she did not have enough time to do them properly. She had to compare all translations not only to the English but also to the French source versions and verify that they were terminologically and contentwise equivalent to the source versions. In addition to this, she often even imaginarily translated the materials from French into Finnish, so as to ensure that a good use was made of the French source versions. Finally, she revised the language of the Finnish versions, making sure that it was fluent and natural.

> Translator C: *You [the first reconciler] had such a horrible amount of work at that time....*
> I: *It takes such a lot of time when you also have a look at those French versions.*
> Translator C: *Of course it does. All this revising and checking, it is so slow. Especially when you have many versions. You have the English one, and then you have the translation, and then there is the French version. So, it's clear that it's much slower than if you only stuck to the English version.*
> I: *And especially as, in a sense, I tried to think…. Because we should have this "double" thing [translation] here. Every now and then I tried to think that if I translated this anew, as it were … in a way, I tried to forget the translation that had come [the first draft] and think how I would translate it tabula rasa, as it were, on the basis of the French version.*

Knowledge of the source languages of the reconciler. As another problem connected with the use of the two source versions, the translators mentioned that it may be difficult for countries to find reconcilers who would have so good a knowledge of both the source languages that they would be able to make a proper use of both source versions. This may also have been the reason why in PISA 2000, Finland only used the French versions for some very few minor checks. This time, however, the first reconciler was competent also in French, and this rendered it possible to make a more extensive use of the French source versions. This, in turn, was felt to improve the quality of the Finnish translations.

> I: *I have thought that all [countries] should be obliged to use both [source versions].*
> Translator C: *Yes. At least as this kind of referee…. And this may be difficult in practice. And this may be the reason they may have been reluctant to [observe] the obligation … Finding such people…. It may quite frankly be difficult to find such people [competent in both source languages]. It is good that we [in Finland] – now that the French versions have come in to a larger extent – have, all in all,*

had improvement. In 2000, I checked a few things in the French versions. Well, it was something like whether "you" refers to one or many. Because it can't be seen [from the English version]. So I had a look at the French version. But they were just a very few rare cases.

Differences between the two source versions. As a third problem with the use of the two source versions, the translators mentioned differences between the versions. These became especially apparent during the verification phase. Finland had used both source versions when translating the PISA 2009 materials. Often the versions were not completely identical. When this was the case, Finland frequently based its translations more on the French versions. However, when verified, the translations were only checked against the English versions. Therefore, some of the renderings that were already equivalent to the French versions but deviated somewhat from the English versions were "corrected" so as to bring them closer to the English versions. This caused much confusion to the first reconciler, whose task it was to check the corrections and suggestions made by the verifier. Also, it forced her to spend extra time to check the translations.

Translator C: *Often it was the translation that had been made on the basis of the French version that was better suited to the Finnish language…. So, I think it was extremely beneficial to make use of the French versions…. But well, how equivalent are these two official source versions in the last resort? This is the question that came to me.*
I: *Really, especially during the verification phase….I noticed what happened in the verification. That they only check against one of the source versions….*
Translator C: *What happened now was that we are expected to use, preferably, both source texts and this double translation system. But in the verification, only one of them is used. It can't be like this. It would be OK, if they [the two source versions] were much closer to each other. But there are differences between them, the English and French versions. So we noticed that it doesn't work.*

Even more importantly, the differences between the source versions left the translators wondering whether or to what extent the two versions were really equivalent to each other. Furthermore, if the two source versions were not equivalent to each other, how did this affect the translations made from them, especially if some countries made their translations only on the basis of the English versions and others only on the basis of the French versions?

Translator C: *Another thing is this usefulness of using the French version. It has really been a huge improvement. And just because it brings us more freedom. Even though, at the same time, it also leads to good questions. Like, to what extent are these two official source versions equivalent to each other?... And if some countries translate only on the basis of the English version and others on the basis of the French version, what is the outcome? How equivalent are these translations?*

4.4.3 Discussion

The findings obtained in this study echo the contradictory findings gained in previous research, suggesting that when two translations are made from two source versions, this can both help to produce better and more equivalent translations and make it more difficult. For example, it can help translators to translate more accurately and more fluently.

However, the study also pointed to problems when following the procedure. Firstly, when two translations are made from two source versions, the reconciler has more tasks to accomplish than when there is only one source version (and especially when there is only one target version). This was certainly the case in Finland, where the first reconciler took care, not only of his or her own duties, but also of those that in the recommended PISA procedure rest with the translator translating the materials from French into the target language (on the other hand, the first reconciler did not need to do much reconciling). However, it seems that more generally, too, a lot of responsibilities and tasks have been vested in reconcilers, in particular, and that the use of the two source versions increases this load. The reconciler has to compare two target versions to two source versions; s/he has to merge together two target versions that have been translated from two different-language source versions and may therefore have differed considerably; and s/he has to revise and finalize the language of the resulting versions, compare it to the language of the source versions and make sure that the translations are equivalent to the source versions also in this respect. Secondly, a proper use of the two versions would in fact require that reconcilers and verifiers be competent in both languages. However, in several countries finding such reconcilers and verifiers has not been possible. Thirdly, the two source versions have not been fully equivalent to each other.

Steps thus seem to be needed to make the practice of making two translations from two source versions more efficient and reliable. To this end, it is important, for example, to see to it that the reconciler has sufficient time to properly perform all the complex and time-consuming tasks that s/he is expected to perform. This might be done, for example, by hiring several reconcilers, each of whom could only work on a few texts. However, the problem with this solution is that the requirements for reconcilers, in particular, are extremely high. For example, they should preferably have a good knowledge of both source languages (not just one). However, finding even one person who would have met this requirement has often been difficult. Therefore, finding several such persons may not be a realistic goal. Another, slightly better solution, then, might be to reserve more time in the translation schedule to the reconciliation phase. Still another option might be to decrease the duties of the reconciler by making, say, the revising and finalizing of the reconciled versions a phase of its own and the responsibility of another translator. Both these latter measures necessitate that more time be allotted to the translation process in the testing schedule.

In addition, whenever two or more different-language source versions are used, it is necessary to ensure that the versions are equivalent to each other. Of course, full equivalence cannot be expected, because, as widely acknowledged in both translation studies (e.g., Chesterman, 1997; Pym, 1995) and test translation (Grisay, Gonzalez & Monseur, 2009), absolute equivalence does not exist. However, a sufficient level of equivalence needs to be ensured, and measures need to be taken to guarantee that this level is attained. This should involve not only a simultaneous production of the parallel versions by a team of translators and content and testing specialists (cf. Harkness et al., 2010a, p. 47), as largely done today in international achievement studies, but also thorough quantitative and judgmental comparisons between all the source versions and preferably even pretesting – and a considerable amount of time to do all this. Then, on the basis of the comparisons and testing, a decision can be made on how equivalent the parallel versions are and whether this level is sufficient to justify the use of both or all of them. If a sufficient level cannot be ensured, it may be safer to use only one source version.

All in all, however, more research is needed on the use of parallel source versions. For instance, given the contradictory experiences gained when making two translations from two different-language source versions, it is important to know how countries actually use the source versions: After countries have made the two target versions from the two source versions, do some of them use more or less exclusively only one of the target versions, with only a few small extracts taken from the other version; and do other countries use the two target versions more equally? Which of these uses yields better translations? Also, what is the best way of translating international achievement tests: double-translation from two different-language source versions; double-translation from one source version; double-translation from one source version which, however, can vary from country to country; single translation from one source version; single translation from one source version which can vary from country to country; or some other procedure (e.g., the procedure followed in Finland)? Likewise, how many different-language source versions should there be? What are the languages in which the source versions should be prepared? To find answers to these questions, comparisons are needed between all the procedures actually followed when translating international achievement tests. In practice this might be done, for example, by conducting discussions with translators, reconcilers, reviewers and verifiers working in the various achievement studies and following the various translation procedures and comparing their experiences and views. Careful linguistic and psychometric comparisons are also needed between the translations made when following the different procedures.

4.5 Revision and Verification

4.5.1 Previous Research

Translation studies. The amount and quality of revision has a significant impact on the quality of the translation. Properly revising a translation requires checking, correcting and improving not only the semantic accuracy and grammaticality of the translation but also, for example, its idiomaticity, fluency, style and textuality. Doing all this takes a considerable amount of time. (Mossop, 2007.) How a translation is revised, in turn, depends on at least five largely interrelated factors, most of which also have an impact on how texts are translated: the translation goal and translation guidelines, the method of revision, the time spent on revising, the number and qualifications of the person(s) making the revision, and the medium for making the revision (paper or screen). Deficiencies in any of these easily result in the reviser not being able to make a proper revision – and the translation ending up containing errors and unduly literal translations, for instance.

For example, if the goal is vague or difficult to grasp and/or if the guidelines do not say clearly and unequivocally how the reviser is to revise to reach the goal, the reviser is left uncertain as to how to proceed and may therefore be tempted to "play safe" and accept literal translations (cf. Pym, 2008, p. 324; see also Chesterman, 2010, p. 42; Jensen, 2009). Also, properly revising a translation necessitates several separate revision rounds. For example, a bilingual revision is needed to check the translation for faithfulness to the source version. However, to check whether the translation is in idiomatic target language, another, monolingual revision is needed. If both were checked at the same time, the source text and its wordings would make the reviser blind to unduly literal renderings, for example, and thus unable to assess the idiomaticity of the translation. (Cf. Englund Dimitrova, 2005a, pp. 32, 233; Mossop, 2007, p. 147.) This is also the reason why the monolingual revision should preferably be carried out before the bilingual revision, if both are made by the same person (Künzli, 2007; Larson, 1998).

The above also largely explains why back translation is often not an effective revision method. When back translating, the test is first translated into the target language and then back into the source language, and the quality of the target text is then judged on the basis of the comparability of the two source language texts. The revision and the assessment of the quality of the translation thus concentrate on the back translated version and its semantic faithfulness to the source text; much less attention is paid to the translation and its idiomaticity (Larson, 1998). Besides, when the translator knows that what s/he translates will be translated back into the source language and that the quality of the translation will be assessed on the basis of its faithfulness to the source text, s/he may think that the best s/he can do is to stay close to the source version (van de Vijver & Leung, 1997, p. 39).

Also, if the reviser lacks time, s/he will not be able to make several revision rounds and to elaborate on and improve the text. Both easily lead to unduly literal translations. (Englund Dimitrova, 2005a; Mossop, 2007, p. 147.) Similarly, if the reviser does not have a good knowledge of the target language, s/he will not be able to ensure that the translation is in good target language; or if s/he does not have knowledge of the principles and theory of translation and revision, s/he may not know how to revise or s/he may even think, for example, that literal translations are good translations. Also, if the reviser works alone, there are fewer eyes to spot errors and unduly literal renderings; in addition, s/he will not be able to divide the tasks so that, for instance, one of the revisers could concentrate solely on the monolingual revision. (Mossop, 2007.) Finally, if the reviser works on screen, s/he may not be able to spot errors and examine the text as a whole (Englund Dimitrova, 2005a, p. 144; Englund Dimitrova, 2005b, p. 36; Mossop, 2007, p. 101).

After a text has been translated and revised, it is sometimes submitted to external evaluation. During this phase, one or more evaluators assess (and possibly even grade) the quality of the translation and verify its compliance with agreed criteria. However, the translation is no longer corrected or improved. (E.g., Brunette, 2000.) Evaluation is largely guided by the same principles as revision (see above).

International achievement studies. The way revision is carried out in international achievement studies varies somewhat depending on the translation approach. In the back translation approach, the revision consists of comparing the back translated version to the source version: The comparison provides the basis on which the target langue version is corrected and improved. At the same time, however, the comparison also serves as an evaluation of the quality of the target version. In the back translation approach, then, revision and evaluation are merged together and take place at the same time. In the forward translation approach, however, there is one phase which is only dedicated to revision. In, for example, IEA studies this phase is called review, and in OECD studies it is reconciliation. However, the review and reconciliation phases differ from each other, in that the former only involves revision, whereas the latter also comprises reconciling.

After the national review or reconciliation has been carried out, the translations are submitted to international verification. The verification phase differs from normal translation evaluation, in that it also involves revising, correcting and improving the translations. The verification is carried out against a checklist, which outlines the purpose of the verification and specifies how verifiers are to proceed and what kinds of problems, for example, they are to look for so as to ensure as high a level of equivalence as possible. The verification is performed by a verifier, who thus acts as a research instrument (Patton, 2002, p. 566). Extremely high requirements are therefore set on the verifier (much higher than those on the other translators). For example, s/he needs an excellent knowledge of all the languages and subject matters concerned, of literary translation, of the principles and theory of translation, and of cognitive tests and test translation. Verifiers are also provided with training and hands-on exercises on how to verify (Dept, Ferrari & Wäyrynen, 2010).

No research proper exists on the review, reconciliation or verification and their effects when translating international achievement tests. However, in questionnaire translation, some studies on review do exist (see e.g., Harkness, Villar & Edwards, 2010, pp. 131–2). When translating questionnaires, the recommended procedure has been slightly different from what it has been when translating international achievement tests: Translation, Review, Adjudication, Pretesting, and Documentation, or TRAPD (e.g., ibid., p. 128). To evaluate the efficiency of this method, Willis et al. (2010) examined documents reporting on the processes followed when translating five questionnaires. They wanted to know to what extent the three revision and evaluation phases included in the method – review (expert review to identify problems and additional translation options), adjudication (making decisions to reconcile options from the first translation drafts and the reviewed versions) and pretesting (cognitive interviewing) – were able to spot problems in the translated questionnaires. They found that great numbers of linguistic translation problems were discovered during both the review and adjudication phases. Both phases were thus needed and played a significant role in revising, correcting and improving the questionnaires.

In international achievement studies, too, the amount of and effort put on revising seems to have had an effect on the quality of the translations (e.g., Solano-Flores, Backhoff & Contreras-Niño, 2009). For example, when Solano-Flores, Contreras-Niño and Backhoff-Escudero (2006) examined the quality of the Mexican Spanish-language version of the TIMSS 1995 test (in TIMSS 1995, the translation procedure was double-translation from English followed by reconciliation), they found that it contained a significant number of translation errors. They concluded that these were largely due to the version not having been properly revised nationally: There were no clear criteria against which to revise, the revisions had to be made in a hurry, the revisers did not make sufficient revision rounds, and there were not enough discussions between the translators and revisers.

Also, there are findings suggesting that the lack of revision – and the resulting lower quality of the translations – may have had to do with the use of the two source versions and the double-translation and reconciliation procedure. For example, the Finnish reconciler in IALS mentioned that when two Finnish versions were made in Finland, one of them from English and the other from French, the versions were often so different that merging them into one Finnish version was extremely difficult. After the merging, it should have been possible to concentrate more on revising and finalizing the versions and making them coherent, fluent and idiomatic. As another factor complicating the reconciliation, the reconciler mentioned that s/he did not know French. (P. Linnakylä, personal communication, November 14, 2008.)

After the IALS study, Finland thus decided to stop following the recommended procedure and modify it so that more attention could be paid to the revising and finalizing of the translations: In most PISA studies (all except PISA 2006, where the procedure was as recommended), Finland has only made one translation (from English) which has then been reworked and revised by two successive national translators (plus the international verifier). Interestingly, however, the quality of these translations has been judged to be very high, one of the best: In psychometric comparisons, they have been found to contain very few flawed items (Grisay, 2002; OECD, 2005); and according to verification reports, they have often not needed any proofreading, because they have been linguistically and grammatically of so high quality (see section 4.6). Of course, more research is needed to disentangle the reasons for the seemingly high quality of the Finnish PISA translations. However, it appears that it is at least partly due to the fact that in Finland much more weight has been put on the revising and finalizing of the national versions than in the recommended PISA procedure.

This seems to be supported by the fact that when the quality of Finland's Swedish materials was verified in PISA 2000, 2003 and 2006[5] (by the same verifier who verified Sweden's materials), it was judged to be much higher than that of Sweden's corresponding materials: Finland's Swedish materials contained fewer errors and more fluent and natural language than Sweden's materials, among other things. Yet, the materials were basically the same, because Finland's materials had been borrowed from Sweden and only adapted for use in Finland. In Sweden, the translation procedure had been as recommended (double translation from English and French followed by reconciliation). In practice, the only difference was thus that in Finland the materials underwent an extra revision round, during which they were checked by an extra person who was able to concentrate solely on the final Swedish versions – without interference from the English and French source versions or the first Swedish drafts – and on finalizing them.

However, there have also been contrary findings. When translating the PISA 2003 materials, Italy followed the recommended procedure, basing its translations on both the English and French source versions. However, when verified, only some minor problems were found in the translations (Siniscalco, 2006, p. 206). The translations also contained very few flawed items (OECD, 2005, p. 78). Siniscalco feels that the good quality of the Italian translations was largely due the great amount of time and effort that was invested in revising and finalizing them.

[5] In PISA 2000, 2003 and 2006, international verification was carried out, when the minority population made 5% of the target population. In PISA 2009, the limit was raised to 10%. In Finland, the proportion of those with Swedish as their mother tongue has been over 5% but clearly under 10%. Therefore, from PISA 2009 onward, Finland's Swedish materials have not been verified.

In addition to there having been problems in revision, there seem to have been problems also in verification. This appears to have been the case during TIMSS 1995, in particular, when the procedure was new and not yet fully developed and when verifiers did not yet receive any verifier training (Dept, Ferrari & Wäyrynen, 2010). Thus, when, for example, Solano-Flores, Contreras-Niño and Backhoff-Escudero (2006) found several severe errors in the Mexican TIMSS 1995 instrument, they felt that these were due, not only to inadequate national revision, but also to deficiencies in international verification, such as a lack of criteria against which to verify, a lack of sufficient revision rounds, and a lack of time.

However, it seems that the TIMSS study is not the only study where there have been problems in verification. For instance, Arffman (2007) attributed some of the problems and non-equivalences in the Finnish PISA 2000 instrument to the verification not having been sufficiently thorough and profound. In PISA, it appears that the problems have sometimes been related to the use of two different-language source versions: Since the two versions cannot be fully equivalent to each other, the verification should be performed against both versions (and verifiers should, consequently, be competent in both languages). However, this does not always seem to have been the case. For example, Dept, Ferrari and Wäyrynen (2010, p. 166) refer to cases where translations have been "corrected" so as to bring them closer, for example, to the English source version, even though they have already been faithful to the French version. This, in turn, may often be expected to have been because countries have been unable able to hire verifiers who would have been proficient in both source languages (B. Halleux-Monseur, personal communication, January 24, 2008).

4.5.2 Findings of This Study

Revision. The translators in this study mentioned revision as one of the factors that has been problematic when translating international achievement tests. One of the reasons for the problems seems to have been that there have not always been enough revisers and/or revision rounds. For example, Translator C mentioned that in PISA 2000, when s/he was the only reconciler and reviser in Finland, a considerable part of his or her time and energy went into comparing the Finnish translations to the source versions, which, in turn, often made him or her unable to properly check and finalize the language of the translations. This time (in PISA 2009), however, there were two reconcilers. And since the first reconciler carefully compared the materials to the source versions, the second reconciler could concentrate more or less entirely on the target versions. This, in turn, helped to improve the idiomaticity of the translations.

I: *There may have been some small issues having to do with grammatical correctness there [in the final Finnish PISA 2000 translations]….*
Translator C: *In a way it's quite amusing that there were still those grammatical correctness issues in them, because both I – I don't remember any more whether [another proofreader] was involved then already, maybe not – had a look at them from that point of view. But the problem here is exactly that when I myself had a look at the English version I went blind…. So, in a way it now [when translating for PISA 2009] worked, because there was one person [the first reconciler] who had a look at the English version and then, in a way, I didn't need to look so much at it that I would have gone blind to the Finnish version. So this time this interference from another language was missing.*

At the same time, however, the workload and time pressure of the first reconciler became heavier (as described in section 4.4.2). This, in turn, made it impossible for her to revise and finalize the translations as carefully as she would have liked to revise and finalize them. Instead, she had to make quick decisions and "compromise", for example, the idiomaticity of the translations.

I: *What often annoyed me was that you can't make these texts as good as you would like to make them…. If I was translating in some other context… We have these limits.*
Translator A: *That's true…. Even though we have this ideal goal [to make good translations], so to speak …. But then there is this other side of the thing … these schedules. The work must go on. You can't think on and on and improve endlessly…. You have to make compromises.*
I: *Well, you really have to make a lot of them.*

Translator C: *Then there is this lack of time….*
I: *You can't think it over as you would like to. And this is another thing that, well, irritates me all the time…. you don't have much time for anything.… Well, I don't know how much anyone would like to spend time solving a small problem. But anyway, you have to make the decision quite quickly. And then it is… I think it has been shown also in translation studies that the less the translator has time, the more literally, word-for-word s/he translates.*

Verification. The translators also mentioned problems in verification. For example, since the verifier works as a part of the quality control system and evaluates the equivalence of the translations, national translators expect him or her to be the highest expert, the authority who helps them to solve translation problems and on whose guidance and decisions they can rely. However, it seems that this has not always been the case. For example, Translator C mentioned that in PISA 2000, the Finnish verifier was not a linguist or translator or not even a student of languages, but a student of psychology.

Translator C: *We just have to trust that, well, compared to us [national translators], the verifier is, so to speak, the higher expert. But it seems that this is not the case…. Namely, I thought back then already … when we had this verifier who lived near [the institute]. It must have been during the PISA 2000 round…. But she was a student of psychology. So of course, she had had that verifier training, and her English was fluent, and so on. But I don't know if she was actually any better than me. So, I thought*

> *back then already that, well… But how I viewed the verification back then and still do is that it is, well, a good external pair of eyes…*
> I: *Yes, exactly. We should not expect too much of it.*

As another problem with the verification, the translators mentioned that it does not always seem to have gone very deep. Mainly it appears to have concentrated on comparing individual words and expressions; at the same time, larger entities and, for example, stylistic considerations have often been ignored. Also, except for the points directly addressed in the question items, relatively little attention seems to have been paid to the stimulus texts, even though especially in reading tests, the stimulus text *per se* has a great impact on how the reader understands, reacts and responds to the text and items. The verification may therefore have failed to spot and correct, for example, cohesion and coherence (i.e., text-level) and stylistic problems.

> Translator C: *In the end, the verification is rather superficial. It can't find the real problems.*
> I: *That's right. And what it mainly focuses on ... is this relationship between the stimulus text and the question items. The texts in themselves receive very little attention. And I think this is surprising because really the text itself also has a very great impact.*
> Translator C: *That's right. It's so mechanical.... And this reflects the same view that could already be seen in the Finnish [reading literacy] text … that there has to be [in every national version, in imitation of the source versions] an expression that contains [three particular nouns]. So, this type of cognitively based word-level examination…. It [verification] could be much more beneficial.*

Still a third problem with the verification phase brought up by the translators was that it has been carried out against only one source version, even though the two versions have not been fully equivalent to each other and even though the Finnish translations have been made on the basis of the two versions (see section 4.4). This, in turn, has led to unnecessary and confusing "corrections" and extra work for the first reconciler.

4.5.3 Discussion

Revision. The findings of this study agree with previous findings, suggesting that there have been problems in revising international achievement tests and that these may have lowered the quality of translations used in these tests. For example, it seems that countries may not always have had enough revisers and that the revisers have often had so many other tasks and so tight a timeline that they have not been able to concentrate on properly revising the translations and on making several revision rounds, for example. This, in turn, appears to be at least partly due to the double-translation and reconciliation procedure, where countries do not always seem to have been able to pay sufficient attention to revision. Since there has been no phase in the procedure during which a person could have concentrated solely on revising but, rather, revision has had to be carried out as the last step of the reconciliation phase, a large portion of the time, energy and attention of the reconciler may easily have gone into comparing the two target versions to the two source versions and merging the two target versions into one.

It thus seems important to ensure that especially in studies following the double-translation and reconciliation procedure, sufficient attention be paid to revision (see also Hambleton, 2002, p. 67; Hambleton & Patsula, 1999; Siniscalco, 2006; Solano-Flores, Contreras-Niño & Backhoff-Escudero, 2006). In practice this might mean splitting the reconciliation phase into two so that during the first phase one person, reconciler, could reconcile the two target versions into one and during the second phase another person, reviser, could revise and finalize the resulting translations. Revision would thus be a phase of its own. In the current double-translation and reconciliation procedure, the reconciled versions are, of course, verified by the verifier. However, knowing as we do the huge number of factors that have to be taken into account when verifying and how difficult, if not impossible, it is for one reviser to spot all unidiomaticicities and errors, especially when working on screen (Mossop, 2007), it would be beneficial to have at least one person properly revise and finalize the translations before they are verified. Obviously, splitting the reconciliation phase into two and making revision a phase of its own means that more time needs to be allocated to translation.

Other ways to guarantee proper revision might include the following: Making sure – by means of a test, for instance – that the revisers have all the needed qualifications. Partly revising the translation guidelines (see section 4.3.3); or, preferably, making separate guidelines and a checklist for revisers, telling them what to revise and how (e.g., that they need to make several revision rounds; see e.g., Hambleton & Zenisky, 2011; Solano-Flores, Backhoff & Contreras-Niño, 2009, p. 82; cf. Colina, 2009; Jeanrie & Betrand, 1999, pp. 280–1; Martínez Melis & Hurtado Albir, 2001). Providing revisers with hands-on training on how to revise. Reserving so much time to revision and using so many parallel revisers that the revisers have sufficient time to make several separate revision rounds, to be creative, and to discuss with other translators and subject matter experts, for example. Encouraging revisers to revise at least partly on paper. This has become increasingly important today, when more and more of the translation work is done in electronic environments. Conducting cognitive laboratories as part of the revision process: Having pilot testees and other outsiders read the translations, complete the test and comment on the language (cf. Ercikan et al., 2010).

Verification. The findings of this study also agree with previous findings in suggesting that there have been problems in verification. For example, the study showed that verifiers may not always have been fully qualified. This appears to have been the case during the first studies, in particular. The verification may also have been carried out superficially. In addition, it has only been performed against one source version. This, in turn, has resulted in unnecessarily "correcting" translations that have already been faithful, for example, to the French version and thus inadvertently nullifying one of the alleged advantages of using the two source versions – the opportunity to see alternative ways of expressing ideas and to translate more freely.

Improvements thus seem to be needed in the verification. For example, since verifiers serve as evaluators and research instruments, it is absolutely necessary that they are highly qualified: They need to be professional translators, well acquainted with translation theory and the principles of translation; they need not only a perfect knowledge of one of the source languages and the target language, but also a sufficient knowledge of the other source language; they need a good knowledge of the topics covered in the test; in reading literacy tests, they also need familiarity with literary translation. A test may be used to check to what extent the prospective verifiers fulfill these requirements. In addition, however, the verifiers also need training in cognitive tests and test translation (as has been the practice since PISA 2000; Dept, Ferrari & Wäyrynen, 2010).

Also, since the purpose of verification is to evaluate and monitor the quality of translated instruments, it is necessary that it be carried out in sufficient depth. This, of course, requires (among other things) that verifiers have sufficient time to do it: that they have time, for example, to make several revision rounds and to pay more attention to the stimulus texts and style (especially in reading literacy tests). It may also require that at least part of the verification be carried out on paper (where spotting, for instance, coherence and cohesions problems is easier than on screen). And it may require at least partly revising and updating the verifier training and the verification checklist currently provided in international achievement studies (cf. Colina, 2009; Martínez Melis & Hurtado Albir, 2001). For example, it may be necessary to put more emphasis on the need to make several revision rounds and the need to pay more attention to the stimulus texts, coherence and style.

Finally, the study suggests that if translations are made on the basis of two source versions, the verification should also be preformed against the two versions. However, such double verification is probably not needed for all materials but mainly only when the verifier feels that the target version may deviate too much from the source version s/he is mainly verifying against (e.g., the English version). Carrying out the verification against two languages, of course, requires that the verifier be competent in both languages and that s/he have sufficient time to compare the translations to both source versions.

All in all, however, much more research is needed on how international achievement tests are revised and verified: How much effort do countries put in revising and finalizing their test versions? What methods do they use when revising (e.g., how many revision rounds?)? What methods do verifiers use when verifying? How does reconciliation and/or the practice of making two translations from two different-language source versions affect the revising, finalizing and verifying of test versions? How do the methods of revising and verifying affect the quality of the versions?

4.6 Time

4.6.1 Previous Research

Translation studies. The amount of time the translator can devote to the translation task also has an impact on the difficulty of the task and the quality of the resulting translation (Jensen, 2000). Generally speaking, high-quality translations necessitate considerable time and processing (Jääskeläinen, 1999, p. 123; Kim, 2006, p. 295). This is because while translating, the translator is repeatedly faced with translation problems (e.g., which of the several meanings of this word is the one meant here; what does this metaphor mean and how can I convey its meaning in the target text; how can I maintain the artistic flavour of this short story; what do these technical terms mean, and are there comparable terms in the target language; how can I say this in a more fluent, natural and idiomatic way?). The translator needs time to solve these problems. S/he needs time to invent and be creative (see e.g., Fontanet, 2005, p. 444), to do research and discuss with others (Mackenzie, 1998), and to elaborate on the text (Jääskeläinen, 1996).

If the translator is in a hurry or if s/he translates hastily, s/he will not have sufficient time to be creative, to do research and discuss with others, and to improve the text. Instead, s/he has to be satisfied with the solutions that first and most effortlessly come to his or her mind (the so-called minimax strategy, "the translator resolves for that one of the possible solutions which promise a máximum of effect with a minimum of effort"; Levý, 1967, p. 1179). These, however, are often the most literal translations, those that follow closely the formulations of the source text (see also Chesterman, 1997; Englung Dimitrova, 2005a, p. 146; Jensen, 2000; Krings, 1986, p. 507). In addition, when in a hurry, the translator lacks cognitive resources, which, in turn, easily leads to errors (cf. Zakay, 1993).

The amount of time that is needed to solve translation problems, however, is not a constant, but varies, for example, between translation and revision. Usually, more time is needed to draft than to revise a translation. However, if the quality of the translation draft is very poor, revision can be extremely time-consuming. (Mossop, 2007.)

The time needed to solve translation problems is also dependent on the translator and his or her competences. The more the translator has experience of translation and the more familiar s/he is with the translation task (e.g., the more experience s/he has of the type text that s/he has to translate and the more knowledge s/he has of the subject matter), the more automatized his or her skills are and the better equipped s/he will be to solve translation problems. In contrast, inexperienced and unqualified translators or translators who are not familiar with the translation task have less automatized skills and therefore need more time to solve problems. (Englund Dimitrova, 2005a; Jensen, 2000; Jääskeläinen, 1999; Kim, 2006.) Yet, research shows they often translate very quickly, because they are not aware of translation problems (Jääskeläinen, 1999).

The text also has an impact on how much time needs to be spent translating. For example, long texts understandably take longer to translate than short texts. Also, more time is usually needed to translate, say, literary and special-language texts than more everyday text types. This is because they typically contain more translation problems than other, more straightforward and popular texts, literary texts because of, for example, their semantic vagueness and the creative processing they require, and special-language texts because of their unfamiliar content and technical terms (see section 4.1). (e.g., Korning Zethsen, 1999, p. 73; Séguinot, 1988, p. 110; Wilss, 1998, p. 58).

International achievement studies. No research proper exists on whether or how time pressure and haste have affected translators translating international achievement tests. Not surprisingly, however, experience and reports show that they have caused problems also in these studies (e.g., Hambleton, 2002, 2005; Hambleton & Berberoglu, 1997; Solano-Flores, Contreras-Niño, & Backhoff-Escudero, 2006). Moreover, their effects seem to have been the same as always when translating in a hurry. For example, in her analyses of Finnish PISA 2000 field trial and main study materials, Arffman (2002, 2007) found numerous mistranslations and unduly literal renderings and interpreted these as being due to the translators having translated in a hurry.

4.6.2 Findings of This Study

The translators in this study had slightly different views as to whether time pressure and haste have caused problems when translating international achievement tests. Four of the five translators felt that they did have enough time to translate, revise or verify the Finnish PISA 2009 materials. Only one, the first reconciler (I), felt that this was not the case.

The explanation the translators themselves gave for the different views was that the amount of work they had differed considerably: All the four translators who felt they had enough time had someone else with whom they could share their task (e.g., two translators shared in making the first Finnish drafts) or on whose work they could draw (the second reconciler could draw on the work of the first reconciler, and the verifier could draw on the work of the national translators), and this, of course, reduced their workload and the time they needed to spend on the tasks. However, the first reconciler had no one.

> Translator C: *In a way I did have enough time.... But this was largely because after you [the first reconciler] had had a look at the texts, they were so ready. So, that I had enough time was largely thanks to and at the expense of you. You had such a horrible amount of work at that time.... But it's clear that it made my work easier....*

> Translator A: *It wouldn't have succeeded, if I had had to do all the texts. In that sense [the other translator]'s contribution was invaluable. So this time it didn't cause me much stress.*

> Translator D: *The [Finnish] translations are of so high-quality that the verifier can concentrate on the verification and does not need to proofread the text first.... For Finland, the process has always worked in an excellent way – the materials are very well translated, the instructions have been followed (and they have been understood), the schedules have been stuck to, etc.... I had enough time.*

How much time, then, did the translators spend working on the materials? This varied enormously. For example, when asked about the minimum time spent on a unit, the answers ranged from 10 minutes to three hours. The least time was spent by the second reviser and the verifier. The first reconciler, however, mentioned having spent "almost two hours" per unit. For the maximum time spent on a unit, the answers varied even more: from one hour to a day and a half. Again, the least time was spent by the second reviser and the verifier, whereas the first reconciler was one of the two translators who reported having spent the most time per unit. Finally, for the average time spent on a unit, the answers ranged from 20 minutes to one day, the second reviser and the verifier, again, being those who spent the least time and the first reconciler one of those who spent the most time. All in all, when taking into account the number of units translated or revised and the time spent per unit by each translator, by far the most time was spent by the first reconciler.

The above figures suggest that the amount of work probably did differ greatly between the translators. They also suggest that, generally speaking, the workload of the translators was heavier than that of those revising them but that the workload of the first reconciler was probably the heaviest. However, in addition to the workload, there may also have been other factors which may have affected the time the translators spent on the materials. For example, the extremely short time (20 minutes on average) one of the translators spent revising one unit might suggest that the reviser may perhaps have had a somewhat simplistic view of revision and may therefore have revised too hastily. In a very short time, it is not possible to go deep into a text and make several revision rounds (e.g., one for comparing the translations to the source versions, and another for checking its fluency and naturalness; see section 4.5). Similarly, the huge amount of time (one day on average) spent by the first reconciler working on the translations might be an indication that her skills were not fully automatized and/or that the quality of the translation drafts she was working on was not always as desired.

The lack of time and time pressure of the first reconciler, in turn, meant that there was not enough time for discussions between the translators and subject matter specialists, which, however, could have helped to solve translation problems. This was a problem in PISA 2006, in particular, where science was the main domain and where subject matter knowledge was therefore especially important.

> Translator A: *There is one such thing I have been thinking about … that we would do [translate] for a certain period, say for a week or two. And then, at some, point, there would be a time for feedback: a discussion, where we would see, well, where there have been problems…. But it wouldn't have worked in this situation either. We wouldn't have had time for that.*
> I: *We wouldn't have had time. And this is exactly what I'm saying, that we would need more time. There would have to be time for such discussions…. It would benefit all of us.*
> Translator A: *It would. Because then you would hear … there would be an exchange of views…. But well, yes, teamwork, it really calls for resources.*
> I: *It really does. And this is the only reason why we didn't do it [have discussions] – actually that I didn't have enough time.*

> I: *[When translating the PISA 2006 materials into Finnish, a specialist in science involved in the translation work] suggested something like this that when they [the first Finnish drafts] come from [the translator], so [they] first [come] to me. And then I send them to [the two science specialists]. And then we all four discuss so that even [the translator who has made the first drafts] is connected by phone.*
> Translator C: *I see. Pretty laborious.*
> I: *Yes. And we had one such discussion – three hours! … And then I said … yes, I said that I haven't got time to do this…. And then it was changed so that … I only sent them [the translations] … to [the two science specialists]. And I don't know what happened to them after that….*
> Translator C: *It is true that there is not time for something like that. Because it always takes much more time when you discuss.*
> I: *Yes, it was such a terribly long time. I mean, three hours! … And there were six texts that we discussed. Six texts!*

In addition, the translators were concerned that because of the time pressure and haste, they would make careless mistakes. Also, they would not have time to think, elaborate on the texts and try to find creative solutions. Instead, they would have to accept the most obvious and "automatic" solutions – those that follow closely the wording of the source version.

> Translator C: *All in all, this haste is irritating. Especially because then there is this stress and you can't sit down and work as you would like to. So it can have an impact on quality.... There will be careless mistakes….*
> I: *And you haven't got time to think and try to find those innovative things, good solutions. You necessarily stay close to the source text.*

Translator B: *Very often the solution will be found after some time by itself. Creative solutions will not be found by force.*

Interestingly enough, even though most of the translators felt that they did have sufficient time to translate the PISA 2009 materials, three of the four translators who had previous experience in translating international achievement tests commented that lack of time and time pressure is always a problem when translating these tests. The reason for this, in turn, is that the time allotted to translation in the testing schedule is insufficient.

Translator C: *I feel this [translation] process has always been characterized by haste, which is disgusting. It is so stressful, but you know that you just have to keep to the schedule.... Time is always a problem.*

4.6.3 Discussion

The findings of this study agree with those of previous studies, showing that time pressure and haste can and often do complicate translation work in international achievement studies and may lower the quality of translations used in these studies. However, it appears that time pressure and haste may not affect all translators and translation phases to a similar degree. For example, in this study, only the first reconciler reported suffering from time pressure. This might appear surprising, because typically revising is known to take less time than translating. However, it is not surprising when we remember the great number of responsibilities the first reconciler had (apart from revising) and the fact that she had to check all the translated materials. Even though the workload and time pressure of the Finnish first reconciler may have been somewhat heavier than those of reconcilers in general, it seems that time pressure may be a problem for reconcilers, in particular, because they have several time-consuming responsibilities (especially in PISA, where translations are made from two different-language source versions) and because they often cannot share their tasks with anyone.

The above further suggests that time pressure may be a problem especially in those studies where reconciliation is needed (as opposed to studies where reconciliation is not needed, because only one target version is produced) and, especially, when the reconciliation involves merging together translations that are made from two different-language source versions (as opposed to studies where the translations are made from only one source version). Also, it may be a more serious problem when translating, for example, science and reading literacy tests than when translating mathematics tests. This is because the former require more processing: science texts because of their specific, technical terminology and literary texts because of their multiple meanings and artistic emphases.

Of course, more research is needed on how much time translators, reconcilers, reviewers and verifiers in international achievement studies really use in the translation work (e.g., per unit) and how this affects their translating and the quality of the translations. Research is also needed on to what extent the time spent in the translation work varies, for example, between the translator roles (translators, reconcilers, reviewers, and verifiers) and according to the domain assessed.

However, given the serious negative effects time pressure and haste have on the quality of translated tests, it is important to ensure that translators, and especially reconcilers, in international achievement studies have sufficient time to perform their tasks properly (and that they also do so). This starts from ensuring that all translators, reconcilers, reviewers and verifiers are qualified. This is important for several reasons. First, unqualified translators and revisers may not be able to produce high-quality translations or make a proper revision within the given time limit. Second unqualified translators and revisers may translate and revise hastily and carelessly. Third, if, for example, the translators are unqualified, this will probably lead to low-quality translations, which, in turn, will increase the time pressure of the reconciler.

Another obvious way to ensure that translators have sufficient time to do their job is to have enough translators. This option, however, mainly works for translators making the first drafts only. The number of the translators, moreover, may have to be higher, for example, in science and reading literacy tests than in mathematics tests. In the case of reconcilers and verifiers, nevertheless, finding several qualified persons is often not possible. Therefore, other, more radical measures may be needed. These might include allotting more time to the reconciliation and verification phases or reducing the workload of the reconciler by making revision a phase of its own. These measures, moreover, may be expected to be needed especially when translating from two source versions and/or when translating tests in more problematic and less translatable subjects.

CHAPTER 5

Conclusion

5.1 Summary of the Study

This study explored the problems translators feel there have been when translating international achievement tests. The purpose was to help to develop translation procedures and practices in these studies and to increase the equivalence of the different-language versions of these tests and the validity of the studies. To explore the problems, an action research study was carried out: Semi-structured face-to-face and email discussions were conducted with the five translators participating in translating the PISA 2009 reading test into Finnish. The discussions were analyzed by means of content analysis.

The discussions showed that there have been problems when translating international achievement tests. Not surprisingly, the problems appear to have been greatest during the first studies, whereas in more recent studies progress has been made. Despite this progress, however, there still seem to be problems in the procedures and practices which complicate the translation work and threaten the equivalence of the different-language test versions. The translators pointed to the following problems:

- *Source instrument*: Lack of translatability, largely because of differences between languages. Comprehension problems, especially in scientific texts. Question items especially problematic. Simplification of literary texts.
- *Translators, reconcilers, and verifiers*: Lack of knowledge of the source language or source languages (in the case of reconcilers and verifiers), lack of subject matter knowledge,

lack of knowledge of the principles and theory of translation, lack of experience in literary translation, lack of familiarity with cognitive tests and test translation.

- *Translation goal:* Vague and elusive; impossible to know how difficult the source and target texts and items are.
- *Translation guidelines:* Indecision as to how many specific linguistic instructions to include in the guidelines: In some studies too many, leading to undue focus on faithfulness to the source text; a problem especially when translating into non-Indo-European languages. In some studies too few, leading to a lack of practicality. Level of freedom allowed left unclear.
- *Translation notes:* Used as a remedy to problems in the source instrument (e.g., differences between languages). Used inconsistently.
- *Parallel source versions:* Differences between the parallel source versions.
- *Revision:* Lack of sufficient revision rounds. The double-translation (from two source versions) and reconciliation procedure: heavy workload and extra demands (language skills) for the reconciler; no phase that would concentrate on revising and finalizing the translations.
- *Verification:* Superficiality and a lack of sufficient revision rounds. Use of two source versions: increased workload (verifying against both source versions) and extra demands (language skills) for the verifier.
- *Time:* Not sufficient time for solving translation problems, being creative, discussing, and making several revision and verification rounds. A problem especially for reconcilers and in tests where a lot of problem-solving and/or elaboration is required when translating the materials.

Deficiencies in any one of the above factors may per se lead to translation problems and non-equivalent test versions. However, what often makes the situation even worse is that frequently there have been deficiencies not only in one but several of the factors at the same time. For example, when translating international achievement tests, the goal is by necessity always vague and elusive. This in itself complicates translation. However, in addition to this, there are often also other problems. For example, the source instrument may contain a lot of untranslatable features and may be written in a language that differs greatly from the language into which it is translated; the translation guidelines may focus on faithfulness to the source version; the translators may lack language skills, subject matter knowledge, experience in literary translation, knowledge of translation theory, and/or familiarity with test translation; the translations may not be revised properly; and/or the translators may translate hastily. Each of these problems increases further the risk of translation problems and low-quality translations.

5.2 Developing Translation Procedures and Practices in International Achievement Studies

Some of the above problems, especially those having to do with differences between languages and untranslatability, are such that nothing can be done to do away with them. However, in most cases, measures can be taken to improve the procedures and practices. These may be summarized as follows:

- *Source instrument:* Making the source instrument, and especially the question items, as translatable as possible: avoiding certain linguistic phenomena and certain types of tasks. Using multicultural teams; making ex ante translations; conducting cognitive laboratories. Using compensation to make up for untranslatable features. Ensuring that construct validity is not jeopardized.
- *Translators, reviewers, reconcilers, and verifiers:* Using, for example, a test to ensure that the translators, reviewers, reconcilers and verifiers have a good knowledge of the source and target language, the subject matter, and the principles and theory of translation; that translators, reviewers, reconcilers and verifiers working on literary texts have experience in literary translation; and that reconcilers and verifiers have knowledge of both source languages. Using different translators to translate different topics and different types of text. Hiring more translators when translating tests in more problematic and less translatable subjects. Partly revising the translator requirements.
- *Translation goal and translation guidelines:* Providing in the translation guidelines a right, clear and unequivocal picture of what is meant by equivalence in difficulty. Providing a limited number of linguistic translation instructions. Emphasizing the need for natural target language; warning against unduly literal translation. Warning against undue improvement and explicitation. Preparing customized instructions for translation into non-Indo-European languages.
- *Translation notes:* Remembering that translation notes cannot solve problems in the source instrument (e.g., they cannot remove differences between languages). Developing clear principles as to when to use the notes and following them consistently. Paying more attention to the tone of the wording of the notes (e.g., making them more cautious). When writing translation notes, knowledge of principles of translation and factors causing translation problems needed.
- *Translator, reviewer, reconciler, and verifier training:* Providing all translators, reviewers, reconcilers and verifiers with training in cognitive tests and test translation. Partly revising the training on the basis of the revised translation guidelines.
- *Parallel source versions:* Ensuring the equivalence of the parallel source versions; using parallel source versions only if they are equivalent to each other.

- *Revision:* Ensuring that there is in the translation procedure a phase dedicated to the revising and finalizing of the translations. Making guidelines and a checklist for revisers. Advising revisers to make several revision rounds. Encouraging revision on paper. Conducting cognitive laboratories.
- *Verification:* Ensuring that the verification is sufficiently profound: Advising verifiers to make several revision rounds; encouraging verification on paper. Partly revising the verifier training and the checklist provided for them. Verifying problem cases against both source versions.
- *Time:* Allocating sufficient time in the testing schedule to the translation, reconciliation, revision and verification of the test materials; to team discussions; and to the development and ensuring of the equivalence of the two/parallel source versions. More time is needed when translating tests in more problematic and less translatable subjects; and when the translation procedure involves reconciling.

5.3 Limitations of the Study and Suggestions for Further Research

The study had a number of limitations (most of which are well known in all action research). First, for both practical and ethical reasons, it was not possible to have all discussions face-to-face; instead, some had to be conducted via email. However, the two discussions that were conducted face-to-face involved the two most experienced translators (one of them making the first drafts and the other working as a reconciler). The discussions were thus truly information-rich and in themselves already provided a fairly extensive picture of the topic discussed. In addition, the outline of the discussion (see Appendix 2) ensured that ample opportunities were given also to the other two translators (commenting via email) to voice their opinions on the topic. Their comments also showed that they largely pointed to the same problems as the translators discussing face-to-face – with the obvious exception, of course, that their comments were much shorter, less detailed, and less thick. All in all, then, face-to-face discussions did prove to be a good technique for collecting thick data from the translators on translation problems in international achievement studies. In future comparable studies, if face-to-face discussions cannot be conducted, it may perhaps be possible to have the discussions by phone or by Skype. Another possibility might be think-aloud protocols (asking translators to think aloud about the translation problems and recording their verbalizations).

Second, there is the risk that my role in the face-to-face discussions not only as the researcher and interviewer but also as one of the translators taking part in the translation process and voicing her opinions might have affected the translators, thereby increasing researcher bias and subjectivity. However, the two translators with whom I discussed face-to-face both had previous experience in test translation and had been involved in international achievement studies longer than me. Both therefore had strong, long-established and well-founded conceptions of translation problems in these studies (as, I feel, can also be seen from their comments). In addition, since both got the outline of the discussion well in advance, they had ample time to think about the questions and their answers to them on their own (unaffected e.g., by me). Besides, in the discussion, the two translators were always the first who expressed their opinions.

In studies where the researcher also acts as one of the participants, collecting thick data from all participants in an objective and consistent way is always a problem (cf. Herr & Anderson, 2005, p. 60; Ladkin, 2005, p. 123). One way to overcome the problem might be to have another person conduct the discussions with all the translators. The problem with this option, however, is that to be able to conduct the discussions, the person should be thoroughly familiar international achievement studies, test development and principles of translation (cf. Kvale, 1996). Finding such a person is not easy. Also, extra investments would be required to hire and train the person. Another option might be to use the think-aloud method. However, this method also has its problems: Thinking aloud (in a laboratory setting) is unnatural, and it easily leads to incomplete and faulty reporting; in addition, people differ in how and to what extent they verbalize their thoughts (see e.g., Ericsson & Simon, 1980, p. 243; Genest & Turk, 1981; Nisbett & Wilson, 1977). The third alternative, then, might be to have face-to-face (or phone or Skype) discussions with all the other translators except the researcher and a think-aloud session with the researcher. This method, however, would no longer be consistent.

Third, the results of this study were only based on experiences and views of translators. Therefore, they cannot say whether or to what extent the problems were actually reflected in the quality of the translations or the performance of the testees. To help to find this out, studies are needed where the views of the translators are combined with analyses of the translations and student responses.

Fourth, only five translators took part in this study who, moreover, were all Finns, translated into Finnish and followed a translation procedure which was different from the procedure recommended in PISA and implemented in several other countries. In addition, the translators only translated reading literacy materials. The results of this study are thus not generalizable. Nor are they as such applicable when translating into languages other than Finnish, when following translation procedures different from the procedure followed in Finland, and/or when translating tests in subjects other than reading literacy. Rather, similar discussions or interviews need to be conducted with greater numbers of translators, translating into different languages, following different translation procedures, and translating tests in different subjects.

Fifth, the materials discussed in this study were all printed texts translated in non-electronic environments. However, in the coming years, international tests will increasingly be both translated and administered on-line. Even though a large number of the problems faced when translating electronic tests in electronic environments may be expected to be similar to those encountered when translating printed tests in more traditional environments, there may also be assumed to be problems that are unique to on-line translation. Future studies will need to address these.

All in all, the study showed that research into translation procedures and practices followed in international achievement studies has been extremely limited and that much more research is needed. Among the topics on which research is most badly needed are, for example, the following: What is the ideal number and ideal way of presenting linguistic translation instructions? How does the practice of making two translations from two source versions affect the revising, finalizing and quality of the translations? To what extent are the parallel source versions equivalent to each other? In what languages should the parallel source versions be provided? What is the best and most effective way of translating international achievement tests: double-translation from two different-language source versions followed by reconciliation, double-translation from one source version followed by reconciliation, translation from one source version with cross-checks against the other source version, or some other procedure?

The study suggested that translation procedures in international achievement studies have developed over the years and that as a consequence also the quality and equivalence of translated instruments has improved. This, of course, is good news. Conversely, however, it also means that the practices followed when translating, for example, the first achievement tests were not as developed as they are today and that the translations made at that time are not of as high quality as they are today. This may even be assumed to be more true of instruments translated into other languages than Finnish, which have always been judged to be of very high quality. The potentially lower quality of the early translations, in turn, casts doubts on the validity not only of those early studies but also of all those studies where materials from those studies have been or will be used as anchors (to provide trend data). Naturally, nothing can be done to improve the validity of the past studies. However, by making close linguistic examinations of the early translations, we can decide whether they are really of a sufficiently high quality to be used in future studies.

Even though translation procedures in international achievement studies have developed and the quality of translated instruments has improved, research in both translation studies and international achievement studies indicates that full equivalence can never be attained (e.g., Chesterman, 1997; Grisay, Gonzalez & Monseur, 2009). In addition, research in translation studies shows that the lack of equivalence is usually reflected in that translations are inferior to original, untranslated texts (see e.g., Chesterman, 2004, pp.

36–39); for example, they often contain less natural language. If this is indeed the case, it means that countries using the source instrument (especially Anglo-American countries, because usually it is the English version from which all other versions, including the French version, are translated) have an advantage over countries using translated instruments (cf. Grisay & Monseur, 2007). Careful linguistic comparisons are thus needed between source instruments and translated instruments. Moreover, if biases are found, steps need to be taken to do away with them. This might involve not only paying extra attention to the translation of the instruments and to the quality of the translated test versions, but possibly also making up for the biases when weighing and assessing the results.

Finally, the undisputed problems connected with translating international achievement tests and the fact that translations can never be fully equivalent to original texts but tend to be inferior to them have led some researchers (see e.g., Bonnet, 2002; Bonnet et al., 2003) to suggest that each country taking part in these studies use indigenous, untranslated materials. The comparability of these materials would then be evaluated by comparing them against a given set of criteria of item and text difficulty. This approach would undoubtedly improve the authenticity and naturalness of the translations, thereby also increasing equivalence. At the same time, however, the materials used by all the different countries would inevitably end up being so different from each other, for example, syntactically, lexically, textually, stylistically and contentwise that ensuring equivalence between them would be even more demanding than in the more traditional approach. Translated materials will thus continue to be needed in international achievement studies, and therefore they and the translation procedures leading to them need to be researched and developed further. Also, more research is required on and comparisons needed between the materials used in these two approaches in order to see whether it would be possible to combine the strengths of each approach.

References

Alderson, C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Allalouf, A. (2003). Revising translated differential item functioning items as a tool for improving cross-lingual assessment. *Applied Measurement in Education, 16* (1), 55–73.

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.

Arffman, I. (2002). *In search of equivalence – Translation problems in international literacy studies*. Unpublished master's thesis, University of Jyväskylä, Jyväskylä, Finland.

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.

Baker, M. (1992). *In other words. A coursebook on translation.* London: Routledge.

Bechger, T., van Schooten, E., de Glopper, C., & Hox, J. (1998). The validity of international surveys of reading literacy: The case of the Reading Literacy Study. *Studies in Educational Evaluation, 24*, 99–125.

Behr, D., & Scholz, E. (2011). Questionnaire transaltion in cross-national survey research. On the types and values of annotations. *Methoden – Daten – Analysen, 5* (2), 157–179.

Bell, R. T. (1991). *Translation and translating: Theory and practice*. London: Longman.

Blum, A., Goldstein, H., & Guérin-Pace, F. (2001). International Adult Literacy Survey (IALS): An analysis of international comparisons of adult literacy. *Assessment in Education, 8* (2), 225–246.

Bohm, D., Factor, D., & Garret, R. (1991). *Dialogue – A proposal.* Retrieved , April, 23, 2012, http://www.infed.org/archives/e-texts/bohm_dialogue.htm.

Bonnet, G. (2002). Reflections in a critical eye: On the pitfalls of international assessment [*Review of the book Knowledge and skills for life: First results from PISA 2000*]. *Assessment in Education, 9*, 387–399.

Bonnet, G., Daems, F., de Clopper, C., Horner, S. Lappalainen, H.-P., Nardi, E., Remond, M., Robin, I., Rosen, M., Solheim, R., Tonnessen, F.-E., Vertecchi, B., Vrignaud, P., Wagner, A., & White, J. (2003). *Culturally balanced assessment of reading (c-bar): A European project.* Paris: European Network of Policy Makers for the Evaluation of Education Systems.

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.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Brunette, L. (2000). Toward a terminology for translation quality assessment: A comparison of TQA practices. *The Translator, 6* (2), 169–182.

Campbell, S. (1999). A cognitive approach to source text: Difficulty in translation. *Target, 11*, 33–63.

Chesterman, A. (1997). *Memes of translation*. Amsterdam: Benjamins.

Chesterman, A. (2004). Beyond the particular. In A. Mauranen & P. Kujamäki (Eds.), *Translation universals? Do they exist* (pp. 33–49). Amsterdam: Benjamins.
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.
Colina, S. (2009). Further evidence for a functionalist approach to translation quality evaluation. *Target, 21* (2), 235–264.
Corey, S.M. (1953). *Action research to improve school practices*. New York: Teachers College Press.
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.
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.
Dorer, B. (2011). *Advance translation in the 5th round of the European Social Survey (ESS).* FORS Working Paper Series, paper 2011–4. Lausanne: FORS.
Elosua, P., & López-Jaúregui, A. (2007). Potential sources of differential item functioning in the adaptation of tests. *International Journal of Testing, 7* (1), 39–52.
Englund Dimitrova, B. (2005a). *Expertise and explicitation in the translation process*. Amsterdam: Benjamins.
Englund Dimitrova, B. (2005b). Literal translation in the translation process of professional translators. *Gothenburg Studies in English, 90*, 23–39.
Ercikan, K. (2002). Disentangling sources of differential item functioning in multilanguage assessments. *International Journal of Testing, 2* (3&4), 199–215.
Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F., & Lacroix, S. (2010). Application of think aloud protocols for examining and confirming sources of differential item functioning identified by expert reviews. *Educational Measurement: Issues and Practice, 29* (2), 24–35.
Ercikan, K., & Koh, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing, 5* (1), 23–25.
Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102* (2), 211–245.
Ericsson K.A., & Simon, H. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251.
European Social Survey. (2010). *ESS round 5 translation guidelines*. Mannheim, European Social Survey GESIS.
Fontanet, M. (2005). Temps de créativité en traduction. *Meta, 50*, 432–447.
Genest, M., & Turk, D. (1981). Think-aloud approaches to cognitive assessment. In T.V. Merluzzi, C.R. Glass & M. Genest (Eds.), *Cognitive assessment* (pp. 233–269). New York: Guilford Press.
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.
Grisay, A. (2002). Translation and cultural appropriateness of the test and survey material. In R. Adams, & M. Wu (Eds.), *PISA 2000 technical report* (pp. 57–70). Paris: OECD.
Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing 20* (2), 225–240.
Grisay, A. (2004). *PISA 2000: Differences in item difficulty between English, French and German countries*. Unpublished manuscript.
Grisay, A., de Jong, J., Gebhardt, E., Berezner, A., & Halleux-Monseur, B. (2007). Translation equivalence across PISA countries. *Journal of Applied Measurement, 8* (3), 249–266.
Grisay, A., Gonzalez, E., & Monseur, C. (2009). Equivalence of item difficulties across national versions in PIRLS and PISA reading assessments. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 2,* 63–84.
Grisay, A., & Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation, 33* (1), 69–86.
Hale, S., & Campbell, S. (2002). The interaction between text difficulty and translation accuracy. *Babel, 48*, 14–33.
Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. London: Falmer.
Hambleton, R. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10*, 229–224.

Hambleton, R. (2001). The next generation of ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*, 164–172.

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.

Hambleton, R. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. Hambleton, P. Merenda & C. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Erlbaum.

Hambleton, R., & Berberoglu, G. (1997). *TIMSS instrument adaptation process: A formative evaluation*. Amherst: University of Massachusetts, School of Education.

Hambleton, R., & 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*, 1–30.

Hambleton, R., & Zenisky, A. (2011). Translating and adapting tests for cross-cultural assessment. In D. Matsumoto & F. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 46–73). Cambridge: Cambridge University Press.

Harkness, J. (2003). Questionnaire translation. In J. Harkness, F. van de Vijver & P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 35–56). Hoboken, NJ: Wiley.

Harkness, J., Edwards, B., Hansen, S.E., Miller, D., & Villar, A. (2010a). Designing questionnaires for multipopulation research. 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. 33–57). Wiley: Hoboken, NJ.

Harkness, J., & Schoua-Glusberg, A. (1998). Questionnaires in translation. In J. Harkness (Ed.), *Cross-cultural survey equivalence* (ZUMA-Nachrichten Spezial, Band 3, pp. 87–126). Mannheim: ZUMA.

Harkness, J., Villar, A., & Edwards, B. (2010b). 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). Wiley: Hoboken, NJ.

Hassan, B. A. (2011). *Literary translation: Aspects of pragmatic meaning*. Cambridge: Cambridge Scholars.

Hatim, B., & Mason, I. (1997). *The translator as communicator*. London: Routledge.

Hayes, J., Flower, L., Schriver, K., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in psycholinguistics. Vol. 2. Reading, writing, and language learning* (pp. 176–240). Cambridge: Cambridge University Press.

Herr, K., & Anderson, G. (2005). *The action research dissertation*. Thousand Oaks: Sage.

Jeanrie , C., & Betrand, R. (1999). Translating tests with the International Test Commission's guidelines: Keeping validity in mind. *European Journal of Psychological Assessment, 15* (3), 277–283.

Jensen, A. (2000). *The effects of time on cognitive processes and strategies in translation*. Copenhagen: Copenhagen Business School, Faculty of Modern Languages.

Jensen, M. (2009). *Professional translators' establishment of skopos – A 'brief' study*. Unpublished master's thesis, Aarhus School of Business, Aarhus University, Denmark.

Johansone, I., & Malak, B. (2008). Translation and national adaptations of the TIMSS 2007 assessment and questionnaires. In J. Olson, M. Martin & I. Mullis (Eds.), *TIMSS 2007 technical report* (pp. 63–75). Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Jääskeläinen, R. (1996). Hard work will bear fruit? A comparison of two think-aloud protocol studies. *Meta, 16*, 60–74.

Jääskeläinen, R. (1999). *Tapping the process. An explorative study of the cognitive and affective factors involved in translating.* Joensuu: University of Joensuu.

Karg, I. (2005). *Mythos PISA. Vermeintliche Vergleichbarkeit und die Wirklichkeit eines Vergleichs* [The myth PISA. Alleged comparability and the reality of a comparison]. Gottingen: V&R unipress.

Kemper, S. (1983). Measuring the inference load of a text. *Journal of Educational Psychology, 75* (3), 391–401.

Kim, R. (2006). Use of extralinguistic knowledge in translation. *Meta, 51*, 284-303.

Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review, 5* (5), 363–394.

Kirsch, I.S. (2001). *The International Adult Literacy Survey (IALS): Understanding what was measured*. Princeton, NJ: Educational Testing Service.

Kleiner, B., Pan, Y., & Bouic, J. (2009). The impact of instructions on survey translation: An experimental study. *Survey Research Methods, 3* (3), 113–122.

Korning Zethsen, K. (1999). The dogmas of technical translation – Are they still valid? *Hermes, 23*, 65–75.
Krings, H. P. (1986). *Was in den Köpfen von Übersetzern vorgeht. Eine empirische Untersuchung der Struktur des Übersetzungsprozesses an fortgeschrittenen Französischlernern* [What goes on in the translator's head. An empirical study of the structure of the translation process in advanced students of French]. Tübingen: Narr.
Kuhiwczak, 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.
Künzli, A. (2007). Translation revision: A study of the performance of ten professional translators revising a legal text. In Y. Gambier, M. Schlesinger & R. Stolze (Eds.), *Doubts and directions in translation studies* (pp. 115–126). Amsterdam: Benjamins.
Kvale, S. (1996) *Interviews: An introduction to qualitative research interviewing*. London: Sage.
Ladkin, D. (2005). The enigma of subjectivity. *Action Research, 3* (1), 108–126.
Larson, M. (1998). *Meaning-based translation. A guide to cross-language equivalence* (2nd rev. ed.). Lanham, MD: University Press of America.
Laviosa-Braithwaite, S. (1998). Universals of translation. In M. Baker (Ed.), *Routledge encyclopedia of translation studies* (pp. 288–291). London: Routledge.
Leech, G. (1983). *Principles of pragmatics*. London: Longman.
Lefevere, A. (1992). *Translating literature. Practice and theory in a comparative literature context*. New York: The Modern Language Association of America.
Levý, J. (1967). Translation as a decision process. In J. Levý (Ed.), *To honor Roman Jakobson: Essays on the occasion of his seventieth birthday II* (pp. 1171–1182). The Hague: Mouton.
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.
Lörscher, W. (2005). The translation process: Methods and problems of its investigation. *Meta, 50*, 597–607.
Mackenzie, R. (1998). Creative problem-solving and translator training. In A. Beylard-Ozeroff, J. Králová & B. Moser-Mercer (Eds.), *Translators' strategies and creativity* (pp. 201–206). Amsterdam: Benjamins.
Martínez Melis, N., & Hurtado Albir, A. (2001). Assessment in translation studies: Research needs. *Meta, 46* (2), 272–287.
McKernan, J. (2000). *Curriculum action research* (2nd ed.). London: Kogan Page.
Mosenthal, P.B., & Kirsch, I.S. (1991). Toward an explanatory model of document literacy. *Discourse Processes, 14*, 147–180.
Mossop, B. (2007). *Revising and editing for translators.* Manchester, UK: St. Jerome.
Mullis, I., Martin, M., Kennedy, A., Trong, K., & Sainsbury, M. (2009). *PIRLS 2011 assessment framework*. Chestnut Hill, MA: Boston College, Lynch School of Education, TIMSS & PIRLS International Study Center.
Neubert, A., & Shreve, G. (1992). *Translation as text*. Kent, OH: Kent State University Press.
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.
Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.
Nord, C. (1991). *Text analysis in translation. Theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam: Rodopi.
Nord, C. (1997). A functional typology of translations. In A. Trosborg (Ed.), *Text typology and translation* (pp. 43–66). Amsterdam: Benjamins.
Nord, C. (2006). Loyalty and fidelity in specialized translation. *Confluências, 4*, 29–41.
OECD. (2005). *PISA 2003 technical report.* Paris: Author. Retrieved, June, 28, 2012, http://www.oecd.org/dataoecd/49/60/35188570.pdf.
OECD. (2007, September). *PISA 2009 translation and adaptation guidelines.* Paper presented at the National Project Managers' Meeting, Dubrovnik, Croatia. Retrieved, June, 28, 2012, http://www.oecd.org/dataoecd/15/35/49023637.pdf.
OECD. (2009a). *PISA 2006 technical report.* Paris: Author. Retrieved, June, 28, 2012, http://www.oecd.org/dataoecd/0/47/42025182.pdf.
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.
OECD. (2009c). *Take the test. Sample questions from OECD's PISA assessment.* Paris: Author. Retrieved, June, 28, 2012, http://www.oecd.org/dataoecd/47/23/41943106.pdf.

OECD. (2009, April). *Outcomes of the JRA pilot study: Annex 1*. Paper presented at the 4th meeting of PIAAC BPC, 2-3 April, Paris. Retrieved, June, 28, 2012, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=COM/DELSA/EDU/PIAAC(2009)2/ANN1&docLanguage=En.

OECD. (2010a). *Computer-based assessment – Field trial PISA translation management system manual*. Paper presented at the National Project Managers' Meeting, Budapest, Hungary. Retrieved, June, 28, 2012, http://www.oecd.org/dataoecd/62/35/49273581.pdf

OECD. (2010b). *PISA 2009 results: What students know and can do – Student performance in reading, mathematics and science*. Volume I. Retrieved, June, 28, 2012, http://www.oecd.org/dataoecd/10/61/48852548.pdf.

OECD. (2010c, October). *Translation and adaptation guidelines for PISA 2012*. Paper presented at the National Project Managers' Meeting, Budapest, Hungary. Retrieved, June, 28, 2012, http://www.oecd.org/dataoecd/63/12/49273486.pdf

OECD. (2012). *PISA 2009 technical report*. PISA, OECD Publishing. Retrieved, June, 28, 2012, http://dx.doi.org/10.1787/9789264167872-en.

PACTE. (2005). Investigating translation competence: Conceptual and methodological issues. *Meta, 50*, 609–619.

Pan, Y., Kleiner, B, & Bouic, J. (2007). *The impact of Instructions on survey translation: An experimental study* (Survey Methodology 18). Retrieved April, 17, 2012, http://www.census.gov/srd/papers/pdf/rsm2007-18.pdf.

Patton, M.Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.

Pym, A. (1995). European translation studies, Une science qui dérange, and Why equivalence needn't be a dirty word. *TTR, 8*, 153–176.

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.

Reiss, K., & Vermeer, H. (1984). *Grundlegung einer allgemeinen Translationstheorie* [Groundwork for a general theory of translation]. Tübingen: Niemeyer.

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.

Rueda, R. (2011). Cultural perspectives in reading: Theory and research. In M. Kamil, P.D. Pearson, E.B. Moje & Afflerbach, P. (Eds.), *Handbook of reading research IV* (pp. 84–104). New York: Routledge.

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.

Séguinot, C. (1989). Pragmatics and the explicitation hypothesis. *TTR, 1* (2), 106–113.

Shreve, G. (1997). Cognition and the evolution of translation competence. In J. Danks, G. Shreve, S. Fountain & M. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 120–136). Thousand Oaks, CA: Sage.

Siniscalco, M. (2006). What are the national costs for a cross-national study? In K. Ross & I. Genevois (Eds.), *Cross-national studies of the quality of education: planning their design and managing their impact* (pp. 185–209). Paris: UNESCO, International Institute for Educational Planning.

Solano-Flores, G., Backhoff, E., & Contreras-Niño, L. (2009). Theory of test translation error. *International Journal of Testing, 9*, 78–91.

Solano-Flores, G., Contreras-Niño, L. & Backhoff-Escudero, E. (2006). Translation and adaptation of tests: Lessons learned and recommendations for countries participating in TIMSS, PISA and other international comparisons. *Revista electrónica de investigación educativa, 8* (2). http://redie.uabc.mx/vol8no2/contents-solano2.html.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12* (2), 257–285.

Tirkkonen-Condit, S. (2005). The monitor model revisited: Evidence from process research. *Meta, 50*, 405–413.

Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

Toury, G. (1995). *Descriptive translation studies and beyond*. Amsterdam: Benjamins.

van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage.

van Dijk, T., & Kintsch, W. (1978). Toward a model of text comprehension and production. *Psychological Review, 85* (5), 363–394.

Vermeer, H. J. (1989). Skopos and commission in translational action. In A. Chesterman (Ed.), *Readings in translation theory* (pp. 173–187). Helsinki: Finn Lectura.

Vinay, J.-P., & Darbelnet, J. (2004). A methodology for translation. In L. Venuti (Ed.) *The translation studies reader* (2nd ed., pp. 84–93). London: Routledge.

Wechsler, R. (1998). *Performing without a stage: The art of literary translation*. North Haven, CT: Catbird.

Willis, G., Stapleton Kudela, M., Levin, K., Norberg, A., Stark, D., Forsyth, B., Brick, P., Berrigan, D., Thompson, F., Lawrence, D., & Hartman, A. (2010). Evaluation of a multiple step translation process. 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. 141–156). Wiley: Hoboken, NJ.

Wilss, W. (1990). Cognitive aspects of the translation process. *Language & Communication, 10*, 19–36.

Wilss, W. (1998). Decision making in translation. In M. Baker & K. Malmkjaer (Eds.), *Routledge encyclopedia of translation studies* (pp. 57–62). London: Routledge.

Zakay, D. (1993). The impact of time perception processes on decision-making under time stress. In O. Svenson & A. J. Maule (Eds.), *Time pressure and stress in human judgment and decision making* (pp. 59–72). New York: Plenum.

APPENDIX 1
EXAMPLE OF A PISA 2009 READING LITERACY UNIT, ACCOMPANIED BY TRANSLATION NOTES AND SCORING RUBRICS

THE PLAY'S THE THING

Takes place in a castle by the beach in Italy.

FIRST ACT

Ornate guest room in a very nice beachside castle. Doors on the right and left. Sitting room set in the middle of the stage: couch, table, and two armchairs. Large windows at the back. Starry night. It is dark on the stage. When the curtain goes up we hear men conversing loudly behind the door on the left. The door opens and three tuxedoed gentlemen enter. One turns the light on immediately. They walk to the centre in silence and stand around the table. They sit down together, Gál in the armchair to the left, Turai in the one on the right, Ádám on the couch in the middle. Very long, almost awkward silence. Comfortable stretches. Silence. Then:

GÁL

Why are you so deep in thought?

TURAI

I'm thinking about how difficult it is to begin a play. To introduce all the principal characters in the beginning, when it all starts.

ÁDÁM

I suppose it must be hard.

TURAI

It is – devilishly hard. The play starts. The audience goes quiet. The actors enter the stage and the torment begins. It's an eternity, sometimes as much as a quarter of an hour before the audience finds out who's who and what they are all up to.

GÁL

Quite a peculiar brain you've got. Can't you forget your profession for a single minute?

TURAI

That cannot be done.

GÁL

Not half an hour passes without you discussing theatre, actors, plays. There are other things in this world.

TURAI

There aren't. I am a dramatist. That is my curse.

GÁL

You shouldn't become such a slave to your profession.

TURAI

If you do not master it, you are its slave. There is no middle ground. Trust me, it's no joke starting a play well. It is one of the toughest problems of stage mechanics. Introducing your characters promptly. Let's look at this scene here, the three of us. Three gentlemen in tuxedoes. Say they enter not this room in this lordly castle, but rather a stage, just when a play begins. They would have to chat about a whole lot of uninteresting topics until it came out who we are. Wouldn't it be much easier to start all this by standing up and introducing ourselves? *Stands up.* Good evening. The three of us are guests in this castle. We have just arrived from the dining room where we had an excellent dinner and drank two bottles of champagne. My name is Sándor Turai, I'm a playwright, I've been writing plays for thirty years, that's my profession. Full stop. Your turn.

GÁL

Stands up. My name is Gál, I'm also a playwright. I write plays as well, all of them in the company of this gentleman here. We are a famous playwright duo. All playbills of good comedies and operettas read: written by Gál and Turai. Naturally, this is my profession as well.

GÁL and TURAI

Together. And this young man …

ÁDÁM

Stands up. This young man is, if you allow me, Albert Ádám, twenty-five years old, composer. I wrote the music for these kind gentlemen for their latest operetta. This is my first work for the stage. These two elderly angels have discovered me and now, with their help, I'd like to become famous. They got me invited to this castle. They got my dress-coat and tuxedo made. In other words, I am poor and unknown, for now. Other than that I'm an orphan and my grandmother raised me. My grandmother has passed away. I am all alone in this world. I have no name, I have no money.

TURAI

But you are young.

GÁL

And gifted.

ÁDÁM

And I am in love with the soloist.

TURAI

You shouldn't have added that. Everyone in the audience would figure that out anyway.

They all sit down.

TURAI

Now wouldn't this be the easiest way to start a play?

GÁL

If we were allowed to do this, it would be easy to write plays.

TURAI

Trust me, it's not that hard. Just think of this whole thing as …

GÁL

All right, all right, all right, just don't start talking about the theatre again. I'm fed up with it. We'll talk tomorrow, if you wish.

Translation Note: Translate the unit heading "The Play's the Thing" as from Act II, Scene ii of Shakespeare's *Hamlet*. If this is not possible in your language, translate the phrase to connote the idea that the play is the focal point.

"The Play's the Thing" is the beginning of a play by the Hungarian dramatist Ferenc Molnár.

Use "The Play's the Thing" on the previous two pages to answer the questions that follow. (Note that line numbers are given in the margin of the script to help you find parts which are referred to in the questions.)

Question 1: THE PLAY'S THE THING

R452Q01

What does Turai say about his profession?

A It is too difficult for him to manage.
B It is a good way to make money.
C It is his whole life.
D It is exciting and wonderful.

THE PLAY'S THE THING SCORING 1

QUESTION INTENT:

Integrate and interpret: Develop an interpretation
Infer the attitude of a character in a play

Full Credit

Code 1: C. It is his whole life.

No Credit

Code 0: Other responses.

Code 9: Missing.

Question 3: THE PLAY'S THE THING

R452Q03 – 0 1 9

What were the characters in the play doing **immediately before** the curtain went up?

....

THE PLAY'S THE THING SCORING 3

QUESTION INTENT:

Access and retrieve: Retrieve information
Locate a reference to action taking place before the events of a play

Full Credit

Code 1: Refers to dinner or drinking champagne. May paraphrase or quote the text directly.

- They have just had dinner and champagne.

- "We have just arrived from the dining room where we had an excellent dinner."*[direct quotation]*
- "An excellent dinner and drank two bottles of champagne." *[direct quotation]*
- Dinner and drinks.
- Dinner.
- Drank champagne.
- Had dinner and drank.
- They were in the dining room.

No Credit

Code 0: Gives an insufficient or vague response.

Shows inaccurate comprehension of the material or gives an implausible or irrelevant response.

- The three of us are guests in this castle.
- They converse loudly behind the door. *[This is part of the first act, not before it.]*
- They got Adam's dress-coat and tuxedo made for him. *[not **immediately** before the events of the text]*
- Got ready to come on stage. *[Refers to the actors rather than the characters.]*
- Takes place in a castle by the beach in Italy.
- Talking about the theatre.

Code 9: Missing.

Question 4: THE PLAY'S THE THING

R452Q04

"It's an eternity, sometimes as much as a quarter of an hour … " (lines 29-30)

According to Turai, why is a quarter of an hour "an eternity"?

A It is a long time to expect an audience to sit still in a crowded theatre.
B It seems to take forever for the situation to be clarified at the beginning of a play.
C It always seems to take a long time for a dramatist to write the beginning of a play.
D It seems that time moves slowly when a significant event is happening in a play.

Translation Note: Adjust line numbers if necessary.

THE PLAY'S THE THING SCORING 4

QUESTION INTENT:

Integrate and interpret: Develop an interpretation

Infer the meaning of a phrase in a play using contextual references

Full Credit

Code 1: B. It seems to take forever for the situation to be clarified at the beginning of a play.

No Credit

Code 0: Other responses.

Code 9: Missing.

Question 6: THE PLAY'S THE THING

R452Q06 – 0 1 9

A reader said, "Ádám is probably the most excited of the three characters about staying at the castle."

What could the reader say to support this opinion? Use the text to give a reason for your answer.

....

....

THE PLAY'S THE THING SCORING 6

QUESTION INTENT:

Integrate and interpret: Develop an interpretation

Support an opinion by construing a character's motivation in a play

Full Credit

Code 1: Indicates a <u>contrast between Ádám and the other two characters</u> by referring to one or more of the following: <u>Ádám's status as the poorest or youngest of the three characters, his inexperience as a celebrity.</u>

- Ádám is poor, he must be excited to stay at a fancy castle.
- He must be happy to be with the two guys who can make him famous.
- He is writing music with two really famous people.
- He is young, and young people just get more excited about things, it's a fact!
- He's young to stay at the castle. *[minimal]*

No Credit

Code 0: Gives an <u>insufficient or vague</u> response.

- He is excited. *[Repeats stem.]*

Shows <u>inaccurate comprehension</u> of the material or gives an <u>implausible or irrelevant</u> response.

- He is an artist.
- He has fallen in love. *[not an explanation of why he is excited to be staying at the castle]*
- Ádám must be excited; surely the soloist will show up. *[no support in the text]*
- He has been given a tuxedo. *[an explanatory detail, not the reason itself]*

Code 9: Missing.

Question 7: THE PLAY'S THE THING

R452Q07

Overall, what is the dramatist Molnár doing in this extract?

A He is showing the way that each character will solve his own problems.
B He is making his characters demonstrate what an eternity in a play is like.
C He is giving an example of a typical and traditional opening scene for a play.
D He is using the characters to act out one of his own creative problems.

Translation Note: Use a verb form for "is doing" etc. that suggests a process rather than a single action.

THE PLAY'S THE THING SCORING 7

QUESTION INTENT:

Integrate and interpret: Form a broad understanding

Recognise the conceptual theme of a play

Full Credit

Code 1: D. He is using the characters to act out one of his own creative problems.

No Credit

Code 0: Other responses.

Code 9: Missing.

Question 9: THE PLAY'S THE THING

R452Q09

Which of the following sentences best describes the way the characters behave in this extract?

A They are sentimental and emotional.
B They are witty and light-hearted.
C They are polite and reserved.
D They are nervous and anxious.

THE PLAY'S THE THING SCORING 9

QUESTION INTENT:

Integrate and interpret: Form a broad understanding

Recognise the tone created by the characters in a play.

Full Credit

Code 1: B. They are witty and light-hearted.

No Credit

Code 0: Other responses.

Code 9: Missing.

APPENDIX 2
OUTLINE OF THE DISCUSSION

A. 1. How did it feel to translate?
   2. How do you think you succeeded?
   3. How do you think the entire translation process went?

B. 1. Which of the texts were the most difficult to translate?
   2. What kinds of questions were the most difficult to translate?

C. 1. Did you have enough time to translate?
   2. How much time (approximately) did you spend translating one text?

D. What were the factors that caused the most difficulties? Why?

E. 1. How well were you able to follow the PISA translation instructions?
   2. Which of the instructions were the most difficult to follow?

F. When faced with a problem what did you do to solve it?

G. Ideas, proposals for improvement, wishes, questions …

H. Something you want to add?

Translating International Achievement Tests

TRANSLATORS' VIEW

TRANSLATING international achievement tests is a responsible and demanding task: For the interpretations made on the basis of the test to be valid, all the translations and different-language versions of the test need to be equivalent, or comparable, to each other. Rigorous translation procedures and practices have been developed to ensure that this would be the case. How, then, have the translation procedures and practices really worked? Have they been easy to follow? Have they helped to ensure high-quality and equivalent translations? Or have there been problems in them?

This book reports on a study that was carried out to answer these questions. Discussions were conducted with the five translators translating the PISA 2009 reading test into Finnish. On the basis of the findings, suggestions are given on how to develop translation procedures and practices in these studies.

The book will provide important reading for all those involved or interested in international achievement studies. It will also be of relevance to those concerned with or intrigued by other types of cross-cultural studies and translation.

Reports 44
ISBN 978-951-39-4811-5

Cover: Martti Minkkinen