Meta
Journal des traducteurs
Translators' Journal

META

Deconstructing the Translation of Psychological Tests

Alicia Bolaños-Medina and Víctor González-Ruiz

Cite this article

Article abstract

The translation of psychological tests has become widespread as the globalization process has led to testing instruments designed in one country being applied in a different one relatively quickly, since it is less expensive and faster to adapt an existing instrument than to devise a new one in another culture. In order to provide a fully functional version of a test in a different language and culture, the whole cultural context within which a particular test is to be used must be considered, and it is not certain that this has always been the case. This review article explains the peculiarities of psychological tests as a genre and documents the process of translation which has traditionally been used. It is shown that while psychologists have increasingly become aware of the role of cultural context, they have mistakenly regarded this as an issue which is separate from the translation process. That concern for the cultural context is part and parcel of translation itself, is illustrated both in modern translation theory and in current translation practice.

Deconstructing the Translation of Psychological Tests

ALICIA BOLAÑOS-MEDINA
Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
abolanos@dfm.ulpgc.es

VÍCTOR GONZÁLEZ-RUIZ
Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
vgonzalez@dfm.ulpgc.es

RÉSUMÉ

La traduction de tests psychologiques ne cesse de se développer, étant donné que la mondialisation favorise assez facilement l'application d'outils conçus pour un pays donné dans un autre. Il est également moins cher et plus rapide d'adapter un instrument déjà existant que d'en créer un nouveau. Afin de produire une version d'un test pleinement fonctionnelle dans une autre langue et une autre culture, le contexte culturel dans lequel il sera utilisé doit être pris en considération, ce qui n'a pas toujours été le cas. Cet article de revue a pour objectif d'expliquer les particularités des tests psychologiques en tant que genre textuel et de décrire le processus de traduction traditionnellement utilisé. Nous y tentons de démontrer que, bien que les psychologues soient de plus en plus conscients du rôle du contexte culturel, ils l'ont malheureusement trop souvent mis à l'écart du processus même de traduction. Or, les théories modernes de la traduction et la pratique sont la preuve que cette préoccupation pour le contexte culturel fait partie intégrante dudit processus.

ABSTRACT

The translation of psychological tests has become widespread as the globalization process has led to testing instruments designed in one country being applied in a different one relatively quickly, since it is less expensive and faster to adapt an existing instrument than to devise a new one in another culture. In order to provide a fully functional version of a test in a different language and culture, the whole cultural context within which a particular test is to be used must be considered, and it is not certain that this has always been the case. This review article explains the peculiarities of psychological tests as a genre and documents the process of translation which has traditionally been used. It is shown that while psychologists have increasingly become aware of the role of cultural context, they have mistakenly regarded this as an issue which is separate from the translation process. That concern for the cultural context is part and parcel of translation itself, is illustrated both in modern translation theory and in current translation practice.

MOTS-CLÉS/KEYWORDS

traduction de tests, rétrotraduction, traduction renversée, équivalence traductionnelle, genre textuel

test translation, back translation, forward translation, translation equivalence, textual genre

1. Introduction

In this article we address the intricacies of the translation of psychological tests, a complex task which involves interdisciplinary competences. Although it is not uncommon to regard psychological testing as a relatively recent Western development, its roots date back to the concepts and practices of ancient China 3,000 years ago (Anastasi 1988). Methods such as observing behavioral changes, measuring response speed as a key intelligence factor, calling forth personality traits across situations, and using interviews in order to gauge mental attributes, allowed the Chinese emperor to assess his officials' fitness for office at the time (Higgins and Zheng 2002). In the Western world, Francis Galton, the renowned English Victorian psychometrician, whose aim was to improve human breeding by selecting the best and brightest individuals (through measuring sensory reactions and reaction times in volunteers), pioneered test assessment in London in the late nineteenth century (Bulmer 2003). Not much later, in France, researchers Alfred Binet and Théodore Simon designed the first intelligence test which was to be used to detect – mentally defective – students, as they were referred to then. In this case, memory, visual traits, imagination and language skills were the mental constructs to be measured (Binet and Simon 1954). From New York, Edward Thorndike, whose work on animal behavior helped lay the scientific foundation for modern educational psychology (Thorndike 1910), contributed to test development not only by devising scales to assess students' performance in reading and mathematics, but also by constructing the tests which sorted out recruits for particular tasks when the First World War broke out in 1914 (Gregory 2010). After the war, psychologists continued to develop new instruments for educational settings; by that time, the notion of testing had gained prestige and was firmly established in the practice of applied psychology, as it remains today.

The use of psychological tests has become more and more widespread as the globalization process, involving all sectors of human activities, has led to testing instruments designed in one country being applied in a different one relatively quickly (Muñiz and Hambleton 1996). This growing interest in tests comes from the fact that they are perceived to be useful in a variety of settings, such as educational institutions, career and counselling services, the workplace or research institutes (Jackson 1996). In every case, the information retrieved is considered with regard to four main uses: classification, evaluation of programmes, the promotion of self-understanding and scientific inquiry (Cronbach 1990). Given the traditional hegemony of Western psychology, with the United States in the lead, the vast majority of tests are usually devised in English and, in the first place, for Western societies. According to Hambleton (1993), there are at least two reasons for translating tests. Firstly, they are translated because it is less expensive and faster to adapt an existing instrument than to devise a new one to measure the same construct in another culture. A second reason is to implement cross-national studies.

Psychological test translation involves the management of different sorts of documents; in fact, the range of translation segments professionals must deal with may be very varied: yes/no questions such as "I feel anxiety about something or someone almost all the time" (Butcher, Dahlstrom *et al.* 1989); a single word on a vocabulary card which must be defined by the test taker (and must present an equivalent degree of difficulty for individuals in different cultures); a complex statis-

tical formula in a technical manual; a question asking individuals to explain the meaning of a particular proverb; or a set of directions for different addressees with various degrees of knowledge about the functioning of test assessment (usually the test administrator and the test taker).

What all these types of segments belonging to different subgenres have in common is that, in order to be adequately translated, the whole cultural context within which a particular test is to be used must be considered and, according to the International Test Commission Guidelines on Adapting Tests, this must be done "taking full account of linguistic and cultural differences among the populations for whom adapted versions of the instrument are intended" (ITC 2000). Furthermore, test publishers must make sure that testing techniques, item formats, test conventions, item content and stimulus materials of adapted versions of a test are familiar to all intended populations. Additionally, if we consider intelligence and personality tests, which are supposed to measure an individual's potential rather than their achievements, it is clearly highly desirable that they should be culturally neutral or even culture-free; however, this has proved so far to be a desideratum that is difficult to achieve.

Although both "test adaptation" and "test translation" are used by researchers, it should be noted that, following Hambleton, we will adopt "test adaptation" in the broader sense, to include all activities which may be necessary to provide a fully functional version of a test in a different language and culture,

> from deciding whether or not a test could measure the same construct in a different language and culture, to selecting translators, to deciding on appropriate accommodations to be made in preparing a test for use in a second language, to adapting the test and checking its equivalence in the adapted form. (Hambleton 2005: 4)

Before describing the process of translation which is traditionally used, we will introduce and describe the peculiarities of this textual genre. Then, we will reflect upon some key factors from the point of view of translatology.

2. Psychological tests as a textual genre

According to the Standards for Educational and Psychological Testing issued by the American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME), a test is defined as "an evaluative device or procedure in which a sample of an examinee's behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process" (AERA, APA and NCME 2008: 3). By "standardized" we understand that instructions provided to examinees, general testing conditions and scoring procedures must follow the same strict and specific scheme. Although this definition could apply to other concepts, such as *scale* (i.e., a set of statements measuring the degree to which people agree or disagree with them), *questionnaire* (i.e., a set of questions for obtaining personal information from individuals), or *inventory* (i.e., a list of traits, preferences, attitudes, interests or abilities used to evaluate personal characteristics or skills), for the purposes of this article, the term *test* will be used to refer to all types of educational and psychological instruments.

There are several categories of psychological tests, in terms of the construct they intend to measure (Bolaños-Medina 2012). Achievement and aptitude tests are usually seen in educational or employment settings, and they attempt to measure either how much the examinee knows about a certain topic, or to what degree he or she has the capacity to master material in a particular area. Intelligence tests aim at measuring an individual's basic and potential ability to understand the world around them, assimilate its functioning, and apply this knowledge to enhance their quality of life. Neuropsychological tests are designed to measure deficits in cognitive functioning (one's ability to reason, construct speech, etc.) which may result from some sort of brain damage. Occupational tests are applied to match the test taker's interests with those of people in known careers, in order to find out which profession suits best. Personality tests attempt to calibrate the personality style of the test taker, and are used for research or diagnosis purposes. Finally, clinical tests measure specific clinical aspects, such as an individual's level of anxiety or depression.

Psychological tests as a whole constitute a technical genre, in other words, a class of communicative event that takes place in a given communicative situation with a particular purpose and which presents a characteristic pattern of textual conventions in terms of schematic structure, style, content and intended audience (Swales 1990). This particular genre presents a primary exhortative contextual focus, according to Gamero Pérez's classification (2001) following Hatim and Mason (1990), who define the aim of this primary focus as the formation of future behaviors through the regulation of action and thought through instructions. In some subgenres, this can be accompanied by a secondary expositive focus (e.g., technical description). As in any other technical genre, the efficacy of communication is fundamental, and it is achieved through a concise, clear and precise language use. Any lack of clarity in the instructions or failure to use comprehensible language could significantly influence an individual's responses.

Many participants interact in the communicative situation of the production and reception of psychological tests: those who develop the test, those who publish and market it, those who administer and score it, those who use test results in order to make a decision, those who interpret results for other clients, those who take the test, whether this is because they need to or by choice or direction, those who sponsor the test and those who compare and select tests for a particular purpose (AERA, APA and NCME 2008). These communicative roles are not always well-defined and some of them may be combined in a single role, for instance, developer and user.

Test producers and translators must be well aware of the fact that each participant is supposed to have a different level of understanding of test development and use procedures, which is best illustrated if we compare the level of understanding of those who administer and score the test with that of test takers. On the other hand, although all the participants should "possess the knowledge skills and abilities relevant to their role in the testing process, as well as awareness of personal and contextual factors that may influence the testing process" (AERA, APA and NCME 2008: 2), this may not be always the case.

Psychological tests as a text genre involve different subgenres. To start with, the materials of the test itself, also known as *stimulus* material, may adopt multiple formats according to the principles of functioning of the test. Together with these elements, as a crucial component for correct test application, tests are accompanied

by supporting documentation, usually in the shape of a manual. For instance, the complete kit of Weschler Adult Intelligence Scale (WAIS-III),[1] includes all necessary equipment (comprising 5 object assemblies, a set of vocabulary cards, 5 puzzles, cartoon cards, 9 small red and white blocks) plus a stimulus book, an administration and scoring manual, a technical manual, 25 record forms, 25 response booklets and a briefcase.

The stimulus book contains all the material provided as part of a test item or task, to which the test taker has to respond. Stimulus material generally includes input texts, questions or illustrations with tokens capable of generating the nervous system activity or response. This material can be recorded in other media; in the past, audiotapes were relatively frequent, but, nowadays, many tests are also available in multimedia format.

Instruments may adopt a single item format, or feature a combination of them: multiple choice items, true-false questions, matching format, completion format, short-answer format, cloze-procedure (i.e., filling the gaps with one of several alternatives), rating scales and checklists (Osterlind 1997; Barbero García, Vila Abad *et al.* 2006). In the case of tests measuring intelligence and cognitive abilities, mostly through performance, other specific types of stimuli exist, most of which involve the use of pictorially presented shapes and concepts and even real objects, besides verbal elements. Following the classification described by Magno (2009), these are verbal analogy, syllogism, number/letter series, topology (test takers are asked to select from different visual options which duplicates the condition presented), visual discrimination, progressive series, classes of visualization (for examinees to figure out how specific pictures will look if rotated, twisted or inverted), orientation (to maintain accurate perception of a pattern when confronted with changing orientations), figure and ground perception, surface development, object assembly and picture completion.

The communicative intention of supporting documents for tests is to provide test users with the information they need to gauge the quality of a test and the interpretation based on test scores. Supporting documents can include test manuals, technical manuals, user's guides, specimen sets, examination kits, directions for test administrators and scorers, or preview materials for test takers. Furthermore, separate documents of sections are often written for certain categories of users (e.g., practitioners, researchers), employing a combination of all the elements mentioned above and usually including the intended test-taking population, the test purpose and other specifications, item formats, scoring procedures, the test development process, technical data and cut scores (AERA, APA and NCME 2008).

As far as supporting materials contents are concerned, according to the Standards for Educational and Psychological Testing, they usually comprise:

> the nature of the test; its intended use; the processes involved in the test's development; technical information related to scoring, interpretation, and evidence of validity and reliability; scaling and norming if appropriate to the instrument; and guidelines for test administration and interpretation. (AERA, APA and NCME 2008: 67)

Supporting information can be presented in more than one manual. As we have said, WAIS-III includes both a technical manual and an administration and scoring manual. The technical manual is a document created by the test authors and publish-

ers in order to provide all the technical and psychometric information needed to use a particular test.

The introductory chapter of the *WAIS Administration and Scoring Manual*[2] explains its basic principles of use (scope, examiners' requirements, standardized procedure, timing, environmental conditions, materials, special instructions for handicapped people, test-retest time span, and aspects concerning abridged versions), its scope, specific aspects of application (order of item presentation, time control, item repetition, among others), general principles of scoring and answering sheet management (chronological age calculation, answers and scoring registration, scoring conversion intelligence quotient, or IQ, calculation, profile obtaining, among others). The second chapter comprises the specific instructions for the application of every subtest.

Record forms are designed to register complete demographic and general behavioral observations, calculate age correctly, record answers to every subtest verbatim, and score subtest items according to the manual. The standardized answer sheets can be hand-scored with templates but, nowadays, most tests are computer-scored and the specific software developed for that purpose also has to be localized.

Although psychological tests are very varied, depending on the constructs they intend to measure and the principles sustaining them, some common discursive traits can still be pinpointed. Instructions are given in a straightforward manner, reflecting the main characteristics of common technical discourse, such as the use of imperative, present tenses and descriptive verbs, in an attempt at an impersonal and objective style. Syntax is plain, frequently comprising several simple independent sentences. Frequent use of general vocabulary, and a tendency to exclude technical terms and acronyms which could create ambiguity, are also common. Examples of how to complete the forms correctly are usually given.

The style is concise, the expression tends to be precise, and the register is far from formal; in fact it is sometimes even colloquial, with the deliberate use of everyday expressions. Very often, text is subordinate to visual and ideographic stimuli, together with real, palpable objects and shapes. Clarity and simplicity prevail, and the avoidance of the use of double negatives and excessively protracted statements has traditionally been recommended, since items must be representative, relevant, diverse, clear, simple and understandable (Muñiz and Fonseca-Pedrero 2009).

3. Translation as an integral part of the process of test adaptation

3.1. Test adaptation from a historical perspective

We have already mentioned the growing interest of psychologists in applying tests cross-culturally. By introducing different national perspectives into their studies, they have two aims: on the one hand, to broaden the quality and scope of the personality, aptitude or achievement measurements; and, on the other, to advance the theoretical and applied scientific knowledge in their field. Likewise, the translation of an existing test into a new language in order to apply it in a different cultural setting may be seen as a convenient shortcut for the target culture as regards economical and technical aspects. From a historical perspective, this cross-cultural trend has increased rapidly, particularly in the second half of the 20th century. As early as 1911,

the *Échelle métrique de l'intelligence*, an intelligence test developed by French researchers Binet and Simon, was translated into the English language (known as the Binet-Simon *Intelligence Scale for Children*); and, by 1916, this scale had already been translated into seven other languages. However, since then, the trend has been to adapt instruments which were originally developed in English into other cultures, as a possible reflection of the dominance of US and British researchers in the field of testing. Among the most popular instruments which have been translated into several languages, Hambleton (1993) draws attention to the *State-Trait Anxiety Inventory for Adults*, reportedly the most widely used self-report measure of anxiety, which was developed, among others, by the American psychologist Charles D. Spielberger; and the *Weschler Adult Intelligence Scale*, a general test of adult intelligence which has already been described in this paper, and which was conceived by another American researcher, David Wechsler.

With respect to particular areas of testing, Hambleton (1993) indicates that the International Association for the Evaluation of Educational Achievement has carried out cross-national studies in the sphere of educational achievement and school attitudes for over thirty years, with the purpose of influencing educational policy in the different countries involved. Besides, Marsella, Dubanosky *et al.* point out that, as far as cultural anthropological and personality studies are concerned, "it was not until the 1970s that psychologists finally began to confront the risks of cultural bias in their studies on a large scale and to evolve conceptual models and methodologies that could minimize these risks" (Marsella, Dubanosky *et al.* 2000: 44). These authors seem to disapprove of the fact that pre-1970s psychologists were generally unaware of cultural differences when they "indiscriminately" administered personality scales to people from non-Western countries, thus ignoring "the possibility of cultural bias in the very nature of [the psychologists'] concepts, scales, and norms" (Marsella, Dubanosky *et al.* 2000: 44). Likewise, Hambleton (1993) refers to Brislin (1970) to indicate that, at least until 1970, there was little evidence that researchers tried to establish translation equivalence or to identify potentially biased items, a defect which would cast some shadows on the validity of significant portions of cross-cultural research studies up through 1970. In a later work, Hambleton adds that this assertion may also be true of research carried out in the 1980s and 1990s, and makes an allusion to some cross-cultural researchers who have suggested that a great part of the investigation in their field should be dismissed as invalid because of the deficiencies in the test adaptation process (Hambleton 2005: 4).

All this notwithstanding, firm steps have been taken in recent years in order to improve cross-national psychological testing. The most outstanding of these steps has been the establishment by the International Test Commission of a committee of psychologists from a number of international associations.[3] In the early 1990s, this committee was commissioned to develop a set of technical guidelines for the translation of tests and the setting of test score equivalence. As a result of their work, the ITC Guidelines for Test Adaptation were unveiled in 2000, covering four sections which account for the whole process of adapting a test into a foreign language/culture: context, test development and adaptation, administration, and documentation/score interpretations. These guidelines, whose creation was led by researcher Ronald K. Hambleton, constitute a thorough approach to the adaptation of psychological instruments, and are often alluded to in this paper.

3.2. *Types of equivalence in test adaptation*

As regards our object of study, "two versions of an item when prepared in different languages are assumed to be equivalent when members of each group of the same ability have the same probability of success on the item"; if the probabilities vary, "the item is labelled 'potentially biased'" (Hambleton 1993: 62). Such equivalence is understood to be composed of several distinguishable layers, which range from the correlation of words and cultural references, to the careful selection of scales and modes of test administration. In this context, therefore, linguistic equivalence is only one of the many aspects to be addressed when assessing the cross-cultural correspondence of instruments.

Even though there is no single accepted nomenclature to refer to the various levels of equivalence, some common features may be inferred. To start with, most researchers emphasize that construct equivalence must occur as a prerequisite for any cross-cultural study. Harkness and Schoua-Glusberg (1998) describe construct equivalence as encompassing conceptual/functional equivalence and equivalence in the way the construct measured by the test is operationalized in each language/cultural group. In other words, Herdman, Fox-Rushby *et al.* (1998: 324), referring to this type of correspondence as "conceptual equivalence," describe it as follows: "[this kind of equivalence] is achieved when the questionnaire has the same relationship to the underlying concept […] in both cultures, primarily in terms of the domains included and the emphasis placed on different domains." By way of illustration, the construct "quality of life" may involve different experiences in different cultures (e.g., abundance of material items vs. access to immediate first aid); as a consequence, a researcher who wishes to adapt a test measuring aspects related to quality of life must ensure, first of all, that the domains under study are conceptualized in the same fashion in both cultures. Likewise, when assessing family-associated characteristics, test adapters must carry out careful research in order to find out whether there is a correspondence between the source and the target cultures as regards the nature and range of familial relations (Herdman, Fox-Rushby *et al.* 1998: 324); if they were not sufficiently similar, the comparison of the results obtained in both versions of the test would be of little value.

Marsella, Dubanosky *et al.* also consider construct equivalence, which they also term "conceptual equivalence," as being "more basic than any of the other equivalencies" (Marsella, Dubanosky *et al.* 2000: 53). From the perspective of personality measurements, these authors refer to a frame of test validation which takes into account four types of equivalence: on the one hand, linguistic equivalence, and, on the other hand, three types of psychometric equivalence, of which conceptual correspondence would be the first. With respect to psychometric properties, besides conceptual equivalence, Marsella, Dubanosky *et al.* (2000) call attention to scale equivalence (i.e., the degree of cultural acceptability in the way in which the instrument is scaled), and to normative equivalence (i.e., the idea that there must be reference data, the so-called *norms*, for the group to be studied, since norms based on a particular source culture group may not be valid for the target population). These types of psychometric equivalence are necessary to ensure the cultural validity of a test, given that, "by itself, the simple translation of materials from one language into another language is no guarantee that the instrument is valid or appropriate for use

in another culture" (Marsella, Dubanosky *et al.* 2000: 53). In the Standards for Educational and Psychological Testing (AERA, APA and NCME 2008), the concept of psychometric equivalence is also associated with the verbal performance of test participants. According to their authors, it should be recognized that "the values associated with the nature and degree of verbal output also may differ across cultures" (AERA, APA and NCME 2008: 97). That is to say, distinct cultural groups may judge certain speech patterns (such as short responses or verbosity) differently, and this diversity should be borne in mind when adapting a test and considering its results.

As Marsella, Dubanosky *et al.* (2000) recognize, as far as cross-cultural research is concerned, linguistic equivalence alone does not suffice. The failure to achieve it, however, will certainly constitute a fatal blow to the validity of the adapted test. Herdman, Fox-Rushby *et al.* allude to "semantic equivalence" when dealing with the correspondence of words in different languages; according to them, "semantic equivalence is concerned with the transfer of meaning across languages, and with achieving a similar effect on respondents in different languages" (Herdman, Fox-Rushby *et al.* 1998: 326). This includes issues such as register and dialect (that is to say, target readers must feel at home with the language used in the test, which, at the same time, is expected to feature the same level of difficulty as the version in the original language). Curiously, Herdman, Fox-Rushby *et al.* (1998) resort to Barnwell (1980), an expert in the field of linguistics, to show the several types of meaning which a translator should take into consideration when translating a test (i.e., referential, connotative, stylistic (social), affective, reflected, collocative and thematic). This level of accuracy at describing linguistic equivalence clashes with the usually vague statements made by cross-cultural psychologists when addressing the problem of translation;[4] they commonly refer to the "preservation of the original meaning" or to "equivalent words and phrases," with no further explanation of what they particularly mean by that. In section 3.4, we will discuss the nature and extent of linguistic equivalence, as well as some misconceptions about the process of translating in the context of psychological testing.

As a recapitulation, Herdman, Fox-Rushby *et al.* refer to a final type of equivalence, "functional equivalence," which "is intended to highlight the fact that all parts of the process outlined here are important in achieving cross-culturally equivalent questionnaires" (Herdman, Fox-Rushby *et al.* 1998: 331). For them, "functional equivalence" would be measured by how well an instrument performs as it is intended in two or more cultures.

As a necessary step in order to assess whether the different types of equivalence have been achieved in a particular instrument, psychologists avail themselves of a set of checking procedures. These may be judgmental (i.e., based on expert opinion) or statistical (i.e., based on the actual item responses of test takers). Both will be reviewed below, but a greater emphasis will be placed on the former, since these involve translation techniques or approaches which are characteristic of this field of knowledge.

3.3. Judgmental and statistical designs for adapting tests

The International Test Commission Guidelines on Adapting Tests (ITC 2000), under the section "Test Development and Adaptation" and as a reflection of the growing concern about the need for multiple evidence in order to establish the validity of an

adapted test, require test developers and publishers to compile both judgmental and statistical evidence in order to support the full efficacy of a given adapted test for a particular population. On the one hand, they state that the collection of linguistic and psychological judgmental evidence will make the adaptation process a more accurate operation. On the other hand, the use of appropriate statistical techniques is also expected in an adaptation procedure with regard to all intended populations, since they help identify potential difficulties and problematic components of the adapted test. In theory, if both types of evidence are adequately collected, all layers of equivalence, as described earlier in this paper, should be secure.

3.3.1. Judgmental designs

Judgmental designs involve complex translation checking techniques, and they are of particular interest as far as this paper is concerned. The two most popular judgmental methods are forward translation and backward translation (back translation). Slight variations of these basic methods may also be found in the literature about test adaptation within the field of psychology.

On the one hand, if a forward translation design is followed, a single translator (or, ideally, a number of them) adapts the test from the source language to the target language. At a second stage, a different group of translators decides on the equivalence of the two versions of the test (Hambleton 2005). Then, two more types of participant may take part as well:

a) a target-language speaker (who may or may not be a translator), in order to smooth out any discrepancies in the language used;
b) a group of target-language examinees, who will give their interpretation of the adapted items (this would involve a "think-aloud" study).

As Hambleton recognizes from the perspective of psychology, the main flaw of forward translation "is associated with the high level of inference that must be made by the translators about the equivalence of the two versions of the test" (Hambleton 2005: 12). In a variation of this design, called "multiple-forward translation," several independent translators are asked to translate the instrument, and then all the translations are compared item by item in order to detect problematic areas.

On the other hand, backward translation is, by far, the judgmental method that is most commonly chosen. As defined by Maxwell, it "is a three-step procedure": firstly, the original version of the test is translated into the target language; secondly, a different translator translates that version back into the source language; finally, the original and back-translated versions are compared by both psychologists and translators in order to consider possible deviations, and correct them (Maxwell 1996: 6). It is generally agreed that, as long as the two versions of the test in the source language look similar, there are enough reasons to argue that the source and target versions of the test are equivalent. Strangely enough, the original test acts as an unusual *tertium comparationis* against which the back-translated version of the test is checked. If both of them employ approximately the same linguistic materials in the source language (a criterion which is evocative of the formal and semantic types of equivalence suggested by Nida [1964] and Newmark [1981], respectively), the target language version of the test, which has not been the focus of the evaluation, is given the green light by the judges.

Even though this is the procedure which is most frequently favoured by psychologists when adapting tests (since it is they who are mainly in charge of assessing the validity of the test by comparing both versions in the source language), some disadvantages have been identified from the point of view of both psychology and linguistics. Most importantly, a proficient back translation may allow defects in the target version of the test to remain hidden. For instance, an adaptation could inappropriately retain linguistic features (e.g., grammar, spelling) from the source language which are alien to the language of the new target population, and it would still pass the back translation check; in other words, in those cases in which the target version is simply a word-by-word rendering of the original version, back translators would have a very easy job transferring the words in it back into the source language, but the adapted version would probably be flawed as regards language fluency and conceptual equivalence. For instance, in a questionnaire measuring the self-confidence of referees (Guillén, Feltz *et al.* 2010), a back translation process would encourage a very literal rendering of the items, as the following example from a hypothetical adaptation into Spanish illustrates:

(1) (You have the ability to) consistently be successful in making correct decisions.

(Tienes la habilidad de) consistentemente ser exitoso en tomar decisiones correctas.

(Literal translation by the author)

When conveying the Spanish version back into English, the back translator is likely to reconstruct the source language item to almost its original form, thus meeting the equivalence requirements of psychologists. However, the literal rendering stands as a syntactic and lexical calque where the Spanish language is used in a very unnatural way. If, on the contrary, the item were translated following more idiomatic criteria, the resulting sentence (e.g., <u>(Eres capaz) de tomar siempre las decisiones correctas</u>), though being much more adequate as a Spanish language utterance, would lead to a back translation which would differ from the original text form-wise. Another example can be found in a test about the confidence of a team regarding an upcoming game or competition (Short, Sullivan *et al.* 2005). Here is the source text item and a literal rendering of it into Spanish, together with a more idiomatic translation:

(2) (Your team has the ability to) stay in the game when it seems like your team isn't getting any breaks.

(a) (Tu equipo tiene la habilidad de) permanecer en el juego cuando parece que no está teniendo ningún descanso.

(Literal translation by the author)

(b) (Tu equipo es capaz de) continuar luchando cuando parece que todo le sale mal.

(Non-literal translation by the author)

As in the previous case, the literal translation would be, in theory, favoured by psychologists since, in the subsequent back translation, it would probably match the form of the source text better than the non-literal version. Yet, in terms of target language usage, the latter seems an item that is more likely to be adequately understood and answered by the Spanish language takers of the test.

Partly for these reasons, back translation is not recommended as a stand-alone procedure, not even by psychologists, since "it may provide an artificial similarity of meaning across languages but not the best version in the new language" (AERA, APA and NCME 2008: 92). In general terms, and taking into account some concerns arising out of translation studies, experts who opt for this procedure tend to lean towards out-of-context literalness, thus producing ill-adapted tests which may make little sense to the subjects who will take them. Furthermore, this simplistic approach delays, by concealing them, the problem-solving stage of, for instance, cultural differences, which will be figured out only in later steps of the process of adaptation, if at all (for example, at the statistical stage of the validation).

Since adapted tests need to be field-tested before carrying out cross-cultural research, judgmental methods are only part of the procedures which have to be put into practice in order to validate the translated instruments. In this sense, most judgmental designs are criticized for not submitting their final adapted versions to samples of the intended population, and for not carrying out the test-taking procedure under actual conditions. With the aim of overcoming this limitation, psychologists make use of statistical and other methods, such as the data collection designs and data analysis.

3.3.2. Statistical designs

These methods are based on the actual item responses of examinees, and provide empirical data which should help developers to verify the equivalence of the source and target language versions of a test. They are regarded as a necessary safety check which complements the judgmental methods which have been described in the previous section. As Hambleton points out, the following are the most commonly implemented statistical or data collection designs:

a) bilingual examinees take both the source and target language versions of the test;
b) monolingual speakers of the source language take the original and back-translated versions of the test;
c) source language monolinguals take the source language version of the test, and target language monolinguals take its target language version.

(Hambleton 1993; 2005)

Experts identify several shortcomings in all of them. For example, in the first one, it is wrongly presumed that bilinguals have the same proficiency in both languages, or that they will answer the test in the same manner as an average monolingual would. With regard to the second design, the main downside is that feedback is only obtained from the source language versions of the test (the original instrument and its back-translated rendering), with no actual data from the target language adaptation. Finally, the third data collection method, despite not showing as many flaws as the other designs, may also be found to be compromised since it takes for granted that the members of the two groups will show identical level of ability. Given these drawbacks, test developers are encouraged to implement more than one statistical procedure in order to confirm the equivalence of the adapted version. Furthermore, the subsequent processing of the data collected through the statistical methods should help recognize these shortcomings and counterbalance them, as some particular analysis and frameworks do (e.g., item response theory framework).

In this context, combining judgmental and statistical specifications, Vallerand's methodology for cross-cultural validation (Vallerand 1989; Vallerand, Blais *et al.* 1989), specially conceived for adapting English psychological scales into Canadian French, has proved particularly influential and consists of seven steps: first, the preparation of a preliminary target language version, preferably involving translation and back translation; the assessment and modification of that preliminary version by a committee of experts is followed by the evaluation of the experimental French version in a pretest, using either "random-probe" or "test-retest" techniques, and/or with a second committee rating the level of ambiguity of each item and providing commentaries and suggestions; at this point, the evaluation of content and concomitant validity is to be performed, by calculating correlations or independent T-tests. Reliability is then gauged by calculating internal consistency and temporal stability indexes. The next step involves studying construct validity, which is evaluated through exploratory or confirmatory factor analysis. Finally, norms for the target language version are prepared so that an estimate of the position of the tested individual in a particular population with respect to the trait being measured can be yielded. Haccoun (1987) also suggested an interesting approach in order to evaluate content and concomitant validity and test-retest reliability in a single step. By administering both the original test and its translation to a group of bilingual individuals twice, at two different moments in time, it is possible to evaluate the relation between the original and the translated instruments and to obtain test-retest correlation coefficients in a simple and methodical manner.

When translated tests are submitted to traditional procedures of analysis based on classical test theory, many methodological shortcomings have been described. For Hambleton, Swaminathan *et al.*, these include:

a) use of item indices whose value depends on the particular group of examinees with which they are obtained;
b) examinee ability estimates that depend on the particular choice of items.

<div align="right">(Hambleton, Swaminathan et al. 1991: ix)</div>

In addition, these procedures involve assumptions that are somewhat stringent and unrealistic (Barbero García, Vila Abad *et al.* 2006). Item Response Theory (IRT) approaches, which entered the scene as a counter reaction to traditional methods, deserve a further explanation, since they have been a particularly enriching area of research which has yielded "numerous models, powerful estimation methods and creative applications […] [providing] compelling and rigorous answers to many measurement problems" (Drasgow and Hulin 1990: 631). Based on the idea that the probability of a correct response to an item is a mathematical function of person and item parameters, IRT relies on the application of related mathematical models to testing data. Whereas the unit of analysis for classical test theory was the test itself as a whole, IRT centers on the individual item, so that researchers are able to address problems beyond the scope of classical test theory easily, such as identifying subjects with inappropriate response patterns or selecting items at appropriate difficulty levels for respondents (Drasgow and Hulin 1990). Thus, psychometric equivalence between a source and a target language test is gauged by equivalence of response probabilities to source and target language items (Hulin 1987), and only if all items in the scale are equivalent is the whole test to be considered as so (Candell and Hulin 1986).

3.4. A critical perspective from translation studies

The range of procedures for use by psychologists and test developers when adapting instruments to different languages manifestly shows how concerned these professionals are about the need to achieve equivalence across cultures. These procedures, in particular the judgmental designs, however, rely on an intuitive approach which seems to confuse and oversimplify a series of concepts which have been pivotal in the development of translation studies, such as culture, translation techniques, and the notions of equivalence and translation themselves. Specifically, experts in the field of multicultural psychological testing often underestimate the actual extent of the process of translation, which is regarded by many of them as a one-way word-by-word substitution procedure which needs to resort to supposedly *non-translational* strategies in order to tackle cultural differences. The "item substitution method" and "decentering" are two such strategies, which Hambleton suggests as procedures to overcome the difficulty of "finding equivalent words or phrases" (Hambleton 1993: 60). The former technique is described as the replacement of "an item which may not translate well […] by a comparable item" (e.g., a reference to a king in a questionnaire should be substituted if the target system is a republic), while the latter refers to "the modifying of words or phrases" at both the development and adaptation stages of the test in order "to alleviate the problem of non-equivalent words or phrases in the source and target languages" (Hambleton 1993: 60). From the perspective of translation studies, phrases such as "which may not translate well" and "the problem of non-equivalent words" take us back to the tentative statements which were made about translation before the first thorough and systematic approaches to the subject were attempted. In this sense, as far as theory is concerned, translation studies have accounted for cultural and linguistic imbalance in a methodical manner since the middle of the 20th century, when, to name but an early example, Vinay and Dalbernet (1958) described their various direct and oblique translation techniques. Literature on test adaptation leaves the general impression that psychologists are somewhat sceptical about translators and their work, but, as we suggested above, we believe that this reservation arises out of a lack of knowledge about the actual workings and potential of translation. For instance, some experts (e.g., Hulin and Mayer 1986) state that a fluent and smooth target version of an item is irreconcilable with an adequate rendering of the different layers of meaning to be found in it (i.e., naturalness vs. equivalence). As a consequence of this idea, a literal or even word-by-word approach is expected of translators, which should not depart from the "plain" meaning of the terms in a source language questionnaire. By following this line of reasoning, we find that another misconception lies in the way in which many developers conceptualize language, which is apparently considered to be a system of words which function irrespectively of any context or cultural environment. This would explain why they think that they have to turn to *special* measures (e.g., the "item substitution method" and "decentering" strategies) when one of their instruments features some non-equivalent cultural element. From this standpoint, translators are likely to be dismissed as accessory participants in the process of adaptation, merely in charge of a superficial transfer of words between languages.

This narrow approach, however, is not shared by all psychologists who wish to adapt an existing test for a foreign population. Some of them acknowledge the expertise of

translators and allow them to play a more significant role in the process of test adaptation. Braun and Harkness, for instance, recognize that

> [the training and skills of translators] help them identify potential ambiguities and translation problems. This is one reason why translators can be very useful proofreaders for draft questionnaires and can be helpful in developing questionnaires intended for comparative use. (Braun and Harkness 1997: 103)

By being more open-minded towards the advantage they can take from the solutions suggested by translators, cross-national researchers are more likely to succeed at validating tests, partly because they deal with many potential problems of adaptation at the translation stage. Furthermore, translators can avail themselves of a full set of techniques and concepts which are founded on contemporary ideas about translation and equivalence drawn from translation studies. By combining the translators' comprehensive approach to adaptation with both the vital judgment of psychologists and their validating statistical techniques, the needs of the target population of a given test would probably be better fulfilled. In particular, among the modern approaches to translation, the functionalist theories are especially well equipped to provide a thorough understanding of the process of test adaptation.

For example, the standpoint of Holz-Mänttäri (1984), who views interlingual translation as an action involving several communicative participants, comfortably fits the circumstances of adapting a test. On the one hand, translation is not seen as a mere act of word juggling, but as an activity which, departing from a source text, is heavily influenced by sociocultural and communicative factors. On the other hand, the "players" that take part in a translational action are believed to determine the outcome of the process, so that, for a given source text, the combination of different players will lead to different (adequate) results. Applying the roles defined by Holz-Mänttäri (1984) to a hypothetical process of test adaptation, the "initiator" would be the institution (e.g., a university) or company that needs the translation; the "commissioner" would be the representative (e.g., the head of a research project, the chief officer of a company's customer relations division) who gets in touch with the translator; the "source text producer" would refer to the author of the original test, who may or may not take part in the adaptation; the "target text producer" would comprise both the translator and the psychologists, working as a team; the "target text user" would be those in charge of administering the adapted test, in many cases the same as the initiator and the commissioner; and, finally, the "target text receiver" would refer to the population that takes the adapted test.

Let us look at two situations in which a test measuring the academic motivation of students is likely to be adapted. In the first of them, a government-funded research group intends to analyze the reasons behind high school dropouts, so they adapt this test in close collaboration with translation scholars acquainted with the field of education, and they address the target text to a population of high school students selected from all socioeconomical backgrounds in the country. In the second situation, an outsourcing company wishes to survey the motivation of top university students as a way to scrutinize the suitability of current executive posts for future elite graduates; for the process of adaptation, they commission the translation to a professional translation services agency, and they administer the resulting test to a group of senior students from the top three universities in the country. Even though

sharing the same source text producer and the same original test, the differing circumstances of the participants in the two situations will mean that the two adapted tests will be unlike. In this context, it does not seem very convenient to follow a back translation procedure which is based on the alleged mirror-like quality of the working languages and on the absence of any extralinguistic influence. Reiss and Vermeer (1984) acknowledge the multiplicity of factors affecting the form of a target text and the way it is received and, when summarizing their translation theory, remark that a *Translat* (i.e., a target text) reproduces an offer of information which is not clearly reversible. Nord, by following a similar line of reasoning, also concludes that "the translation process is irreversible." (Nord 1997: 32)

A proficient translator is well aware of the impact which varying players have on the form and the content of a target text. Furthermore, they are capable of coping with these dissimilar circumstances, and of devising their strategies and methods by focusing on the purpose (*skopos*) of the translation. In this sense, following the seminal work of Reiss and Vermeer (1984), functionalist translators have replaced the concept of "equivalence" with that of "adequacy," so they consider that a target text is functionally and communicatively adequate if it conforms to the *skopos* of the translation assignment. In the context of test adaptation, arriving at a properly adapted test is the same as producing a target text which fulfills the *skopos* outlined by a commission (i.e., which is adequate in the new communicative situation). This means that translators, when adapting tests, can attend to the source linguistic material from the perspective of the needs of the commissioner, rather than –as is often demanded of them– in a vacuum. If we applied these considerations to the adaptation of tests, a better defined working environment would arise: on the one hand, translators would be in charge of supplying a communicatively adequate rendering of the source language test; and, on the other hand, psychologists would devote their time to taking care of the psychometric properties of the adapted test by briefing translators and by applying the subsequent statistical techniques.

In order to produce an adequate target language test, translators can look into the "intratextual factors" which are part of Nord's analysis strategy (1991: 79-130). By addressing them, translators can identify potential problems from the perspective of the particular translation commission.[5] Among them, the factors which are most likely to pose a threat to adequacy in the process of test adaptation are the presuppositions and the lexis. All of these problems may be tackled by the translators themselves, as is the case in many other specialized fields.

In the case of presuppositions, these "comprise all the information that the sender expects (i.e., presupposes) to be part of the recipient's 'horizon'" (Nord 1991: 96). In this sense, they are directly related to the primal "conceptual equivalence" that psychologists pursue when adapting a test. In questionnaires where all or some of the items are bound to the source culture, and are not relevant in the target culture, a cultural shift should be implemented (in close association with psychologists), or, in some instances, it could be even recommended not to adapt the test. For example, in a questionnaire about physical self-description, items measuring negative self-perception in the form of "I am fat" or "My thighs are too big" would be interpreted differently in countries where bodily fat is applauded by society. Likewise, an item which illustrates the range of usual diseases in a given source culture (e.g., flu, virus, cold) would not be pertinent in a community where malaria or AIDS are the main

reasons to visit a hospital. In both cases, translators would be able to suggest how to reconfigure the items in the process of adaptation, leaving psychologists in charge of examining the psychometric qualities of the proposed solutions.

In some circumstances, the cultural imbalance is aggravated by the intervention of different forms of "patronage" (Lefevere 1985). For instance, if a test gauging academic motivation were to be adapted for a student population in a foreign-language communist regime, target users and receivers would not find it relevant to include items which measure capitalism-related incentives (e.g., earning more money or prestige, leading a "good life" in a material sense). This would also happen where motivation to put effort into one's job is analyzed. In order to adjust to the patronage to be exercised by the communist apparatus with regard to the contents of the adapted test, as well as to the expectations of the target test takers, translators could propose a reconfiguration of the inadequate items. Having been advised about the problems and the suggested solutions by the translator, the initiator or commissioner could decide to delete the affected items, or even dismiss the whole task of adaptation.

Regarding lexical factors, translators are particularly aware of the choice of register, and the use of slang or group-related formulas associated with the particular test takers (e.g., professional athletes, university students). The question as to the words and the style which translators will use is crucial, since the target population must feel at home with the language they read if an adapted test is to be successful. For any given test, several sets of target receivers and users can be distinguished. This is the case even in very specific tests which are addressed to a very narrow population. Let us take a questionnaire which assesses the self-confidence of referees as an example. To start with, a difference could be made between amateur and professional referees, a distinction which is closely related to the age range of the users (referees of games involving professional athletes are usually older than those refereeing games in under-18 competitions). This difference in the age range should result in a different register being used in the adapted test; it could even produce the alteration of the content of some items. Thus, a test addressed to young amateur referees should feature a simpler and more informal style and vocabulary, whereas one for older professional referees should include more formal statements and words.

Psycholinguistic and cognitive approaches to translation have similarly highlighted the many qualities that the translator needs, by identifying the subprocesses involved in translation (see for example De Groot 1997: 28). Empirical methodology borrowed from psychology to research into the translator's black box, like Think-Aloud Protocols[6] (TAPs) or, more recently, keystroke logging (Jakobsen 2006) among others, has proved to be very helpful in the study of process-oriented phenomena. For instance, the study of differences between experts and novice translators has been particularly productive. It seems that professional translators process larger translation units, are mainly "sense-oriented" instead of "form-oriented," take into account stylistic and text-type adequacy and "have a larger number of variants at their disposal" (Kussmaul and Tirkkonen-Condit 1995: 187). All these qualities seem to favour the role of professional translators versus amateurs or individuals versed in languages and who have a psychological background, as ideal linguistic counsellors in test adaptation projects.

It has also been acknowledged that the comprehension of the particular traits of successful translation performance and professionals is "valuable […] in any efforts

aimed at reforming existing professional practices" (Kussmaul and Tirkkonen-Condit 1995: 189) in any field of translation, and psychological test adaptation is no exception. Thus, findings suggest that successful translators seem to subordinate local decisions to global ones, they do not always aim at an optimal result but at a text product which is adequate and sufficient for a particular communicative situation; and are ready to use their world knowledge and inferences about the text in general, and text type in particular, in order to make decisions. Furthermore, they apparently have "relative articulate subjective theories of translation" and "they focus their attention, their conscious decision-making and their use of translation aids so that their investment in effort results in sufficient communicational gains" (Kussmaul and Tirkkonen-Condit 1995: 190). Finally, some personal characteristics (Jääskeläinen 2000) such as flexibility, realism, tolerance of ambiguity (Tirkkonen-Condit 2000) and intellectual curiosity have also been associated with successful translators. All these data can be of help when it comes to defining the ideal test translator profile, to revising a translation test or to conducting an interview prior to translator recruitment.

Another critical issue within this framework of study of translation is the relationship between automatic and effortful processing and the way it changes between novices and experts over time (Shreve 1997; Shreve and Diamond 1997). According to some studies, it seems that as the level of professionalism grows, the translator's conscious decision-making alters since "while some decisions become non-conscious, or 'automatic,' the translator becomes sensitized to new aspects of the task which require conscious decision-making" (Jääskeläinen and Tirkkonen-Condit 1991: 106). In this way, more controversial aspects of a particular commission are identified and cost-effectiveness is improved since time and efforts are devoted to the resolution of key problems.

On the other hand, every important cognitive paradigm has in some way influenced translation research. The early cognitive symbol manipulation approach assumed that the brain controls intelligent action and that it stores information by using mental representations, with the functioning of a human being's mind being compared to the information processing of computers which manipulates symbols according to fixed rules. Under the influence of this approach, translator researchers took to comparing syntactic and semantic structures of different linguistic systems. In this framework, techniques such as back translation might have been considered as suitable. Next came Connectionism, in which the mind became a dynamic, holistic network (Risku 2010: 96) and meaning was determined by "patterns that emerge in an unpredictable way from the parallel activation of neural connections" (Martín de León 2012); an individual's experience was recognized as playing an important role in language use and understanding, which, in the discipline of translation studies, entailed a growing interest in the cultural and contextual aspects of translation (Martín de León 2012).

During the last 20 years a new translation paradigm, which found its "milestones" in the works of Hönig, Kussmaul, Reiss and Vermeer, and Holz-Mänttäri (Risku 2002) has developed, based on situated and embodied cognition. Instead of restricting the object of research to internal representations, it "emphasises the role played by physical and social context in cognition" (Risku 2002: 523), and studies the interaction with artefacts such as language and the social environment. As Risku clearly puts it:

The question "What happens in the translator's brain?" should be supplemented by others, such as "What happens in the hands, in the computers, on the desks, in the languages, in the dialogues of translators?" Translation is done not only by the brain, but also by complex systems, systems which include people, their specific social and physical environments and all their cultural artefacts. (Risku 2002: 530)

Prototypes, cultural norms and conventions are seen as merely initial hypotheses which must be adapted to an anticipated situation (Risku 2010). The conception that the translators' task is to provide text recipients with the required tools to construct their own meanings in their own situation (Risku 2004; Martín de León 2012) seems particularly relevant to test adaptation: figuring out the anticipated situation of recipients' test taking within their particular environment could lead to more appropriate decision-making.

As a whole, during the last 30 years, researchers have managed to sustain the fact that psycholinguistic and cognitive approaches are "both appropriate and fascinating, and that they may have an enormous impact on translation and interpreting quality" (Muñoz-Martín 2012). Coupling both representational and situated frameworks of study could be the only way to completely understand translation in its complex professional environment (Martín de León 2012).

In summary, while acknowledging that many test adapters are still far from understanding and taking advantage of the full range of techniques which translators may offer them (or, in some cases, are unaware that they are taking advantage of them), it seems evident that a close collaboration between translation and test experts is needed in order to accomplish a successful adaptation (e.g., Braun and Harkness 1997). The psychometric properties of a test, in particular, can only be adequately adapted to the target population as a result of the strict monitoring and manipulation of translations by psychologists. In connection with this idea, it is commonly accepted by test adapters (e.g., Moreno Rosset 2005) that, while doing a translation may take a relatively short time, the whole process of adaptation and validation of a test could imply years of work. Perhaps not paradoxically, the simultaneous drafting of the test in the several languages of the populations to be surveyed (i.e., the interaction of experts from the conception of the instrument) would be the most efficacious, if very often impractical, method of adaptation. In such a situation, translators would play a significant role as far as communication between the experts is concerned, but all versions of the test would be developed as source language instruments which would fully take into account the cultural constraints and peculiarities of the intended population.

4. Final considerations: requirements for professionals and test adaptation as test localization

As in many other interdisciplinary fields which must ideally benefit from team work, all test translation professionals, regardless of their background, need to have a thorough understanding of each other's tasks. While translators require some training in test and scale construction, psychologists need to assimilate the scope and main concepts underlying translation. However, despite their importance for the final quality of the adapted instrument, the requirements for professional translators to be selected in a test adaptation project are not explicit in the International Test

Commission Guidelines on Adapting Tests (ITC 2000). From the standpoint of psychologists, Hambleton warns about the risks of engaging translators just because they "happened to be available – a friend, a wife of a colleague, someone who could be hired cheaply, and so on" (Hambleton 2005: 10), which has been a relatively common practice in the past, to the detriment of the quality of the final translated product. According to this author (1993; 2005), in order to provide different perspectives designed to solve the difficulties which arise during the process, test translation must be accomplished by more than one translator with a deep knowledge of the cultures involved, especially as far as the target culture is concerned. Likewise, they must also be familiar with the subject matter and they should have some training in test and scale construction, so that they are able to avoid mistakes[7] which could affect the validity of the translated instrument. Additionally, special attention is needed to guarantee the correct functioning of the translation team during the whole process (Muñiz and Hambleton 1996). Furthermore, from the point of view of translators, it would also be advisable to require other characteristics besides the mastery of the languages and cultures involved in any given commission and of the other fields of competence which have been already suggested. Along these lines, Computer-Assisted Translation software skills, ability for teamwork, linguistic sensitivity, previous knowledge or experience in project management, knowledge and experience in terminology management and predisposition for long-life learning are to be considered.

Although steps have been taken in the last few years in order to improve cross-national psychological testing, there is still room to reflect on several aspects in order to refine the process and improve the quality of the final product. As translators know, translation is not just a transcoding process, but a form of human action which has been addressed from varied theoretical angles. All these academic trends have their counterpart in a translation practice which could be enriching for psychologists and translators involved in psychometric test translation. For instance, as we have seen, functionalist authors have argued that a translated text is not exclusively determined by the source text and that its own purpose or *skopos* must be borne in mind. In contrast with theories which focus on prescriptions limited by the source text, translations could thus be described in terms of original text production instead of in the more traditional terms of equivalence with another text in another language (Schäffner 1998). Thus, the notion of equivalence, while all-important in earlier works about test translation, is controversial in translation studies, and has even been rejected by some authors, while others try to categorize it as involving denotative, connotative, text-normative, pragmatic, formal, textual and functional equivalence. As we have illustrated, from another perspective, cognitive approaches could shed light on the features which distinguish expert and successful professional translators from novices, the balance of automatic versus conscious decision-making, and the key role of the specific social and physical environments and their cultural artefacts, to name but a few.

Nowadays, both the increasing widespread need for psychometric tests translation and the rapid growth in computer-based testing instruments, seem to demand a wider framework in which to transcend earlier limitations, and this framework would be the broader process of localization, a concept which arose in the 1980s with deep roots in the computer software sector. According to the Localization Industry

Standards Association, a not-for-profit organization formed in 1990, this process implies "modifying products or services to account for differences in distinct markets" (LISA 2003: 13), and addresses "significant, non-textual components of products or services in addition to strict translation" (LISA 2007: 19). Involving as it does not only linguistic transfer but also content, cultural, regulatory, ethical and technical issues, localization accounts for changes in information, functionality, software codes and even product design in an organized manner, thus facilitating quality assurance and control. If translation is taken into consideration from the early stages of test development, it is more likely that all resources will be more efficiently assigned, delays will be more easily maintained and the result will be more culturally appropriate.

Furthermore, for the localization industry, quality assurance constitutes a priority (LISA 2004), and psychological test translation could benefit from its vast experience in this area. The Localization Industry Standards Association has devised a standardized and exhaustive quality assurance model for product localization that covers all aspects in the process (from language issues to documentation) and which has been implemented as a software application with a stand-alone interface (LISA 2007: 57). For instance, it provides an extensive list of localization error categories with examples, and objective measures of error severity. Although most suitable for the translation of computer-based testing instruments, an adapted form of this systematic approach could also be of much use for conventional test translation. In line with this suggestion, the test translation guidelines drafted by the US Census Bureau (Pan and De la Puente 2005) and the European Social Survey (ESS 2010), together with the test adaptation procedures implemented by the Survey of Health, Ageing and Retirement in Europe (SHARE) already provide an exhaustive range of tools to assess the routines of translators and the quality of their work. In particular, the latter has availed itself of the expertise of translation researchers Hans Hönig and Paul Kussmaul, and advocates a team translation model, TRAPD, which comprehensively embraces the tasks of translation, review, adjudication, pretesting and documentation (Harkness 2005). The efforts by the ESS and SHARE, as large cross-national survey development projects, and the US Census Bureau, in their intent to account equally for the realities of the multinational population of the US, are examples of good practice in test adaptation, where some of the principles of localization are efficiently followed. Likewise, the work of some cross-national researchers (e.g. Harkness and Schoua-Glusberg 1998; Harkness, Van de Vijver *et al.* 2003; Harkness 2007; Harkness, Villar *et al.* 2010) signals a change of direction in psychological test adaptation, where systematic translation procedures and assessment are regarded as crucial contributors to the quality of the final product, that is, the data which is eventually collected from the administration of the adapted tests.

NOTES

1. This is one of the most frequently used and translated tests worldwide (Hambleton 1993), and it is the most common psychological test in Spanish- and Portuguese-speaking countries (Prieto, Muñiz *et al.* 1999).
2. THE PSYCHOLOGICAL CORPORATION (1997): *Weschler Adult Intelligence Scale (WAIS-III) Administration and Scoring Manual.* San Antonio: The Psychological Corporation.
3. These associations were the International Association for the Evaluation of Educational Achievement, the European Association of Psychological Assessment, the International Association of

Applied Psychology, the International Association of Cross-Cultural Psychology, and the International Union of Psychological Science.

4. Likewise, these authors show a sensibility towards translation which is unheard of in other works about test adaptation. For example, they claim that "much of the translator's task cannot be reduced to a mere technical exercise and there is almost always an element of 'art' involved. The ability to express the original message as accurately, clearly and naturally as possible, the ability to find the right tone and register (level of language) and an awareness of the impression that his or her translation will make on the reader in the target language are all hallmarks of a good translator" (Herdman, Fox-Rushby *et al.* 1998: 328).

5. A "translation problem," in the sense given by Nord (1997: 141), is one which has "to be solved by the translator in the translation process in order to produce a functionally adequate target text and which can be verified objectively or at least intersubjectively."

6. Despite the recent controversy about the validity of TAPs, it has been recognized that although the information gathered "may well be edited, partial, communication-oriented constructions by subjects […] the fact that they do not yield reports on actual processes does not mean that they do not offer valuable insights on them" (Muñoz-Martín 2012).

7. See Solano-Flores, Contreras-Niño *et al.* (2006: 6) for a description of the ten dimensions of translation errors in test adaptation (ranging from style and format to conventions and register).

REFERENCES

Aera, apa and ncme (2008): *Standards for Educational and Psychological Testing*. Washington: American Educational Research Association, American Psychological Association and National Council on Measurement in Education, and American Educational Research Association.

Anastasi, Anne (1988): *Psychological Testing*. New York: Macmillan.

Barbero garcía, María Isabel, Vila Abad, Enrique and Suárez Falcón, Juan Carlos (2006): *Psicometría*. Madrid: Universidad Nacional de Educación a Distancia.

Barnwell, Katharine (1980): *Introduction to Semantics and Translation*. City: High Wycombe Summer Institute of Linguistics.

Binet, Alfred and Simon, Théodore (1954): *La mesure du développement de l'intelligence chez les jeunes enfants.* Publications de la Société Alfred Binet. Paris: Bourrelier.

Bolaños Medina, Alicia (2012): Las pruebas psicométricas en la investigación empírica sobre los procesos cognitivos del traductor. *In*: Susana Cruces Colado, Ana Luna Alonso, Maribel del Pozo Triviño and Alberto Álvarez Lugrís, eds.: *Traducir en la Frontera*. Actas del IV Congreso de AIETI. Granada: Atrio, 469-482.

Braun, Michael and Harkness, Janet A. (1997): Text and Context. Changes to Comparability in Survey Questions. *In*: Jürgen H. P. Hoffmeyer-Zlotnik and Janet A. Harkness, eds. *Methodological Aspects in Cross-Cultural Research*. Mannheim: ZUMA, 95-108.

Brislin, Richard W. (1970): Back-Translation for Cross-Cultural Research. *Cross-Cultural Psychology*. 1:185-216.

Bulmer, Michael (2003): *Francis Galton: Pioneer of Heredity and Biometry*. Baltimore: The Johns Hopkins University Press.

Butcher, James N., Dahlstrom, W. Grant, Graham, John R., *et al.* (1989): *The Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

Candell, Gregory L. and Hulin, Charles L. (1986): Cross-Language and Cross-Cultural Comparisons in Scale Translations: Independent Sources of Information about Item Non-Equivalence. *Journal of Cross-Cultural Psychology*. 17(4):417-440.

Cronbach, Lee J. (1990): *Essentials of Psychological Testing*. New York: Harpercollins College Division.

De Groot, Annette M. B. (1997): The Cognitive Study of Translation and Interpretation. *In*: Joseph Danks, Gregory Shreve, Stephen Fountain, *et al.*, eds. *Cognitive processes in Translation and Interpreting*. London: Sage, 25-56.

DRASGOW, Fritz and HULIN, Charles L. (1990): Item Response Theory. *In*: Marvin D. DUNNETTE and Leaetta M. HOUGH, eds. *Handbook of industrial and organizational psychology.* Palo Alto: Consulting Psychologists Press, 577-636.

ESS (European Social Survey) (2010): *ESS Round 5 Translation Guidelines.* Mannheim: European Social Survey GESIS.

GAMERO PÉREZ, Silvia (2001): *La traducción de textos técnicos.* Barcelona: Ariel.

GREGORY, Robert J. (2010): *Psychological Testing: History, Principles and Applications.* New York: Prentice Hall.

GUILLÉN, Félix, FELTZ, Deborah, and DITHURBIDE, Lori (2010): The Refficacy Scale: A premilinary investigation to develop a referee efficacy scale. *2010 Conference Canadian Society for Psychomotor Learning and Sport Psychology*, Ottawa, Canada, 28-30 October. Unpublished communication.

HACCOUN, Robert R. (1987): Une nouvelle technique de vérification de l'équivalence de mesures psychologiques traduites. *Revue québécoise de psychologie.* 8(3):30-39.

HAMBLETON, Ronald K. (1993): Translating Achievement Tests for Use in Cross-National Studies. *European Journal of Psychological Assessment.* 9(1):57-68.

HAMBLETON, Ronald K. (2005): Issues, Designs and Technical Guidelines for Adapting Tests into Multiple Languages and Cultures. *In*: Ronald K. HAMBLETON, Peter. F. MERENDA and Charles D. SPIELBERGER, eds. *Adapting Educational and Psychological Tests for Cross-Cultural Assessment.* Mahwah/New Jersey/London: Lawrence Erlbaum Associates, 3-38.

HAMBLETON, Ronald K., SWAMINATHAN, Hariharan and ROGERS, Jane (1991): *Fundamentals of Item Response Theory.* Newbury Park: Sage.

HARKNESS, Janet A. (2005): Translation Procedures and Translation Assessment. *In*: Axel BÖRSCH-SUPAN and Hendrik JÜRGES, eds. *The Survey of Health, Aging, and Retirement in Europe – Methodology*, 24-27. Visited on 15 September 2010, <http://www.share-project.org>.

HARKNESS, Janet A. (2007): Improving the comparability of translations. *In*: Roger JOWELL, Carolyn ROBERTS, Rory FITZGERALD, *et al.*, eds. *Measuring Attitudes Cross-Nationally: Lessons from the European Social Survey.* Sage: Los Angeles, 79-93.

HARKNESS, Janet A. and SCHOUA-GLUSBERG, Alicia (1998): Questionnaires in Translation. *In*: Janet A. HARKNESS, ed. *ZUMA-Nachrichten Spezial No. 3. Cross-Cultural Survey Equivalence.* Mannheim: ZUMA, 87-127.

HARKNESS, Janet A., VAN DE VIJVER, Fons J. R. and MOHLER, Peter P., eds. (2003): *Cross-Cultural Survey Methods.* Hoboken, New Jersey: John Wiley & Sons.

HARKNESS, Janet A., VILLAR, Ana and EDWARDS, Brad (2010): Translation, Adaptation, and Design. *In*: Janet A. HARKNESS, Michael BRAUN, Brad EDWARDS, *et al.*, eds. *Survey Methods in Multinational, Multiregional, and Multicultural Contexts.* Hoboken: John Wiley & Sons, 117-140.

HATIM, Basil and MASON, Ian (1990): *Discourse and the Translator.* London: Longman.

HERDMAN, Michael, FOX-RUSHBY, Julia and BADIA, Xavier (1998): A model of equivalence in the cultural adaptation of HRQoL instruments: the universalist approach. *Quality of Life Research.* 7:323-335.

HIGGINS, Louise T. and ZHENG, Mo (2002): An Introduction to Chinese Psychology. Its Historical Roots Until the Present Day. *The Journal of Psychology.* 136(2):225-239.

HOLZ-MÄNTTÄRI, Justa (1984): *Translatorisches Handeln: Theorie und Methode.* Helsinki: Suomalainen Tiedeakatemia.

HULIN, Charles L. (1987): A Psychometric Theory of Evaluations of Item and Scale Translations: Fidelity Across Languages. *Journal of Cross-Cultural Psychology.* 18(2):115-142.

HULIN, Charles L. and MAYER, Laura J. (1986): Psychometric equivalence of a translation of the Job Descriptive Index into Hebrew. *Journal of Applied Psychology.* 71(1):83-94.

JÄÄSKELÄINEN, Riitta (2000): Focus on Methodology in Think-aloud Studies on Translating. *In*: Sonja TIRKKONEN-CONDIT and Riitta JÄÄSKELÄINEN, eds. *Tapping and Mapping the Processes of Translation and Interpreting.* Amsterdam/Philadelphia: John Benjamins, 71-81.

Jääskeläinen, Riitta and Tirkkonen-Condit, Sonja (1991): Automatised Processes in Professional vs. Non-Professional Translation: A Think-Aloud Protocol Study. *In*: Sonja Tirkkonen-Condit, ed. *Empirical Research in Translation and Intercultural Studies*. Tübingen: Narr, 89-109.

Jackson, Charles (1996): *Understanding Psychological Tests*. Leicester: British Psychological Society.

Jakobsen, Arnt L. (2006): Research Methods in Translation – Translog. *In*: Kirk Sullivan and Eva Lindgren, eds. *Computer Keystroke Logging and Writing*. Oxford/Amsterdam: Elsevier, 95-106.

Kussmaul, Paul and Tirkkonen-Condit, Sonja (1995): Think-Aloud Protocol Analysis in Translation Studies. *TTR*. 8(1):177-199.

Lefevere, André (1985): Why Waste Our Time on Rewrites? The Trouble with Interpretation and the Role of Rewriting in an Alternative Paradigm. *In:* Theo Hermans, ed. *The Manipulation of Literature: Studies in Literary Translation*. Kent: Croon Helm, 215-243.

LISA (LOCALISATION INDUSTRY STANDARDS ASSOCIATION) (2003): *The Localisation Industry Primer*. 2nd Edition. Visited on 30 September 2009, <http://www.lisa.org>.

LISA (LOCALISATION INDUSTRY STANDARDS ASSOCIATION) (2004): *LISA Best Practice Guide: Quality Assurance. The Client Perspective*. Visited on 16 October 2010, <http://www.lisa.org>.

LISA (LOCALISATION INDUSTRY STANDARDS ASSOCIATION) (2007): *The Globalization Industry Primer. An Introduction to Preparing Your Business and Products for Success in International Markets*. Visited on 30 September 2010, <http://www.lisa.org>.

Magno, Carlo (2009): Taxonomy of Aptitude Test Items: A Guide for Item Writers. *The International Journal of Educational and Psychological Assessment*. 2:39-53.

Marsella, Anthony J., Dubanoski, Joan, Hamada, Winter C., *et al.* (2000): The Measurement of Personality across Cultures: Historical, Conceptual and Methodological Issues and Considerations. *American Behavioral Scientists*. 44(1):41-62.

Martín de León, Celia (forthcoming): Who cares if the cat is on the mat? Contributions of cognitive models of meaning to translation. *In*: Ana Rojo and Iraide Ibarretxe-Antuñano, eds. *Cognitive Linguistics and Translation: Advances in Some Theoretical Models and Applications*. Berlin: Mouton de Gruyter.

Maxwell, Beverley (1996): Translation and Cultural Adaptation of the Survey Instruments (Chapter 8). *In*: Michael O. Martin and Dana L. Kelly, eds. *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development*. Chestnut Hill: Boston College, 1-10.

Moreno rosset, Carmen, ed. (2005): *Evaluación psicológica. Concepto, proceso y aplicación en las áreas de desarrollo y de la inteligencia*. Madrid: Sanz y Torres.

Muñiz, José and Fonseca-Pedrero, Eduardo (2009): Construcción de documentos de medida en Psicología. *Focad (Formación continuada a distancia del Consejo General de Colegios Oficiales de Psicólogos)*. 1:22 p. Visited on 15 September 2009, <http://www.cop.es/focad/>.

Muñiz, José and Hambleton, Ronald K. (1996): Directrices para la traducción y adaptación de los tests. *Papeles del Psicólogo*. 66:63-70.

Muñoz-Martín, Ricardo (2013): Cognitive and Psycholinguistic Approaches. *In*: Carmen Milán and Francesca Bartrina, eds. *Routledge Handbook of Translation Studies*. London/New York: Routledge, 252-267.

Newmark, Peter (1981): *Approaches to Translation*. Oxford/New York: Pergamon.

Nida, Eugene A. (1964): *Toward a Science of Translating*. Leiden: E. J. Brill.

Nord, Christiane (1991): *Text Analysis in Translation: Theory, Methodology and Didactic Application of a Model for Translation-Oriented Text Analysis*. Amsterdam: Rodopi.

Nord, Christiane (1997): *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Manchester: St. Jerome.

Osterlind, Steven J. (1997): *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance and Other Formats*. Boston: Kluwer Academics Publishers.

PAN, Yuling and DE LA PUENTE, Manuel (2005): *Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials: Documentation on how the Guideline Was Developed*. Washington: Statistical Research Division of the US Bureau of the Census.

PRIETO, Gerardo, MUÑIZ, José, ALMEIDA, Leandro S., *et al.* (1999): Uso de los tests psicológicos en España, Portugal e Iberoamérica. *Revista Iberoamericana de Diagnóstico y Evaluación psicológica*. 8(2):67-82.

REISS, Katharina and VERMEER, Hans J. (1984): *Grundlegung einer allgemeinen Translations-theorie*. Tübingen: Niemeyer.

RISKU, Hanna (2002): Situatedness in translation studies. *Cognitive Systems Research*. 3(3):523-533.

RISKU, Hanna (2004): *Translationsmanagement: Interkulturelle Fachkommunikation im Informationszeitalter*. Tübingen: Narr.

RISKU, Hanna (2010): A Cognitive Scientific View on Technical Communication and Translation. Do Embodiment and Situatedness Really Make a Difference? *Target*. 22(1):94-111.

SCHÄFFNER, Christina (1998): Skopos Theory. *In*: Mona BAKER, ed. *Routledge Encyclopedia of Translation Studies*. London/New York: Routledge, 235-238.

SHORT, Sandra E., SULLIVAN, Philip and FELTZ, Deborah L. (2005): Development and Preliminary Validation of the Collective Efficacy Questionnaire for Sports. *Measurement in Physical Education and Exercise Science*. 9(3):181-202.

SHREVE, Gregory M. (1997): Cognition and the Evolution of Translation Competence. *In*: Joseph H. DANKS, Gregory M. SHREVE, Stephen B. FOUNTAIN, *et al.*, eds. *Cognitive Processes in Translation and Interpreting*. London: Sage, 120-136.

SHREVE, Gregory M. and DIAMOND, Bruce J. (1997): Cognitive Processes: Critical Issues. *In*: Joseph H. DANKS, Gregory M. SHREVE, Stephen B. FOUNTAIN, *et al.*, eds. *Cognitive Processes in Translation and Interpreting*. London: Sage, 233-251.

SOLANO-FLORES, Guillermo, CONTRERAS-NIÑO, Luis Á. and BACKOFF-ESCUDERO, Eduardo (2006): Translation and Adaptation of Tests: Learned Lessons and Recommendations for Participant Countries in TIMSS, PISA and other International Comparisons. *Revista electrónica de investigación educativa*. 8(2). Visited on 18 September 2009, <http://redie.uabc.mx/vol8no2/contents-solano2.html>.

SWALES, John M. (1990): *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.

THORNDIKE, Edward L. (1910): The contribution of psychology to education. *Journal of Educational Psychology*. 1:5-12.

TIRKKONEN-CONDIT, Sonja (2000): Uncertainty in Translation Processes. *In*: Sonja TIRKKONEN-CONDIT and Riitta JÄÄSKELÄINEN, eds. *Tapping and Mapping the Processes of Translation and Interpreting*. Amsterdam/Philadelphia: John Benjamins, 123-142.

VALLERAND, Robert J. (1989): Vers une méthodologie de validation transculturelle de questionnaires psychologiques: implications pour la recherche en langue française. *Canadian Psychology*. 30(4):662-680.

VALLERAND, Robert J., BLAIS, Marc R., BRIÈRE, Natalie M., *et al.* (1989): Construction et validation de l'Échelle de Motivation en Éducation (EME). *Canadian Journal of Behavioural Sciences*. 21:323-349.

VINAY, Jean-Paul and DALBERNET, Jean (1958): *Stylistique comparée du français et de l'anglais: Méthode de traduction*. Paris: Didier.