City Research Online

City, University of London Institutional Repository

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Title: Using Cognitive Interviewing for the Semantic Enhancement of Multi-Lingual Versions of Personality Questionnaires

Short Title: Using Cognitive Interviewing for Test Adaptation

Abstract

We discuss the use of cognitive interviewing with bilinguals as an integral part of cross-cultural adaptation of personality questionnaires. The aim is to maximize semantic equivalence to increase the likelihood of items maintaining the intended structure and meaning in the target language. We refer to this part of adaptation as semantic enhancement, and integrate cognitive interviewing within it as a tool for scrutinizing translations, the connotative meaning, and the psychological impact of items across languages. During the adaptation of a work-based personality questionnaire from English to Arabic, Chinese (Mandarin), and Spanish, we cognitively interviewed 12 bilingual participants about 136 items in different languages (17% of all items), of which 67 were changed. A content analysis categorizing the reasons for amending items elicited eleven errors that affect two identified forms of semantic equivalence. We provide the resultant coding scheme as a framework for designing cognitive interviewing protocols and propose a procedure for implementing them. We discuss implications for theory and practice.

Keywords

Cognitive interviewing; Test adaptation; Cross-cultural assessment; Test translation; Equivalence

and bias.

Introduction

   The use of multilingual versions of questionnaires has become increasingly essential with the rise in economic interdependence between countries, prominent migration streams, and the rapid demographic changes around the world (van de Vijver, & Phalet, 2004). Assessment processes nowadays rarely include individuals from one cultural or ethnic background (Byrne, Leong, Hambleton, Oakland, van de Vijver, & Cheung, 2009; Daouk, Rust, & McDowall, 2005). Yet, many cross-cultural researchers continue to compare mean scores across cultures without ensuring that multi-lingual versions of tests are equivalent (Byrne et al., 2009).

   The International Test Commission (ITC) guidelines (Hambleton, 1994 and 2001), van de Vijver and Leung's (1997) Theory of Equivalence and Bias, and the Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials (Pan & de la Puente, 2005) laid the foundations for unifying practice in cross-cultural assessment. Such guidelines have been put in place in order to ensure that multilingual versions of tests are equivalent. Equivalence is related to whether 1) test takers taking an adapted version of a test can be meaningfully compared and / or 2) test takers taking different language versions can be compared (van de Vijver, 1998). In personality assessment across cultures, two multilingual versions of a test can be equivalent on construct, measurement unit, and scalar levels. Construct equivalence is concerned with ensuring that the construct exists in the target culture and whether it is defined and manifested in the same way as in the original culture. Ho (1996), for example, explains that the behaviors associated with being a good son or daughter, known as filial piety, are much broader in China than in most Western countries (as cited in Byrne & Watkins, 2003; van de Vijver & Hambleton, 1996). Therefore the questionnaire used to assess filial piety in China should contain a broader set of questions than the one used in Western countries. Once

construct equivalence is established, measurement unit equivalence can be assessed. This type of equivalence refers to situations where participants from different cultures perceive and interpret observed measures (items) similarly (Byrne & Watkins, 2003; Muller, 1995). To illustrate, van de Vijver and Tanzer (2004) explain that two scales could be measuring the same construct, say temperature, but using different scales, say Kelvin and Celsius. Similarly, multi-lingual versions of tests could be measuring the same construct though not necessarily using the same scale. Scalar equivalence, on the other hand, could be seen as the ultimate goal to be reached for assuming full score comparability between different language versions (van de Vijver, 1998; van de Vijver & Leung, 1997; van de Vijver & Tanzer, 2004). Scalar equivalence is full equivalence between two measures, indicating that they are functioning in the same manner across any cultures of interest. If scalar equivalence is achieved, tests are assumed to be bias free. For a full review of equivalence, see van de Vijver and Leung (1997).

Establishing any of these types of equivalence, however, depends on multi-lingual items being semantically equivalent; that is, equivalent in meaning (Behling & Law, 2001). Semantic equivalence is concerned with establishing the same meaning between source and target language statements by carefully choosing terms as well as sentence structures (Pan & Del Puente, 2005). Differences in the structure or wording of items in different language may lead to differences in interpretation and item responses, thus rendering the comparability of scores questionable.

The traditional back-translation method (Brislin, 1980) has been widely used across disciplines as the main tool for scrutinizing multi-lingual versions of questionnaires (Hambleton, 1993; Geisinger, 1994; Hambleton & van de Vijver, 1996; van de Vijver & Tanzer, 2004; Daouk, Rust, & McDowall, 2005). Although back translation offers valuable insights for

assessing the quality of the translation and for detecting potential problems, this method is fallible and is considered to be somewhat misleading when used on its own (Geisinger, 1994; Hambleton & Patsula, 1999). van de Vijver and Hambleton (1996) argue that this technique favors literal translations at the expense of readability and naturalness, which often go unnoticed. For example, a close match between the original and translated versions could be the result of a word-for-word translation, which often leads to nonsensical sentences in the target language. To illustrate, let us consider the idiom "everything is coming together" which implies that a situation is working out well in every way. Should this item be translated literally to Arabic, the back translation would reproduce the same wording in English. However, the literal translation of this idiom in Arabic implies that everything bad is happening at the same time. Hambleton (1993) highlights further criticisms about this judgment technique such as the difficulties arising from words or expressions with multiple meaning in the target language, and the discrepancy between the translation skills of the forward and back translators. For example, the word "sense" has several meanings in English. The sentence "something makes sense" could be mistranslated to "something makes feeling" in the target language, yet, a back translation may still come back as "something makes sense", hence the change in meaning in the target language may go undetected through this process. As another example of problems with back-translation, a more proficient back translator might correct a grammatical mistake committed by the forward translation, thus hindering the identification of this mistake in translation.

Other techniques for comparing cross-cultural versions of questionnaires are statistical, such as log linear modeling, confirmatory factor analysis, and item response theory (e.g., DeVellis, 2003, Griel, Jodoin & Ackerman, 2000; Swaminathan & Rogers, 1990; van de Vijver & Leung 1997; Zumbo, 1999) and are typically used to assess construct, measurement, and

scalar equivalence through test takers' responses quantitatively. Such techniques can detect the presence of discrepancies in responding, however, they fall short in that they do not provide insight into the source of such problems. It is therefore advantageous to complement them with qualitative techniques that can help identify the source of such discrepancies (Willis & Miller, 2011) since such approaches are better suited to investigate semantic equivalence. Generally, there is a need for the integration of in depth qualitative methods, focusing on the complex cognitive process that test-takers engage in when filling out multilingual versions of tests.

Cognitive Interviewing (CI) has been used extensively in survey development as a tool for detecting errors and increasing the quality of surveys (Dietrich & Ehrlenspiel, 2010; Redline, Smiley, Lee, DeMaio, & Dillman, 1998; Willis, 2005). The use of this technique, however, remains underexplored in the literature of cross-cultural test adaptation. Some studies have reported the use of CI for developing multi-lingual versions of national census surveys (e.g., Carrasco, 2003; Goerman, & Casper, 2010; Martinez, Marin, & Schoua-Glusberg, 2005; Pan, Sha, Park, & Schoua-Glusberg, 2009), and of health related surveys (Fujishiro et al., 2010; Kudela, Forsyth, Levin, Lawrence, & Willis, 2006; Willis et al., 2008). To our knowledge, the use of CI has not been explored in the adaptation of self-report personality measures to other languages and cultures. The application of this technique to cross-cultural research is relatively new and there is yet no uniform guide for practice (Willis et al., 2008). This prompts the need for more research reporting on the use of CI for adapting instruments to other languages and cultures.

It is the aim of our paper to contribute to the methodological aspect of the literature on cross-cultural test adaptation, as well as reporting results from an adaptation study. Particularly, we focus on improving the semantic equivalence of multi-lingual versions of personality

questionnaires using CI as an addition to the traditional forward and back translation process. Whilst recognizing the limited but growing use of CI in survey adaptation from one language to another, we propose utilizing CI with bilinguals, as opposed to monolinguals, as a cost and time effective approach to enhancing the quality of the cross-cultural adaptation of questionnaires. We operationalize this through the adaptation of a work-based personality questionnaire derived from the Five Factor Model of Personality, Orpheus© (Rust, 1996), from English to Arabic, Chinese (Mandarin), and Spanish. Although in this study we use a self-report measure specifically designed to assess personality in the workplace, the findings could be relevant to all other types of self-report measures.

Translation and Adaptation

Poor test adaptation is the main and most common source of lack of validity of inferences drawn from translated tests (Hambleton, 2001) and consequently lack of equivalence between them (van de Vijver, 1998). The literature provides a range of procedures that can be used to adapt questionnaires (Schweizer, 2010), such as forward translation, back translation, and expert panel reviews. While it is generally agreed that a combination of methods is essential for achieving equivalence between multilingual versions of tests, this cannot be always guaranteed (van de Vijver & Hambleton, 1996).

Methods of cross-cultural test adaptation are continuing to advance; yet they fall short in that they do not necessarily involve the test taker in the process of test adaptation. Given the complexity of any cognitive process the respondent goes through while answering questions (Murtagh, Addington-Hall, & Higginson, 2007), this appears as an omission that needs to be addressed. Cognitive pretesting is one option. Cognitive interviewing (CI) has been mainly

associated with test development and more recently, though in a limited manner, with cross-cultural test adaptation where understanding a statement as intended by the test developer is of extreme importance. The number of potentially flawed items detected during field-testing greatly impacts the likelihood of reaching overall equivalence between multilingual versions of a test (Grisay, 2003). However, investment in pretesting multilingual versions of survey instruments before field testing is not common partly due to the associated cost and time (Goerman, 2006). For impactful methods of cross-cultural test adaptation to become more widely used, it is paramount that they are perceived by test developers as, among other criteria, practical. It is a reality that the field of testing and assessment moves swiftly, responding to market needs, and test developers need to be in a position to respond to these. It is therefore our aim to report on the use of CI as a tool for detecting, from the test taker's perspective, potentially rectifiable item level anomalies when adapting personality questionnaires simultaneously to other languages. Since CI has been implemented in a number of ways (Willis, et al., 2008), we propose an approach that is focused on practicality as defined by Behling and Law (2001); that is, one that could be implemented quickly, cheaply, and easily, whilst maintaining methodological rigor and transparency.

Cognitive Interviewing techniques and their role in survey and test adaptation

Typically, CI involves in-depth interviews that aim to understand the cognitive process that goes through participants' minds when answering certain questions or items (Willis, 2005). This technique can be used by test developers for examining the respondents' interpretation of items in different languages and consequently detecting problems that could potentially be rectified (Wolley, Bowen, & Bowen, 2006). For example, Martinez et al. (2005) identified,

through the use of CI, specific improvements that were needed for establishing better semantic equivalence between the English and Spanish versions of the 2002 National Survey of Family Growth. Similarly, Levin et al. (2009) utilized the CI technique for semantically enhancing the Spanish language version of a dietary intake questionnaire and detected translation, culture specific, and general design issues that affect speakers of either language. CI could therefore be used as a tool of *semantic enhancement* of multi-lingual versions of adapted tests. We define semantic enhancement as the process of scrutinizing the technical qualities of translations, the connotative meaning, as well as the psychological impact of items on respondents across language versions.

Generally, there are two main techniques of CI: think aloud and verbal probing. For an overview of these two techniques, see Willis (2005). In the think aloud technique, participants are instructed to think out loud while attempting to answer questions. The interviewer's role centers on encouraging the participants to say what they are thinking with minimal interference (Redline, Smiley, Lee, DeMaio, & Dillman, 1998). The alternative verbal probing technique is in contrast based on continuous interaction between the interviewer and interviewee. The interviewer probes predefined semi structured questions that can help the interviewees verbalize their mental processes (Beatty & Willis, 2007; Willis, 2005). The cognitive interviewing method most commonly applied in practice is a combination of both techniques, whereby interviewers ask open-ended questions that can encourage the interviewee to think aloud while probing when necessary (DeMaio, Rothgeb, & Hess, 1998).

Another distinguishing characteristic of CI is the approach that could be coupled with the technique used. Any CI technique can adopt concurrent or retrospective probing approaches (Ericsson& Simon, 1980). In the concurrent approach, the cognitive interviewing takes place

9

while participants encounter the questions for the first time (Dietrich & Ehrlenspiel, 2010). In retrospective probing, participants first take the test and then are cognitively interviewed about their thinking processes (DeMaio, et al., 1998).

Research reporting on the use of CI for adapting questionnaires highlights some other variations in the use of this tool such as the number of interviews that the researchers conduct as well as the characteristics of the sample. These are important considerations that impact the practicality of CI (Behling, & Law, 2001) as they may render it expensive, time consuming and difficult to accomplish. That is, if the number of interviews is very high and the criteria for selecting participants are stringent, it may become difficult and expensive for test developers to recruit participants and time consuming to conduct interviews with all of them. Eventually, this may contribute to limiting the application of this tool in the context of cross-cultural test adaptation.

The number of interviewees reported in studies that pretested multi-lingual versions of surveys using CI ranged from 2 (Yam, Chow, & Ronen, 2005) to 35 (Carrasco, 2003) with the majority using around 20 (e.g., Goerman, & Caspar, 2003; Martinez, Marin, & Schua-Glusberg, 2006; Pan, Sha, Park, & Schua-Glusberg, 2009). Gathering cognitive data from a comparatively large sample can help to identify problems that are common to many respondents, disregarding those that are based on personal preferences alone. Going through the process with many participants may be both costly and time consuming however, particularly when adapting an instrument to more than one language. It may also be challenging for test developers to access large numbers of native speakers of the target language for 2 hours of in depth interviewing, especially if they are not based in that country.

Typically, research using CI in cross-cultural test adaptation involves monolingual

interviewees (e.g., Fujishiro et al., 2010; Kudela et al., 2006; Pan et al., 2009). Although there is great advantage to including monolinguals, such as their in depth understanding and familiarity with the target culture, there are methodological challenges that could be associated with this approach. After implementing CI with monolinguals regarding the Spanish translation of their survey, Goerman and Caspar (2010) found that they were still unable to evaluate the equivalence between the English and Spanish versions. As a follow up to this study, they tested both Spanish and English versions simultaneously using CI and found that they were better able to examine the questions across languages.

One other potential challenge of using CI in cross-cultural settings is that several investigators are often involved. As Willis and Miller (2011) argue, interviews conducted in other languages can only be analyzed by speakers of that language, which can be a potential source of bias. The authors also explain that for any information to be communicated between the investigators of the different languages, these have to be translated.

We therefore propose the use of CI with bilinguals, in addition to other methods of semantic enhancement, as an approach for facilitating the coordination between all investigators involved and maximizing the likelihood of semantic equivalence across language versions. We recognize that using bilinguals for evaluating two language versions of a test has advantages and disadvantages (Sireci, 2005). The most serious problem is that bilinguals tend to be very different to monolinguals in terms of their academic achievement (Sireci, 2005). Such differences are relevant to educational testing but to our knowledge no differences in personality have been documented between monolinguals and bilinguals.

Considering the length of personality questionnaires as well as the time and cost involved in CI, we also propose relying on one bilingual interviewee for every set of items in each target

11

language. This will make this method more practical to use, however, it can also have limitations. Therefore, we incorporate panel reviews following the CI as part of the semantic enhancement process to ensure any changes are not based on the opinion of one individual only. We will describe this procedure through the simultaneous adaptation of Orpheus© (Rust, 1996) personality questionnaire from English to Arabic, Chinese (Mandarin), and Spanish.

Method

The CI reported here was part of a larger test adaptation process of a work-based Five Factor Model questionnaire, Orpheus© (Rust, 1996) from English to Arabic, Chinese (Mandarin), and Spanish. The test adaptation process had two main phases: i) Semantic Enhancement and ii) Field-Testing. Cognitive interviewing was part of the Semantic Enhancement process, which was designed to increase the accuracy of the adaption in order to minimize the chance of having malfunctioning items after field testing (Sireci, 2005). Figure 1 illustrates this process. Each of the steps in this process is distinct and can be executed once the previous one is completed.

Comment to editor: Insert Figure 1 about here

For each target language (TL), Orpheus© (Rust, 1996) was first forward translated then reviewed by a panel of three experts that included the first author with a background in psychology and psychometrics as well as two native speakers of the TL. The two native speakers were either Masters or PhD students and at least one of them had a background in psychology and personality assessment and experience in translation. Native speakers were from different geographical areas in their home country. The versions revised by the expert panels were then back translated to English by a different translator followed by a review by the same panel of experts. The outcome of this translation phase was three adapted versions of Orpheus© (Rust,

12

1996) in the three target languages.

In the monitoring phase, the adapted versions consisting of 198 items each were pre-tested with approximately sixty bilingual participants in each target culture (Arab world n=62, China n=61, and Spain n=64). An electronic version of the questionnaire was developed and circulated via email to students in different universities through the student services office. Once 30 participants filled out the English version, the link was updated so that the other half would fill out the TL version. Within each culture, statistical mean comparisons were conducted in order to identify differences in item endorsement (Behling & Law, 2001). Although more sophisticated techniques exist for comparing differential item functioning (DIF; e.g. Zumbo, 1999), which typically involve the use of logistic regressions to determine and compare functioning across languages, t-tests can be used in this case in order to flag items as potentially malfunctioning, whether wrongly or rightly, given that they will be further scrutinized using CI. All items that were flagged as functioning differently, just over 17% of the original item pool (i.e. 792 items, 198 in each language), were then investigated further during cognitive interviewing to identify whether this was due to a problem arising from the adaptation process.

Participants

We recruited a convenience sample through personal contacts of the researcher and advertisements in higher education institutions (N = 12; 4 native Arabic speakers, 4 native Mandarin Chinese speakers, and 4 native Spanish speakers). In order to ensure sufficient command of both UK English and the TL, we stipulated that participants had to have lived in their home country most of their lives, had lived in the UK for at least 2 years, and had completed (or were currently completing) at least an undergraduate degree at a UK institution. Participants' age ranged from 23 to 61 years ($\bar{x}$=30.42; SD= 10.37 years). In each culture group,

there was one male and three females.

Materials

The material used in this study consisted of the items in English and in the TL (Arabic, Chinese, or Spanish) as well as the cognitive interviewing protocol discussed below.

Procedure

We adopted a concurrent approach with a combination of think aloud and verbal probing techniques for the cognitive interviewing protocol (Willis, 2005). In order to avoid cognitive fatigue, four interviews were held in each TL, whereby the items (48 Arabic, 45 Chinese, and 43 Spanish), were randomly divided into four batches. The first author conducted all interviews, each lasted approximately 2 hours and 12 to 16 items were probed in depth. This was followed by a panel review to ensure that any suggested amendments proposed by the participant were not due to individual preference.

During the first half of the interview, participants were presented with all items, one at a time, but in alternating languages so that each item was presented in one language only. Participants were instructed to paraphrase any item when presented in English, or to translate it to English when presented in the TL. Participants then rated their endorsement of the item (Strongly Disagree, Disagree, Agree, and Strongly Agree) and described their thinking process through structured questions such as: *How did you come up with this answer? Can you tell me a specific example that you thought about and that made you come to this decision?* The interviewer relied on pre-defined probes when the participant did not reveal enough information about the thinking process. For example, if the participant explained that they chose "*agree*" to *"I enjoy talking to my friends about work"* because *"I like to talk to my friends about work"*, the interviewer would probe further.

The same process of questioning and probing continued during the second half of the interview using the same items but in the alternate language. After reviewing each item, participants were presented with both versions and rated the similarity between them. They were asked to explain their choice through structured questions such as: *What does "A Word" mean to you?* They also provided amendments for the items where the similarity rating was anything other than "exactly the same".

All interviews were audio recorded and then transcribed once informed consent was obtained from each participant.

Analysis and Results

The total number of items analyzed during the cognitive interviewing was 136 (48 Arabic, 45 Chinese, and 43 Spanish), and participants proposed changes to 83 of them (37 Arabic, 16 Chinese, and 30 Spanish). These amendments were discussed in the panel of experts described earlier that include the first authors as well as two native speakers of the target language, with knowledge in personality assessment and experience in translation. The panel agreed on the proposed amendments to 67 items (33 Arabic, 10 Chinese, and 24 Spanish) and rejected 14 suggested amendments since they were perceived as either unnecessary for achieving semantic equivalence or representing the interviewer's personal preference.

In the second part of the analysis, the transcriptions of participants' CI of the 67 items were analyzed using Content Analysis (Stemler, 2001) in order to identify common sources of cross-cultural adaptation problems that can be detected by CI.

Data coding. Two researchers, with experience in qualitative analysis, worked independently on coding the data and devising a coding scheme. This entailed reading the participants' answers to the interview protocol and identifying the reason for the necessary amendments to the item.

These reasons where then coded into different categories, following a thorough discussion of coding differences to reach eventual agreement. As an example, during an interview with a male Spanish participant regarding the item "*it's often necessary to break the rules in order to get things done*", the participant interpreted the Spanish version as *"often, it is necessary to break the rules to do things"*. The participant indicated that he agreed to the statement but insisted that "*I agree but not strongly agree. You can do things without breaking the rule if you have plenty of time*". When responding to the English version, he strongly agreed with the item because "*many times when you really need to finish something, you need to break some rules otherwise you cannot get what you want in time*". When prompted further about the differences in his responding to the same item in the two languages, he pointed out that "*to do things*" and "*to get things done*" are very different because *"one might break the rules in order to meet deadlines, but they would not necessarily break the rules all the time to do things"*. Although a literal equivalent was available in the target language, the translation did not reflect this. The reason for amending this item was therefore coded as 'Not Literal Translation'.

During a different interview, a Chinese participant was asked to explain in her own words what she understood by the English item "*it always pays to tell the truth*" to which she responded:*"there's a price to pay if you try to promise to tell the truth"*. Since the participant misunderstood the intended meaning from the English item due to the idiosyncratic nature of this expression, the interviewer provided her with the correct explanation. Interestingly, when she encountered the Chinese version of this item, she discovered that it was translated in the same way she first misunderstood the English item. The reason for amending this item was thus coded as 'Idioms' since the term "*it pays to tell the truth*" was not translated in its idiomatic sense.

As a final example, when presented with the item "*I am sometimes too rash in making*

16

decisions" in English, a native Arabic speaking participant disagreed with it: *"I don't agree. I usually think my decision especially if they are big."* However, when presented with the Arabic version, the participant indicated that she agreed with the statement. Upon further probing and presenting the two language versions, the participant indicated that *"the two versions are not very similar, the Arabic version is lighter and I would be more likely to say yes to it".* When probed further about the meaning of "lighter", she explained that "*the translation of "rash" in Arabic is the same as the English word if you look it up in the dictionary but I don't see them as the same".* The proposed amendment to the item was approved by the expert panel as a reasonable argument and the item was re-worded in order to convey a comparable psychological impact in both languages. The reason for this amendment was therefore coded as "Different Magnitude" because the terms were linguistically the same but had differing psychological magnitude in either language.

Categorization and Reconciliation. We then reviewed all our categories to ensure consistency in coding and wording by reviewing all items in each category, and to assess whether coding was accurate (Cohen's Kappa= 0.67). This process resulted in eleven categories that represent sources of adaptation errors that impact semantic equivalence from two perspectives, namely: linguistic and connotative. The outcome of this exercise was the coding scheme presented in Table 1.

<div align="center">Comment to editor: Insert Table 1 about here</div>

Linguistic equivalence

The term "linguistic equivalence" has previously been used to refer to similarity of wording that reproduces the same meaning in the target language (i.e., Butcher, Cheung, & Lim, 2003). Since this definition is congruent with that of Semantic Equivalence as defined by

Behling & Law (2001), we use the term Linguistic Equivalence as one of two subtypes of Semantic Equivalence. The linguistic form of semantic equivalence was the result of mistranslations, syntactic, or stylistic errors that hindered the reproduction of technically comparable multi-lingual versions. Six errors that affected linguistic equivalence were identified, namely: 1) literal translation available but not used, 2) words or sentence grammatically not equivalent, 3) other grammatical mistakes, 4) clumsy wording, 5) wrongly added or omitted words, and 6) context dependent synonyms. For example, "写作风格" and "工作作风" both mean style in Chinese but the first one refers to the style of writing whereas the latter refers to style of working. This is an example of context dependent synonyms that illustrates a subtlety in a particular language that impacts the quality of the translation in the TL; yet, the meaning could still be captured by the respondents. This type of anomaly is the result of inaccuracy in translation that could be rectified by applying similar grammatical rules or vocabulary in both languages. Although participants are usually able to understand such items even before amending them, this type of error can affect potential test takers' perception of the test and their faith in the validity of the inferences drawn from it.

Connotative equivalence

The connotative form of semantic equivalence on the other hand was the result of language idiosyncrasies particular to one language but not the other. This led to differences in meaning and consequently differential psychological impact on participants in the two language versions. van de Vijver and Jeanrie (2004), Butcher (2004), and Butcher, Cheung, and Lim (2003) also recognize the importance of the comparability of psychological impact of items in reaching full equivalence between them. We identified 5 errors that affect connotative equivalence between items: 1) wrong translation of words with multiple meaning, 2) composed

words, 3) words nonexistent in TL, 4) idioms and 5) different magnitude. For example, the literal Arabic translation of the word "discouraged" is " محبط"(pronounced "mouhbat"). These two words, however, resulted in a different reaction from the same participant. As suggested by the participant and agreed by the expert panel, the Arabic version carried a more negative connotation than its English equivalent as it was closer to "depression" than "discouragement". This is an example of words having "different magnitude" in different languages. In such cases, the meaning of the items is affected and would be understood differently in the TL than in the source language. In order to semantically enhance the connotative meaning of the item, different grammatical rules or vocabulary may have to be applied.

Discussion

This study contributes to the literature of cross-cultural adaptation of personality tests in several ways. First, we present an example of a systematic process of adapting personality questionnaires into other languages with focus on semantic enhancement. We highlight the role of CI as an integral part of this process for incorporating the test taker's perspective qualitatively, arguing that this approach has many advantages over the traditional use of back translation alone. We propose the resultant coding scheme as a framework for designing CI protocols to be integrated into a practical approach for conducting cognitive interviewing relying on bilingual interviewees.

Systematic process of personality questionnaire adaptation

We adapted Orpheus© (Rust, 1996) personality questionnaire from English to Arabic, Chinese (Mandarin), and Spanish using a systematic process of semantic enhancement. The process is divided into two parts, translation and monitoring, as outlined in Figure 1. This study

focuses on CI as an integral process for monitoring the quality of test adaptations at the item level. Typically, cross-cultural adaptation processes rely on qualitative input from the translator and the expert judges. Whilst pertinent guidelines and recommendations exist in the literature (Hambleton, 2001; Hambleton, Merenda, & Spielberger, 2005), they consider the perspective of potential users solely quantitatively through pilot studies. Statistical results can only point to differences in item functioning but not to the reasons that lead to these differences (Slocum, Gelin, & Zumbo, 2003). An observed discrepancy is not always a reflection of problems with cross-cultural adaptation processes as it could also be the result of real differences between the two groups (Zumbo, 1999). When discrepancies are detected, items need to be scrutinized in order to unveil the origins of the difference in performance between groups (SIOP, 2003; Zumbo, 1999; Slocum, Gelin & Zumbo, 2003).

CI as a means to semantic enhancement adds value to the cross-cultural adaptation process because it provides clues for rectifying items and thus increases the chance of them functioning similarly to their English equivalent. The CI technique can therefore be used as a structured tool for interpreting statistically significant item-level differences and identifying problems that could be overcome by manipulating grammatical rules or vocabulary in order to reach better linguistic and connotative forms of semantic equivalence. Specifically, we advocate the use of CI as a tool for modifying personality questionnaires that will be used cross-culturally to increase the degree of validity of the inferences drawn across those data. The psychological impact of items is crucial for such tools to ensure that appropriate inferences can be drawn about the degree to which they measure the underlying construct.

In this regard, CI offers distinct advantages over traditional back translation followed by statistical scrutiny. We do not discount the value of back translation which we also used as one

20

process to assist in the development and subsequent review of multilingual versions. Our analysis underlines however the potential fallibility of back translation as a *standalone* process, since this method could not have identified the translation errors elicited through the CI interviewing technique. Magnitude is a pertinent example here. Words such as 'never' differ in subtle ways in their emphasis across languages. Only the involvement of our CI interviewees allowed us to highlight and record this aspect, as the technique allows users to verbalise their understanding of the item, even where endorsement may not necessarily differ. We now offer practical suggestions for doing so in future research based on our findings, including the emphasis of connotative equivalence, and the involvement of bilingual individuals.

The value of coding adaptation errors for screening equivalent versions

Six adaptation errors presented in the coding scheme in Table 1 impact linguistic equivalence; whereas five errors impact connotative equivalence. These findings may offer a framework for semantically enhancing multilingual version of items during test adaptation. We propose two ways for doing so. First, the reasons for amending items, as coded above, can be used to screen items that will need to be cognitively interviewed in subsequent studies, such as those with idioms or complex grammatical structures. Failing to limit the number of items cognitively interviewed may impose a burden on respondents (Napoles-Springer, Santoyo-Olsson, O'Brien, & Stewart, 2006) that can make CI more time consuming, costly, and difficult to accomplish. When investigators pre-test items statistically before CI, this can be an alternative way of limiting the number of items to be cognitively interviewed. Therefore, another way of using the coding scheme is for designing the CI protocol. The protocol could be scripted in a way that captures linguistic and connotative meanings, whereby probes could be designed to prompt respondents about, say, wrongly omitted words or grammatical equivalence, to ensure such

mistakes do not go unnoticed.

Connotative and linguistic equivalence

As alluded to above, we became sensitized during the course of this research to the potential issues arising from translating items in too literal a sense. Whilst literal translation can initially seem important for insuring full equivalence, in some cases this may lead to misinterpretation or misunderstanding in the target language. Investigators involved in adapting questionnaires need a paradigm for balancing between linguistic and connotative types of semantic equivalence. A case in point is the following example of adaptation errors. "Literal translation available but not used" was noted as an error that challenges reaching linguistic equivalence whereas "literal equivalent not most appropriate" was categorized as affecting connotative equivalence. This finding suggests that, the linguistic equivalence between two multi-lingual items could be reached if the literal translation is used. In some instances, however, the literal equivalent may not be appropriate if it impacts the connotative equivalence between items. Using the literal equivalent may therefore result in different psychological impacts between the two language versions.

In personality assessment, equivalence in the psychological impact of an item is critical for reaching full equivalence between multi-lingual versions. This can be remediated by adopting words and sentence structures that are more closely aligned with the TL. Therefore, reaching semantic equivalence between multi-lingual versions of tests is a function of both linguistic and connotative equivalence. Although both forms of semantic equivalence identified in this study are essential, we argue that connotative equivalence prevails over linguistic equivalence. Maintaining linguistic equivalence is important to ensure that individual preferences of investigators are minimized. This specific type of semantic equivalence becomes less important

when connotative equivalence is not achieved. In such cases, investigators should risk linguistic equivalence at the expense of reproducing the same meaning or psychological impact in both languages.

The involvement of Bilinguals

Behling and Law (2001) identify four criteria for evaluating methods of cross-cultural test adaptation: Informativeness, Security, Source language transparency and Practicality. In general, cognitive interviewing satisfies the first two criteria because it provides valuable information pertaining to problems with semantic equivalence (informativeness) and it creates an opportunity to scrutinize the quality of the translator's work (security). The other two criteria, however, may not be satisfied when CI is used with monolinguals. In the case of adaptation into more than one language, one is unlikely to find a lead investigator proficient in all target languages. Therefore, evaluating source language transparency is difficult given that the investigator may not be able to understand and solve problems with the TL instrument.

Using CI with bilinguals assisted us in tackling this issue. Their presence provided the researchers who lack fluency in the TL with the opportunity to understand item level anomalies across languages and the suggested solutions. This understanding may in turn facilitate the transfer of knowledge between the parallel developments of the multi-lingual versions of the instrument. Therefore, involving bilinguals allows for the direct comparability of semantic equivalence, which otherwise would not have been possible (Goerman, & Caspar, 2010). The interview process and resulting analysis and discussion thus facilitates a skill building element for the research team who, whilst not fluent in all adaptation languages, nevertheless glean fuller insight into the translation issues and difficulties particular to certain items in certain languages.

Having documented and analysed these issues in a systematic way, our framework for

identifying errors may now assist other researchers in their own adaptations. As for practicality, recruiting monolingual interviewees can sometimes be difficult to achieve. Investigators interested in developing multi-lingual versions of questionnaires may be located in countries where access to monolinguals is challenging. For example, it may be easier for investigators in the US to recruit Spanish monolinguals than for researchers working in the UK where the demographics of immigrants are different. Therefore, including the perspectives of bilinguals is a practical way for scrutinizing the semantic equivalence of questionnaires. Nonetheless, subsequent expert panel reviews are still essential for minimizing individual preferences bias. We advocate that the combination of back translation and CI should always be followed by such a review. In all, using CI with bilinguals facilitated the interviewers' understanding of how the instrument may be understood by test takers across languages thus providing Informativeness, Security, Source language transparency, and Practicality.

Limitations and future directions

We recruited 4 bilingual interviewees in every language, and divided the items between them, so that every item was reviewed once but in both source and target languages. Whilst recognizing the benefits of having more than one interviewee review each item, this would have required additional cost, time, and commitment on behalf of participants. We argue that having one interviewee review a handful of items facilitated the recruitment of participants and minimized the cost associated with CI. We addressed the issue of a potentially one-sided perspective through the expert panel review. Each suggestion by the bilingual interviewee for amending an item was scrutinized and either agreed or refuted. This discussion process ensured that the reasoning was focused on semantically enhancing the item with sound justification,

rather than a matter of personal preference. This may strengthen the practicality of incorporating CI by cross-cultural test developers around the world. A fruitful area of future research could focus on comparing bilinguals to monolinguals for enhancing the semantic equivalence between multi-lingual versions of personality questionnaires. It may also be worthwhile to investigate the number of interviewees that need to be recruited for reaching saturation when CI is as part of a systematic process of semantic enhancement that includes forward and back translation as well as several expert panel reviews. Recruiting expert panels in itself may pose a logistic challenge, as individuals are required to have very specific linguistic skills, as well as be available at certain times. Nevertheless, we tackled this challenge in the present context through reiterating the potential benefits. Feedback from the expert panel indicated that they indeed learned from the experience, and would be willing to take part again. If face to face recruitment poses a challenge, future research may also make use of new technologies to facilitate panel meetings.

A further observation arising from this study is that particular items could not be rectified in the target language. For example, a word might be seen as milder in one language compared to another or may not have an equivalent word in the target language; yet, there may be no suitable alternative words that could create an equivalent semantic meaning and thus psychological impact in both languages. An example is the Arabic word for "fantasy", which could at best be translated to "imagination" since there is no equivalent for that word in Arabic. Items that include such words are by nature idiosyncratic and are unlikely to ever achieve equivalence. This is of course particularly true for emic questionnaires, that have been developed in a specific language and cultural context, perhaps long before adaptation was even conceived as a design consideration. Such items might have to be dropped after the field test and will impact the degree of validity of inferences drawn from questionnaires.

An etic approach focused on the concurrent development of multilingual versions of tests might be the solution for this limitation as opposed to developing the test in one language followed by adaption to other TLs. Finally, we note that we employed CI as an enhancement method prior to field testing. This study does not report DIF analysis to test out if rectified items are equivalent in all languages, as a) this would be beyond the scope of the current paper and b) where differences are detected these could point to genuine group differences, rather than mere translation errors. Thus, future research may embed a second iteration of CI following a field testing phase.

Another worthwhile direction for future research could be the application of non-statistical methods for classifying items that may benefit from further investigation through cognitive interviewing. In the current study, we referred items to the CI process based on quantitative differences in ratings. However, future studies may also include a qualitative review process by referring for instance items containing words with multiple meaning and/ or idioms as a matter of course. It will be imperative in this instance to work closely with the translation professionals and seek their feedback and input, too. Finally, our check list as presented in Table 1 may also be adapted to serve as a guide during simultaneous test construction. In the present paper, we adapted an existing questionnaire which has been used in the original language for some time. Other researchers or practitioners may wish to develop new instruments with parallel version in different languages from the outset. Our classification may also serve as a blue print for a check during such a parallel development process, to ensure that semantic features that as language idiosyncratic idioms are avoided from the outset.

Conclusion

As demonstrated in this study, CI complements other methods of cross-cultural

adaptation in order to create a more rigorous process of semantic enhancement. The greater the number of items preserved after field testing, the more likely one is to achieve full equivalence between multi-lingual versions of an instrument (Grisay, 2003).

CI can be a practical, integral, and rigorous methodological part of multi-lingual test adaptation processes by involving bilingual test takers to enhance items prior to field testing. We argue that CI has advantages over and above traditional methods such as back translation and statistical testing for DIF due to its effectiveness in detecting even subtle differences in meaning and understanding through structured discussion with interview participants. The coding scheme that resulted from our CI process has the potential to inform the design of subsequent CI studies as well as standardizing test adaptation processes by identifying common translation errors. Thus, we recommend that CI should be used as an integral part of multi-lingual test adaptation for personality measures in conjunction with more established methods.

References

Behling, O., & Law, K.S. (2001). *Translating Questionnaires and Other Research Instruments: Problems and Solutions.* London: Sage Publications.

Beatty, P.C., Willis, G.B. (2007). Research Synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71,* 287–311.

Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology*, *1*, 389-444. Boston: Allyn & Bacon.

Butcher, J. N. (2004). Personality assessment without borders: Adaptation of the MMPI–2 across cultures. *Journal of Personality Assessment, 83,* 90–104.

Butcher, J. N., Cheung, F. M., & Lim, J. (2003). Use of the MMPI–2 with Asian populations. *Psychological Assessment, 15,* 248–256.

Byrne, B. M., Leong, F. T., Hambleton, R. K., Oakland, T., van de Vijver, F. J., Cheung, F. M. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Education in Professional Psychology*, *3,* 2, 94-105.

Carrasco, L. (2003). *The American Community Survey (ACS) en Español: Using Cognitive Interviews to Test the Functional Equivalency of Questionnaire Translations.* Study Series: Survey Methodology #2003-17. Methodology and Standards Directorate, U.S. Census Bureau.

Daouk, L., Rust, J., & McDowall, A. (2005). Testing across languages and cultures: challenges for the development and administration of tests in the internet era. *Selection and Development*

Review, 21, 4, 11-1.

DeMaio, T.J., Rothgeb, J., & Hess, J. (1998). *Improving survey methods through pretesting*.
Washington, DC: US Bureau of the Census.

DeVellis, R.F. (2003). *Scale Development: Theory and Applications.* Sage, Thousand Oaks, CA,
USA.

Dietrich, H., & Ehrlenspiel, F. (2010). Cognitive interviewing: A qualitative tool for improving
questionnaires in sport science. *Measurement in Physical Education and Exercise Science*,
14, 51 – 60.

Ericsson, K., & Simon, H. (1984). *Protocol Analysis: Verbal Reports as Data.* Cambridge, Mass.:
MIT Press.

Fujishiro, K., Gong, F., Baron, S., Jacobson Jr, C.J., DeLaney, S., Flynn, M., & Eggerth, D.E.
(2010). Translating Questionnaire Items for a Multi-Lingual Worker Population: The
Iterative Process of Translation and Cognitive Interviews with English-, Spanish-, and
Chinese-Speaking Workers. *American Journal of Industrial Medicine, 53,* 2, 194-203.

Geisinger, K.F. (1994). Cross-cultural normative assessment: Transition and adaptation issues
influencing the normative interpretation of assessment instruments. *Psychological
Assessment, 6*, 304–312.

Gierl, M. J., Jodoin, M. G., & Ackerman, T. A. (2000, April). *Performance of Mantel-Haenszel,
simultaneous item bias test, and logistic regression when the proportion of DIF items is
Large.* Paper presented at the annual meeting of the American Educational Research
Association, New Orleans, LA.

Goerman, P. (2006, March). *An Examination of Pretesting Methods for Multicultural, Multilingual
Surveys*. Paper presented at the Third International Workshop on Comparative Survey

Design and Implementation, Madrid, Spain.

Goerman, P.L., & Caspar, R.A. (2010). A Preferred Approach for the Cognitive Testing of

Translated Materials: Testing the Source Version as a Basis for Comparison. *International*

Journal of Social Research Methodology, 13, 4, 303-316.

Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international Assessment.

Language Testing, *20,* 2, 225-240.

Hambleton, R. K. (1993). Translating achievement tests for use in cross-national studies.

European Journal of Psychological Assessment, 9, 57-68.

Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress

report. *European Journal of Psychological Assessment*, *10,* 229-244.

Hambleton, R. K. (2001). The next generation of the ITC test translation and application

guidelines. *European Journal of Psychological Assessment*, *17,* 3, 164-172.

Hambleton, R. K., & Kanjee, A. (1995). Increasing the validity of cross-cultural assessments: Use

of improved methods for test adaptations. *European Journal of Psychological Assessment*,

11, 147-157.

Hambleton, R. K., Merenda, P. F., Spielberger, C. D. (Eds.) (2005). *Adapting educational and*

psychological tests for cross-cultural assessment. Mahwah: Lawrence Erlbaum Associates.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be

avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing*

Technology, 1, 1-30.

Hughes, K.A., & DeMaio, T.J. (2002). Hughes, K., & DeMaio, T. (2001). *Does this question*

work? Evaluating cognitive interview results using respondent debriefing questions. Paper

presented at the Annual Meeting of the American Association for Public Opinion Research, St. Petersburg, Florida.

Kudela, M. S., Forsyth, B. H., Levin, K., Lawrence, D., & Willis, G. (2006, May). *Cognitive interviewing versus behavior coding.* Paper presented to the American Association of Public Opinion Research, Montreal, Canada.

Levin, K., Willis, G.B., Forsyth, B.H., Norberg, A., Kudela, M.S, Stark, D. & Thompson, F.E. (2009). Using Cognitive Interviews to evaluate the Spanish-Language Translation of a Dietary Questionnaire. *Survey Research Methods, 3,* 1, 13-25

Martinez, G., Marín, B. V., & Schoua-Glusberg, A. (2006). Translating From English to Spanish: The 2002 National Survey of Family Growth. *Hispanic Journal of Behavioral Sciences, 28,* 531-545.

Martinez, G., Marín, B.V., Schoua-Glusberg, A. (2011). Analyzing Cognitive Interview Data Using the Constant Comparative Method of Analysis to Understand Cross-Cultural Patterns in Survey Data. *Field Methods, 23,* 420-438.

Murtagh, F.E.M., Addington-Hall, J.M., & Higginson, I.J. (2007). The value of cognitive interviewing techniques in palliative care research. *Palliative Medicine*, *21,* 87-93.

Napoles-Springer A.M., Santoyo-Olsson J., O'Brien H., & Stewart, H.B.A. (2006). Using cognitive interviews to develop surveys in diverse populations. *Medical care*, *44,* 21–30.

Pan, Y., & de la Puente, M. (2005). *Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials: Documentation on how the Guideline Was Developed.* Statistical Research Division's Research Report Series (Survey Methodology #2005-06). Washington, DC: U.S. Census Bureau. http://www.census.gov/srd/www/byname.html#panyuling.

Pan, Y., Sha, M., Park, H., & Schoua-Glusberg, A. (2009). 2010 *Census language program: Pretesting of census 2010 questionnaire in five languages*. Research report series (survey methodology #2009-01). Statistical Research Division, US Census Bureau. Retrieved January 15, 2012, from http://www.census.gov/srd/papers/pdf/rsm2009-01.pdf

Redline, C., Smiley, R., Lee, M., DeMaio, T., & Dillman, D. (1998). *Beyond concurrent interviews: an evaluation of cognitive interviewing techniques for self- administered questionnaires*. Retrieved September 15, 2007, from US Census Bureau Website: http://www.census.gov/srd/papers/pdf/sm98-06.pdf.

Rothgeb, J., Willis, G., & Forsyth, B. (2001). *Questionnaire pretesting methods: do different techniques and different organizations produce similar results*. Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Montreal, Canada.

Rust, J. (1996). *Orpheus Handbook*. The Psychological Corporation, London and San Antonio.

Schweizer, K. (2010). The adaptation of assessment instruments to the various European languages. *European Journal of Psychological Assessment, 26,* 75–76.

Society for Industrial and Organizational Psychology (SIOP), Inc (2003). *Principles for the validation and use of personnel selection procedures*. Retrieved July 6, 2007, from http://www.siop.org/_Principles/principles.pdf

Sireci, S. G. (2005). Using bilinguals to evaluate the comparability of different language versions of a test. In R.K. Hambleton, P. Merenda, & C. Spielberger (Eds.). *Adapting educational and psychological tests for cross-cultural assessment* (pp. 117-138). Hillsdale, NJ: Lawrence Erlbaum.

Slocum, S. L., Gelin, M. N., & Zumbo, B. D. (2003). Statistical and graphical modelling to investigate differential item functioning for rating scale and Likert item formats. In B. D.

Zumbo (Ed.), *Developments in the Theories and Applications of Measurement, Evaluation, and Research Methodology Across the Disciplines*, *Vol. 1*. Vancouver: Edgeworth Laboratory, University of British Columbia.

Stemler, S. (2010). An overview of content analysis. *Practical Assessment, Research and Evaluation, 7, 17*. Available from: http://PAREonline.net/getvn.asp?v=7&n=17

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27,* 361-370.

van de Vijver, F. J. R. (1998). Towards a Theory of Bias and Equivalence. *ZUMA Nachrichten Spezial*, *3,* 41–65.

van de Vijver, F. R., & Hambleton, R. K. (1996). Translating tests: some practical guidelines. *European Psychologist, 1,* 2, 89–99.

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

van de Vijver, F., & Phalet, K. (2004). Assessment in multicultural groups: The role of acculturation. *Applied Psychology: An International Review, 53,* 2, 215-236

Willis, G.B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.

Willis, G., Lawrence, D., Hartman, A., Kudela, M.S., Levin, K., & Forsyth, B. (2008). Translation of a tobacco survey into Spanish and Asian languages: The tobacco use supplement to the Current Population Survey. *Nicotine & Tobacco Research, 10,* 1075-1084.

Willis, G., & Miller, K. (2011). Cross-Cultural Cognitive Interviewing: Seeking Comparability and Enhancing Understanding. *Field Methods, 23,* 4, 331-341.

Woolley, M.E., Bowen G.L., & Bowen N.K. (2006). The development and evaluation of
procedures to assess child self-report item validity. *Educational and Psychological
Measurement, 66,* 687-700.

Yam, W.K.L., Chow, S.M.K., & Ronen, G.M. (2005). Chinese version of the Parent-Proxy
Health-Related Quality of Life Measure for Children with Epilepsy: Translation, cross-
cultural adaptation, and reliability studies. *Epilepsy & Behavior, 7,* 4, 697–707.

Zumbo B.D. (1999). *A handbook on the theory and methods of differential item functioning (DIF):
Logistic regression modelling as a unitary framework for binary and Likert-type (ordinal)
item scores.* Ottawa, ON: Directorate of Human Resources Research and Evaluation,
Department of National Defence.

Figure 1: The semantic enhancement process during test adaptation

Table 1: The two types of semantic equivalence and eleven adaptation errors that impact them

Type of Semantic Equivalence	Definition	Adaptation error	Exemplars	Solution
Linguistic Equivalence	Any mistranslations, syntactic or stylistic errors.	**1. Literal equivalent available but not used**	Using *"boring" instead of "tedious"* when *"tedioso"* exists	Could be rectified by applying similar grammatical rules or vocabulary in both languages in order to reach better linguistic equivalence
		2. Word(s) or sentence grammatically not equivalent	*"I take risks"* and *" I am a risk taker"* have similar meanings but different grammatical use of words	
		3. Other grammatical mistake	*" I has eaten"* instead of *"I have eaten"*	
		4. Clumsy wording	*"Always, I prefer to be prepared"* rather than *"I prefer to always be prepared"*	
		5. Wrongly added or omitted word(s)	Omitting the word *"almost"* from the following sentence *"My work is more important to me than almost anything else"*	
		6. Context dependent synonym(s)	"التعاطي" (altaati) means dealing but usually drugs, or commerce. "التعامل" (altaamoul) refers to dealing but with	

people

Connotative Equivalence	Any differences in meaning between two language versions due to idiosyncrasies particular to one language but not the other.	**7. Wrong translation of words with multiple meaning**	"*Change*" meaning money wrongly translated to "*change*" as in modification	Could be rectified by applying different grammatical rules or vocabulary in the two languages in order to reach better comparability of meaning
		8. Composed word(s)	"*Punctuality*" is "*precision in time*" in another language	
		9. Word(s) nonexistent	"*fantasy*" does not exist in one language	
		10. Idiom(s)	"*fly-by-night schemes*" or "*go out of my way*"	
		11. Different magnitude	"nunca" and "jamás" both mean "*never*" but the latter is closer to "*never ever*"	