Cultural Adaptation of Research Instruments – The Case of Materialism Scales Culturally Adapted for Use in China

Bogdan Robert Ioane[1]

[1] Higher Colleges of Technology, Al Ain, Abu Dhabi, UAE

Correspondence: Bogdan Robert Ioane, Business Faculty at Higher Colleges of Technology, Al Ain Colleges, Zayed Al Awwal Street, Falaj Hazza, Al Ain, Abu Dhabi, P.O.Box 17258, United Arab Emirates.

Abstract

Introduction: This study seeks to understand whether a back translation process can produce culturally adapted statements. Economic materialism levels are often studied using materials designed in English for the Western subject. This study revolves around culturally adapting, for use in China, economic materialism research instruments that were originally created in English for use in the West.

Methods: This study used an instrument translation process followed by a validation step in order to culturally adapt research constructs not originally designed for subjects in China. The study consisted of six main steps: forward translation, reconciliation, blind back translation, expert committee review, validation, and statistical analysis.

Results: Inadequate back translated items were identified. The analysis revealed several items that should be redesigned specifically for the Chinese cultural context. This study shows that the back translation process may be insufficient, and that Western-developed research instruments can be improved through a process of cultural adaptation.

Keywords: translation, cultural adaptation, materialism, China

1. Introduction

China today is a far different place from the China of yesteryear. Where once its society was rabidly anti-materialistic, as it practiced "puritan communism" (Zhao, 1997, p. 42), today the Chinese people are hungry for consumer goods, as though in response to the many years of deprivation (Zhao, 1997). Chan and Prendergast (2007) define materialism as "a set of attitudes which regard possessions a symbols of success, where possessions occupy a central part of life, and which include holding the belief that more possessions lead to more happiness" (p. 213). However, efforts to research materialism and the values of material possession and acquisition, whether in China or elsewhere, commonly centre on research instruments such as surveys that have been written originally in English and designed for a Western subject. Given the substantial differences in both culture and language between China and Western countries (Doctoroff, 2005; Nisbett, 2003), this approach is problematic. Research instruments that have been identified, validated, and declared reliable for Western contexts may generate invalid results in China, due to important differences between the Western and the Chinese cultures (Chan & Prendergast, 2007). The translation process used to adapt these Western research constructs for the Chinese population commonly involves a verification step, back translation, by which the material that has been translated in Chinese is then translated back into English, to verify its adherence to the original document, which means that the English language and Western culture remain the hegemonic force in these translations. Nonetheless, back translations are commonly used in cross-cultural research.

The purpose of this study was to analyze two materialism value scales (MVS) designed for Western consumers, and to translate and culturally adapt them for use with Chinese consumers, with the goal of achieving more reliable results. Materialism has been scrutinized in some depth by researchers, but cross-cultural research concerning the Asian consumer market is now becoming important due to the rise of Asia in the global markets. The cultural and language differences between the East and West call for new approaches to understanding the values of the population (Podoshen, et al., 2011; Watchravesringkan, 2012; McCarthy et al., 2013).

This study translated, and culturally adapted for use in China, the 18-item Materialism Value Scale MVS (Richins & Dawson, 1992) and the 24-item Materialism Scale (Belk, 1984). These are the two scales that are most frequently used by researchers (Watchravesringkan, 2012). Some researchers have recently suggested a shortening of the scales, such as Opree, et al. (2011) in their work with children, and Trinh and Phau (2012), in their development of a 16-item scale. Nonetheless, the 18-item MVS of Richins and Dawson, and Belk's 24-item MVS remain the classic scales used in research. In addition, there is a positive relationship between the two materialism instruments, as the Richins MVS scale correlates with Belk's instrument of materialism (Richins, 2004). Richins MVS approaches materialism from the values perspective - that materialism and the desire for possession and acquisition of goods result from commonly-held values in a society; Belk's MVS approaches materialism as resulting from character traits (Zarco, 2014). Therefore, the two scales were expected to measure materialism from complementary, overlapping perspectives. The intention was to observe whether a simple back translation is sufficient for translating research instruments appropriately for research.

2. Literature Review

a. Translation and Cultural Adaptation of Research Instruments

Translation processes can include two approaches: forward translation and back translation. Forward translation is the process of translating the research document, or instrument, from its original language into a target language. Back translation is the process of translating the result of the forward translation back to the original language. When a different culture is involved, however, forward and back translation are not sufficient to attain cultural equivalence with the original instrument (Cannino, Bird, Rubo-Stipe, & Bravo, 1997; Watchravesringkan, 2012). Each item has to contain the nuances of the original instrument, and if the items are not translated exactly the same way, there is a risk that the translated item will refer to an adjacent concept or meaning. In some cases, only part of the original meaning is able to be preserved after translation (Cannino et al., 1997). The loss of meaning, the narrowing down of item meaning, or the expansion and broadening of the coverage of an item can be significant issues during both forward and back translation. To avoid such issues, a process of cultural adaptation that follows a mixed qualitative and quantitative approach is superior to arbitrary decision making done by individuals during research because it quantifies the extent of agreement on the appropriateness of the translation and its cultural fit (Sumathipala & Murray, 2000).

The cultural adaptation of research instruments is the most challenging and sensitive issue of translation, in part because of the multitude of languages and cultures. South Africa has nine official languages, as does India. The French language is significantly different in France than in Canada, and Spanish is different across Spanish-speaking countries. German is also spoken differently in Belgium, Switzerland, and Germany.

In China, there are five mutually unintelligible varieties of Chinese language, and under each are multiple dialects that are variations and modifications of the five large groups. There is also distinction between the spoken and the written language: Zhongwen (中文), written language, and Hanyu (汉语/漢語), spoken language. The Chinese language is supported by a pictographic script, and the language needs to provide pictures even in speech (Ozolins, 2009). For example, concepts like 'length' are explained as 'how short' or 'how long,' whereas 'very difficult' is a concept translated to the more pictographic description of 'large difficulties' (Ozolins, 2009).

In addition to these differences, there are nuances, idioms, and significant differences in local culture, understanding of concepts, literacy levels; and other issues (Andriesen, 2008). As well, cultural differences such as that of individualism versus collectivism may play a role in the equivalency of concepts. Although it is generally believed that "materialistic values are common traits that exist in individuals across cultures" (Watchravesringkan, 2012, p. 236), nuances of these values may not be as common as we think. For instance, "[w]hile the United States is an individualistic culture where people view themselves as independent of others, Thailand is a collectivist society where individuals view themselves as interdependent with others and their behaviors are mainly driven by social norms" (Watchravesringkan, 2012, p. 237). Cultural differences such as this may influence the reception of a given value cross-culturally. In China, recent historical and cultural changes have altered certain time-honored values, such as that of the head of the household, generally the man, being the dominant driver of familial behaviors. Whereas the most highly-desirable goods are those which most benefit the family, and family-centeredness remains of prime importance in China, the official policy of the one-child family which began in the early 80s has resulted in "an unprecedented generation of single children in cities, nicknamed the 'little emperors of China' because of the popular belief about the 'imperial' position they occupy within the family. This policy has, in a way, turned the traditional power structure of a Chinese family 'upside down'" (Zhao, 1997, p. 8).

A culturally equivalent research instrument has to take into account the culture's shared norms and socially desirable behaviors, together with shared beliefs, ideas, and assumptions about the world. Moreover, the translated

instrument has to be in line with the local culture's expectations and moral standards, and, of course, it is desirable to consider linguistic appropriateness for the local target population's comprehension, reading skills, and education level. The simplest terms can have significantly low equivalence; for example, intensity words such as 'moderate' and 'frequent' can be understood differently in different cultures, resulting in conceptual lack of equivalence unless such words are carefully analyzed.

Translation may stand as support for any ethnographic research, and translating a research instrument is a critical task. Researchers in different fields need consistency in their research across multiple populations of different languages, and have only recently begun to identify the best practices for translating research instruments into other languages and adapting them for other cultures (Harkness & Schoua-Glusberg, 1998; Harkness, 2003). Standards for translating, and for evaluating translated instruments, are variable across researchers and fields, and, at times, budgets and focus are not dedicated to translation efforts (Harkness, 2003; Rode, 2005; Montoya, 2011). This can result in poorly translated instruments that do not capture the entire plateau of differences between languages and cultures (Harkness, Pennell, & Schoua-Glusberg, 2004; Harkness & Schoua-Glusberg, 1998; Harkness, 2003; Rode, 2005; Montoya, 2011).

b. Translating Research Data

Data collection for the purposes of research can be costly and time consuming, and questionnaire development requires specialized knowledge; therefore, researchers often rely on data collection instruments already in existence. Adopting an available instrument is easier than creating one, and a number of validated instruments in most fields of research are used repeatedly by different researchers. A major problem arises when these instruments need to be used in different countries, requiring the instrument to be translated. The research questionnaires or other documents have been prepared for a population different from the one the researchers may be studying. Researchers use the' ask-the-same-question' model as a means of preserving the properties of the instrument. The "translated items must present the same stimulus as the source questionnaire items and do so by referring to the same entities (abstract and concrete) as do the source items" (Harkness, van der Vijver, & Johnson, 2003, p. 37).

Generally, this results in a verbatim translation of the questions, which is precisely where the problem lies. Different cultures have different cultural schemas and worldviews, different lenses through which they see the world, as well as different labels and descriptions of concepts. A simple back translation is insufficient in numerous cases, including in medical back translation (Rode, 2005; Smith, 2003; Rosdolsky, 2010; Muller et al., 2013). The wording of items within surveys is culturally anchored (Harkness, van der Vijver, & Johnson, 2003). Therefore, some questions may only be appropriate in some cultural contexts, and the change of context may render the scale invalid due to more or less subtle cultural differences. Furthermore, connotations can be lost after translation, and words may have historical connotations in certain countries (Rode, 2005). The context can also have a significant influence on meaning (Smith, 2003).

All these issues can result in using a back-translated instrument that does not measure the intended concept. Back translation is not a sufficient solution to these issues, as it focuses mostly on semantics and less on comprehensibility, naturalness, or connotations (Van der Vijver & Leung, 1997). Current practices of back translation have been criticized as insufficient in cross-cultural research. Only parts of the original meaning can be expressed in different languages (Sumathipala & Murray, 2000). The simple back translation may result in translated items that are too narrow, capturing only a portion of the initial question, or items that have been extended to contain more than the original item.

Sumathipala & Murray (2000) and Harkness (2003) both conclude that one individual producing a translation is far less effective than a group; "A group is better placed to translate, modify, or eliminate inadequate or ambiguous items and generate culturally appropriate translation with semantic and conceptual equivalents" (Sumathipala & Murray, 2000, p. 87). In the committee approach proposed by Harkness (2003), following the first back translation, a group of experts reviews the translation, produces different versions of the instrument, and then collaborates to reach a consensus on which version of the instrument is best. At the end of the process, they pretest the instrument, documenting the issues and problems encountered while providing solutions for future use (Harkness, 2003). The problem with this method is the costly, time consuming process usually avoided in everyday research. Many researchers consider that simple back translations are "often successful enough for the purpose of the research at hand," and "if some of the aforementioned problems occur due to mistakes in translation, there is no immediate way to see them so researchers can pretend they do not exist" (Rode, 2005, p. 17). This path is wrong for research for multiple reasons, as it destroys reputations, it is unethical, and it leads knowledge on a path towards nothing. Rode's suggestion that a rigorous manner and direction must be preserved in research is correct: "No matter how

low the budget for research is, special care should be taken to check the quality of the translated instrument" (Rode, 2005, p. 25).

The present study follows a slightly modified process that was tested and recommended by Harkness (2003). An additional step was added to the process to improve the possibility of identifying not-adapted items. The added validation step was used in conjunction with a different group of bilingual translators, and the results were analyzed statistically. This process resulted in the identification of several items considered to lack conceptual or normative equivalence, which happened in cases where terms or concepts were not possible to translate, when the translation was considered difficult as the exact idiom or concept was not the same in the local language and culture, and when the modification of meaning led to a loss of connotation, making items too narrow or too wide (Sumathipala & Murray, 2000). To identify the items that lacked cultural adaptation, the committee had to fully agree on the quality of each translation; after long discussions, lack of consensus resulted in revealing items that were not translated to every member's satisfaction.

This study also used more translators than the guidelines generally recommend. In general, it is considered that several translators are sufficient for each direction of translation, whereas the present study used more than 100 translators for the two translation directions combined. Their suggestions, notes, observations, and ideas helped to improve final instrument items, and allowed for a more in-depth analysis of each item.

In addition, since an in-depth literature review is considered to be a vital step for cultural adaptation (Gjersing et al., 2010), this study contains a review of literature in the traits and values domain. Each item of a research instrument has to be analyzed separately and assessed as to whether it is equally relevant and acceptable in the target population as in the original target population (Borsa, Damasio, & Bandeira, 2012). Such conceptual and item equivalence cannot be assessed without an in-depth literature review.

Beaton, Bombardier, Guillemin, and Ferraz, (2002), Harkness (2003), and Solano-Flores et al. (2005), as well as institutions such as the U.S. Census bureau, have developed guidelines and translation processes to be used in cross-cultural research to avoid validity problems. The steps, then, the principles of good practice as recommended by Wild et al. (2005), are instrument creation, forward translation, followed by reconciliation, back translation, and expert committee. Usually, recommendations are made to return to the original target population and update, adjusting the adapted instrument after consensus is reached; however, in the present study this was not possible as the original and the target population were different (Gjersing et al., 2010). Rather, this study included the additional validation step described above, in which the research instrument was tested, with the aim of evaluating the understanding and the similarity of each item.

c. Instrument Translation Processes and Validity

The cultural adaptation of research instruments is of vital importance, as is clear from the literature that focuses on medical research (Beaton, et al., 1998; Bullinger, 2003). However, this principle is equally as true when psychological questions are at issue (Borsa, Damasio, & Bandeira, 2012; Watchravesringkan, 2012; McCarthy et al., 2013). Gjersing, Caplehorn and Clausen (2010) used an instrument that was not adapted in Norway, and the instrument failed confirmatory analysis. The authors explained that instruments need to be adapted from the perspectives of language, time, and context. Indeed, considering the importance of cultural adaptation, there is little agreement as to how to achieve this adaptation. Gjersing et al. (2010) emphasized that previously validated instruments do not signify that the instrument is valid in another time, culture, or context. Linguistic translation does not necessarily ensure the validity and reliability of an instrument, and so it is vital that each item is adapted properly (Gjersing et al., 2010).

Back translation has been defined as translating while trying to change as little as possible in the final version (Drennan, Levett, & Swarts, 1991). Back translation has not been not sufficiently studied even though this procedure is ubiquitous in business, technical, and medical instrument translation (Ozolins, 2009). The generic guidelines that exist, for example the World Health Organization (1994) guidelines that insist on a process of forward translation, synthesis, back translation and pilot testing, are challenged by the literature coming from sensitive areas of research, such as cross-cultural psychological studies, including Bullinger's (2003) review of methods and debate over cultural adaptation and Ozolins's (2008) overview of methods describe this criticism. The criticism describes the most common issues, namely grammar discrepancies, translation noise, and cultural issues that can be ultimately solved more or less rigorously, depending on the degree of confidence that authors want to have in the final version of the instrument. This differs across researchers, as some may take translation as a less important step in their research process, and not observe every nuance and variation that can exist or be missed in translation (Ozolins, 2009).

In studying translation, researchers have identified specific issues for cross-culture research. For example, Behling and Law (2000) explored three levels of equivalence and found differences depending on the types of questions asked. They identified different types of questions, such as demographic, behavioral, and knowledge related, and concluded that semantic and conceptual equivalence is relatively easy to achieve, especially in the case of demographic questions where ideas and words are more universal and commonly used. However, they warned that it is more difficult to achieve normative equivalence, as cultures differ radically. For example, even in the case of demographic questions, while semantic and conceptual equivalence can be achieved, cultures differ on their willingness to share personal information. From a normative perspective, equivalence is more difficult to achieve, as abstract concepts and ideas may not be equivalent or equally relevant across cultures (Behling & Law, 2000). Furthermore, there are often no equivalent terms for a given concept, and Western ideas of risk, health, and need, among others, may be viewed differently and be less dominant in other cultures (Hunt & Bhopal, 2004; Watchravesringkan, 2012).

A properly translated instrument must have semantic, conceptual, and normative equivalences to the original instrument. Semantics refers to words and syntax of sentences, an important dimension, but less important than conceptual or normative dimensions (Borsa, Damasio, & Bandeira, 2012). Conceptual equivalence refers to measuring the same concept as the source instrument, a dimension that might force changes in syntax and wording for the sake of moving closer to the original concept (Borsa et al., 2012). Normative equivalence refers to the successful measurement of different social norms across cultures, a dimension that creates further incentives to modify the instrument items for cultural considerations (Borsa et al., 2012). In addition to these dimensions of assessment, other types of assessment were identified, such as experiential equivalence, or whether each item is applicable in the target culture (Borsa et al., 2012). The final choice of item translation to be used has to be made through a consensus among the judges (Gjersing et al., 2010). It was recommended that the entire process to be transcribed and item choice or removal be described to provide a qualitative overview of the process; the present study followed this recommendation by providing the entire set of information related to the translation and cultural adaptation process. Full scripts describing requirements for each stage of the process are available in Beaton, Bombardier, Guillemin, and Ferraz (2002).

Researchers have identified specific problems concerning validity in translation research instruments across cultures. During translation, there are five kinds of validity problems (Bravo, Woodbury-Farina, Canino, & Rubio-Stipec, 1993; Gaviria et al., 1984): content validity, semantic validity, technical validity, criterion validity, and conceptual validity. Content validity refers to the fact that each item assesses concepts that are equally relevant in the investigated culture. Semantic validity refers to the words being used that contain a similar meaning to the original version. Technical validity refers to the achievement of similar effects in different cultures. Criterion validity refers to measuring the same concept. Finally, conceptual validity refers to asking questions that are relevant to the concept construct in both initial and the investigated culture.

3. Research Design

The objective of this study was to adapt two Western materialism scales for use in China, the 18-item Materialism Value Scale MVS (Richins & Dawson, 1992) and the 24-item Materialism Scale (Belk, 1984). To do so, the instrument translation process as recommended by Harkness was followed (Harkness, 2003), improved with the validation step adapted from Montoya, Llopis, and Gilaberte (2011) and an additional test was added using the Median and the Mann-Whitney analyses. This study consisted of six main steps: forward translation, reconciliation, blind back translation, expert committee review, validation, and statistical analysis.

In the first step of the study, the materialism scales were translated into Chinese by Chinese students, from large universities in Shanghai, who had good levels of English. During this forward translation step, Chinese students provided notes, observations regarding confusing items, issues, and suggestions. After the forward translation, the reconciliation step consisted of analysis for each translation, comments and notes, and with the help of two assistants, the assembly of the best synthesized Chinese translation for each item. This step took 2 days; the help of native Chinese assistants was required as the researcher is neither native Chinese, to understand cultural nuances, nor fluent in the Chinese language.

Next, during the blind back translation, other Chinese students, with similarly high English language skills but unfamiliar with the original English version, translated the Chinese version that resulted from the forward translation back into the English language, providing the researcher with another set of notes, observations, issues and suggestions. The intention was to have an in-depth image regarding each item's issues, discrepancies, and translation concerns.

After the back translation, an expert committee, composed of seven members with similarly good English skills, was summoned for a translation review and ample discussion regarding the translation and cultural differences, student notes, observations, and suggestions; this committee discussed each item in both the forward and back translations. Based on absolute consensus, the expert committee chose the best translation option for each item so that it fit best with the Chinese culture, worldview, and life concept. The committee received the materials before the committee meeting, and consensually agreed on each translation; if no consensus was found, then the ambiguous or inadequate items were eliminated. The process whereby equivocal response or disagreement was considered to be failure to reach consensus is similar to the one used by Sumathipala and Murray (2000). Each translation stage (forward and back) took around 40 minutes, and the expert committee discussion took approximately 4 hours.

The items that remained after this committee deliberation were further assessed during an added validation step to evaluate the comparability of the language. Specifically, this involved evaluating the formal similarity of words by looking at the similarity of interpretability; the degree to which the original and translated versions engendered the same response, even if the wording was different; and the degree of understandability – that is, the degree of comprehension of the two versions, even if the wording was different. Two statistical methods were used to identify whether all items could be considered to be fully culturally adapted.

a. Support and Rationale

Consensus building and consensus measurement techniques were central in the cultural adaptation of the research instruments in this study. The committee approach was the means to consensus development, and the validity step was a statistical measurement for consensus (Jones & Hunter, 1995). Furthermore, the selection of committee members was completed in line with Hunter et al.'s (1994) guidelines: members should have very good knowledge of the English language, have expressed interest in the topic, and, further, should represent a variety of backgrounds, to avoid situations where a particular opinion or interest would dominate (Hunter, McKie, Sanderson, & Black, 1994). The researcher of this study participated as facilitator for group meetings but did not participate as a decision maker due to insufficient experience in the Chinese culture.

This study followed a methodology similar to one used in the medical field for the translation of medical instruments used in research (Wild et al., 2005). The lesson learned in the medical world was that, when language issues are not well understood, things can be taken for granted until things go wrong. Unfortunately, in medicine, one error can result in serious consequences and immense costs. Likewise, in business, errors can also result in large costs and unexpected problems. Accordingly, brokering knowledge from the medical world is appropriate to ensure both efficiency and effectiveness during the translation and adaptation of instruments for business research.

The present study followed the best practices of most translating guidelines (Behling & Law, 2000; Solano-Flores & Hurtado, 2005). As suggested by the guidelines, at least two bilingual translators created independent forward translations. Written documentation was kept for each item and for each step of the process. The bilingual translators used in the second translation were unaware of the original version of the instrument and did not have access to the first translation. An expert committee comprised of seven experienced bilingual speakers met to discuss and consolidate the best version of the two translations (Beaton, et al., 2002; Harkness, 2003; Solano-Flores & Hurtado, 2005; United States Census Bureau, 2015). Moreover, the translators were provided with a description of the scope of the project, as indicated by the US Census Bureau guidelines.

The procedure followed during the present study rested on the principle that no one individual should decide anything alone. Therefore, at least two people were involved in most steps. For example, after the translation from English to Chinese, both an assistant and an expert participated in deciding the best translation of each item into Chinese, resulting in a synthesized translated version. The expert used in this stage was a native speaker who had not been involved in the forward translation stage. This is according to Wild et al.'s (2005) description of best practices for the translation and cultural adaptation process. Wild et al. (2005 specified that, in reconciliation stages, it is acceptable to use "an independent native speaker of the target language who had not been involved in any of the forward translations" (p. 99). This study used two independent native speakers - the expert and an assistant. No participant was remunerated for participating in this study, as it was considered that material compensation could impact negatively on the results.

The purpose of the large number of translators selected for this study was to provide a variety of interpretations and views, observations and notes, avoiding the personal interpretations that could be possible if only one or two translators had been used. The Belk (1984) MVS trait scale contains several items that are potentially ambiguous, especially in a Chinese context - Chan and Prendergast (2007) specify the difference, for instance, between the

Western and the Eastern sense of hierarchy and group norms; therefore, this study deemed it preferable to use more translators in order to adjust more accurately to cultural context.

In addition, the back translation method has been criticized for the bluntness of the method, and for its inability to identify subtle errors. The use of many translators, combined with the expert committee and the statistical validation, was intended as a guard against many of the limitations of the back translation method (Harkness & Schoua-Glusberg, 1998; Wild, et al., 2005).

As well, this study aimed to adapt the Western materialism research instruments for use in both urban and rural areas of China. Therefore, the adaptation needed to be appropriate for several reading and educational levels. In this case, the use of many translators with different levels of understanding was meant to reduce demographic limitations (Solano-Flores & Hurtado, 2005).

The use of multiple translators, committee members, and the statistical validation of consensus eliminated additional potential issues signaled by other studies (Gjersing et al., 2010), such as personal idiosyncrasies added by a limited number of translators, or a rushed back translation in which the researcher might have total decision power. It was necessary to remove the idiosyncrasies from the source instruments, as well as the outdated items, as the instruments were created in the 80s and 90s, and new settings due to changes in society over time had to be taken into consideration (Gjersing et al., 2010).

b. Methods

This study adapted the methodology of Harkness (2003) for translations. In this approach, in addition to forward translation, synthesis, and back translation, a group of experts reviewed translations, produced different versions of the instrument, and then collaborated to reach a consensus regarding which version of the instrument was best; at the end of the process, they pretested the instrument and documented the issues and problems encountered, while providing solutions for future use (Harkness, 2003). The back translation was performed by a large number of students, and the expert committee collaborated to reach a consensus regarding the best version for each item.

During the first stage, seventy-four Chinese college students translated the questionnaire from English to Chinese. The students answered the form, *Scale Translation Form ENG to CN*, and translated the original items from English into the Chinese language. They were asked to make notes wherever the items, phrasing, or the concepts were confusing. After that, the synthesis of the translations was done with the help of a Chinese University faculty member and a Chinese assistant, in order to choose the most common translations to create the first Chinese questionnaire. During the back translation step, forty-four college students from another university translated the synthesized questionnaire from Chinese back into English. These were different students from a different university, and they had no access to the original version of the questionnaires. This was a double blind translation, as the students did not see the original English version of the items. Moreover, each stage was performed with different groups of students from different universities to assure the reliability of translations. During this stage, the students used the *Scale Translation Form CN to Eng*. The two translations led to a number of discrepancies, as some items confused the translators; some items were difficult to understand conceptually and were difficult to translate into Chinese; and other items were not offering equal choices (e.g., everybody would have chosen only one option).

After translation, an expert committee was summoned to decide the best translation for each item. The committee method is one of the most widely used methods in cross-cultural validation of instruments (Harkness, 2003). For this stage, the expert committee was composed of faculty members from two large public universities in Shanghai, all of them native Chinese with excellent English language skills. The committee members were seven Chinese natives with an average age of thirty, all working in higher education, all college graduates.

The committee's objective was to review the generated scale translation and choose the best possible translation adapted for Chinese customs and culture. For each questionnaire item, there were seventy-four choices from English to Chinese, and forty-four translations from Chinese to English. The committee members were given lists with the most common translations for each item. The session took four hours. The agenda of the meeting had three points: (a) discuss problematic items, define, and choose the best translation for each; (b) discuss and answer any questions regarding any item; and (c) review all items and agree on a final translation.

The condition for accepting translated items in the present study was full consensus between the committee members, and the committee members were asked to explain whether they fully agree or fully disagree with the resulted translation. Disagreements were discussed and alternative translations were sought. Translations failing to achieve consensus were discussed in depth, and the initial translations were again scrutinized for these items.

A number of questions emerged from the translation stage, as many translators pointed towards confusing items, issues in equivalence of translations, and unclear translations. First, a number of expressions seemed more difficult to translate and comprehend, given the structure and rules of the Chinese language. English idioms or expressions and even words were problematic during translation. For example, "pay much attention" (6), "all that important" (9), "don't mind" (32), "having luxury in one's life" (12), "wouldn't be any happier" (16), and "appealing" (19) caused difficulties for committee members.

Furthermore, some items raised questions regarding the meaning of the item in English and were problematic for translation into Chinese. For example, Item 25: "I worry about people taking my possessions" uses the word "taking," while the meaning may be "stealing." Similar wording confusion exists between Item 21 and Item 22: "I get very upset if something is stolen from me, even if it has little monetary value" and "I don't get particularly upset when I lose things" (stolen versus lost, respectively). Moreover, Item 33: "I don't like to have anyone in my home when I'm not there" confused both translators and committee members as to whether the question asks about friends or strangers. The questions had to be as specific as possible to give no room for interpretation or discussion. Therefore, it was decided that items should be adapted towards a simpler, less confusing phrasing than the English version. This recommendation is reflected in the final translation of the items, in the way that most items were simplified for easier and better understanding. This change may reduce the comparability of language between the original and the translated version, but this approach was necessary to increase the similarity of interpretability and the degree of understandability.

The committee members received a form in which each original item was described in English, together with the synthesized version in Chinese as it had been agreed upon during the reconciliation, as in the following example:

1.   I admire people who own expensive homes, cars, and clothes

我羡慕那些拥有豪宅，名车，华美衣服的人.

They also received the detailed translations from both stages (Eng to Cn and Cn to Eng). The discussion looked at four criteria: (a) the appropriateness of statements for China, (b) the comparability of language, (c) the similarity of interpretability, and (d) the degree of understandability. During the meeting, some items were seen as forcing the respondents towards only one possible answer. Some items were slightly modified for better understanding in the Chinese language.

c. The Median Test

The Median test was used in this study to validate the survey items that had been agreed upon following the translation and cultural adaptation process (Tabachnick & Fidell, 2007). The median test determines whether the items have equal medians according to these hypotheses: H0: The k items have equal medians; H1: The k items do not have equal medians. The median test's methodology follows these steps: First, determine the common median; second, for each group count the number of observations that lie above and below the common median; third, split each of the k groups of scores at the combine median, and cast the resulting frequencies in $k{\times}2$ table (see Figure 2).

Test statistic. Using the $k{\times}2$ table, compute the value of the test statistic:

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} (O_{ij} - E_{ij})^2 / E_{ij} \qquad \sim \chi_{(k-1)}^2$$

Where $E_{ij} = (j^{th} \text{ row total} \times i^{th} \text{ column total}) / (\text{Grand Total})$

$j =1,2…k$ / $i = 1,2$	O_{ij}				i^{th} row total
	$j=1$	$j=2$		$j=k$	
Below	O_{11}			O_{12}	1st row total = $(O_{11} + O_{12})$
Above	O_{21}			O_{22}	2nd row total = $(O_{21} + O_{22})$
j^{th} column total	1st column total = $O_{11} + O_{21} + …+ O_{k1}$			k^{th} column total = $O_{12} + O_{22} +…+ O_{k2}$	Grand Total

Figure 1. The Median Test

The χ^2 table is used to find the critical value at 5% significance level to test the hypothesis. If the calculated value is greater than the critical value, then H0 (i.e., k items have equal medians) is rejected. Furthermore, at least one item has a different median than others. If the calculated value is lower than the critical value at 5% significance level, H0 is accepted (i.e., k items have equal medians). If the median test is significant (calculated > critical value at 5% significance level), then the item that consists of the most number of 'poor' observations is removed, and the Median test is repeated on the rest of the items until the H0 is accepted.

d. The Mann Whitney Test

The Mann Whitney test can be used to check whether two groups are identical. This powerful nonparametric test is an alternative to the parametric t test when the sample size is low or t-test assumptions need to be avoided. This test encompasses the following hypotheses: Suppose we have two groups, A and B; H0: There are no differences between the responses of the two groups; H1: There are differences between the responses of the two groups. The Mann Whitney test's methodology follows these steps: First, determine the sample sizes of the two groups (e.g., let n = sample size of Group A and m = sample size of Group B); second, pool the data and rank together the scores of both groups, assigning the rank of one to the score which is algebraically lowest.

Test statistics. The test statistic U is given by focusing on one of the groups,

$U\_1 = nm + (n+1)n/2 - R\_1$ Where R1 – sum of the ranks assigned to the group of the size n

$U\_2 = nm + (m+1)m/2 - R\_2$ Where R2 – sum of the ranks assigned to the group of the size m

The smaller of the two U values should be considered as test statistics.

Decision rule. If the smaller $U \leq$ critical value at 5% significance level, then the null hypothesis is rejected. When sample sizes (n, m) increase in size, the sampling distribution of U rapidly approaches the normal distribution with mean $\mu = nm/2$ and variance $\sigma2 = (nm(n+m+1))/12$ then the test statistic $Z = (U-\mu)/\sigma \sim N(0,1)$ can be used to test the hypothesis. If the p value of the Z is less than 0.05, then the null hypothesis is rejected at 5% significance level (i.e., the compared item is significantly different than Item 1 for that particular characteristic). Note: If the p value of the Z is less than 0.1, then the null hypothesis is rejected at 10% significance level.

4. Details of Analysis and Results

During the first translation from English to Chinese, the seventy-four translators had to read and translate each item from the questionnaire. They were given ample space for notes and observations and were asked to explain whether the English item made sense for them, to avoid items that led all respondents to only one choice. The respondents were also asked to explain confusing items, reasons for confusion, and potential interpretations and translations into Chinese. The notes were compiled during reconciliation with the help of two native Chinese people who were not involved in the forward translation. After reconciliation, the best translation was synthesized into the first version of the questionnaire in Chinese. This synthesized version was used in the back translation into English.

a. Translations from Chinese to English

After deciding the most frequent, best translation from English to Chinese, a new form was compiled. A group of forty-three translators translated the new Chinese questionnaire into English. This was a double-blind translation, as the students translating the new Chinese questionnaire into English did not see the original English version of the items that resulted in the development of the new Chinese questionnaire. Moreover, each stage was performed with groups of students from different universities to assure the reliability of translations, and to conform to the requirements of a blind back translation. In translating from Chinese back to English, there were significantly fewer issues than in the earlier English to Chinese translation.

b. Expert Committee

After the back translations were completed, a group of experts reviewed these translations and collaborated to reach a consensus regarding which version of the instrument was best. This review happened in a session that took approximately four hours. The committee members received all material in advance and were asked to bring their notes at the meeting. Because of their interest in the topic, the experts spent many hours reading the materials and preparing for the meeting. The agenda of the meeting had three points: (a) discuss all items, define, and agree on each translation; (b) brainstorm, ask and answer any questions regarding any item; and (c) review all problematic items. The following issues emerged from the discussion: First, a number of expressions seemed more difficult to translate and comprehend, given the rigor/rigidity of the Chinese language. Expressions such as "pay much attention" (6), "all that important" (9), "don't mind" (32), "having luxury in one's life" (12), "wouldn't be any happier" (16), and "appealing" (19) posed concerns for committee members. Second, some items raised questions

regarding meaning in English, as previously discussed. The committee reached a consensus for each item analyzed. The problematic items where consensus for translation was not reached were considered inadequate.

c. Translation Validation

After the committee meeting, forty-four students not involved in the earlier translation processes evaluated the original English items and the resulting Chinese translation in terms of equivalence. The criteria used included the following: Comparability of language, which refers to formal similarity of words; similarity of interpretability, which refers to the degree to which the original and the translated versions engender the same response, even if the wording is different; and degree of understandability, which refers to the degree of comprehension of the two versions, even if the wording is different.

Participants were asked to evaluate the translated items, and the validation form presented each item in both the original version and back-translated version. The scale used for the evaluation of each item was as follows: (1) *extremely comparable, similar, or understandable*; (4) *moderately comparable, similar, or understandable*; and (7) *not at all comparable, similar, or understandable*.

Typically, the middle value should be neutral, such as *don't know/not sure/no opinion*; however, the researcher intended to use the validation methodology as proposed by Montoya, Llopis and Gilaberte (2011), avoiding the potential for central tendency bias where many respondents choose the neutral, middle value, regardless of how they think. In addition, especially for the validation stage, the intention was to use the respondents' responses in a meaningful way; accordingly, the decision was made to drop the midpoint for this stage. As this stage was testing the level of quality of a translation, it made sense to eliminate the middle value and look in more depth at quality by adding a third positive. In addition, if the *strongly disagree* to *strongly agree* scale is a continuum theoretically from -1 to +1, in the case of this study (and in general in cases of quality evaluation, as this study concerns a quality evaluation scale), the scale is a continuum from 0 to 1: It is good (and how good it is, on various levels) or not. There is no equivalent of *strongly disagree* or -1. *Poor* or *very poor* is the equivalent of 0, unsatisfactory quality, while excellent is the +1. Forty-four respondents, all Chinese students studying English language at the university level, completed the validation form. The 32 initial items were numbered in order.

d. Validation Methodology

The quality of each item is scaled using five categories: poor, fair, good, very good, and excellent. These categories are assigned values to aid in analysis. The categories are as follows:

Poor – 1, fair – 2, good – 3, very good – 4, excellent – 5

Step 1 - All k items are separately compared for three characteristics using the Median test (i.e. comparability of language, similarity of interpretability and degree of understandability of each item).

Step 2 - a) If the Median test is significant (the null hypothesis is rejected) for a particular characteristic, step 1 is repeated by removing the item which has the most observations below a score of 3 (i.e. the item that has more 'poor' and 'fair' observations.)

   b) If the Median test is not significant; Step 3 is repeated by removing the significant items from Step 1.

Step 3 - Each remaining item is compared using the Mann Whitney test. These are compared with Item 1, which has no 'poor' or 'fair' observations. Since the number of observations are larger than 20, Z statistic can be used to test the hypothesis.

If an item concerned is significant (different) in a particular characteristic when compared to Item 1, then that item is marked as such.

The Median test results for each characteristic are presented in Figure 3:

$K=32$	Comparability of Language	Similarity of interpretability	Degree of understandability
$\sum_{i=1}^{2}\sum_{=1}^{k}\frac{\left(O-E\right)^2}{E}$	237.6	200.7	237.8
χ^2_{-1} at 5% significance level	44.98534	44.98534	44.98534
Result of hypothesis test	Reject H0	Reject H0	Reject H0

Figure 2. Median Test Results

The Median test was considered significant at 5% significance level. The test was repeated on the remaining data set after removing Item 21, which had the highest number of 'poor' and 'fair' observations for all three characteristics.

$K=31$	Comparability of Language	Similarity of interpretability	Degree of understandability
$\sum_{i=1}^{2}\sum_{=1}^{k}\frac{\left(O-E\right)^2}{E}$	63.33783	65.17773	51.00565
χ^2_{-1} at 5% significance level	43.773	43.773	43.773
Result of hypothesis test	Reject H0	Reject H0	Reject H0

Figure 3. Median Test Results, First Repeat

The Median test was again considered significant at 5% significance level. The test was repeated again after removing Item 23, which has the next highest number of 'poor' and 'fair' observations for all three characteristics.

K=30	Comparability of Language	Similarity of interpretability	Degree of understandability
$\sum_{i=1}^{2}\sum_{=1}^{k}\frac{\left(O-E\right)^2}{E}$	34.09228	38.70629	24.37522
χ^2_{-1} at 5% significance level	42.55697	42.55697	42.55697
Result of Hypothesis test	Do not Reject H0	Do not Reject H0	Do not Reject H0

Figure 4. Median Test, Second Repeat

The Median test concluded that, apart from Items 21 and 23, other items have similar medians (i.e. the other 30 items are not significantly different from the median of 3). This observation is based on the number of observations for comparability of language, similarity of interpretability and degree of understandability. Items 21 and 23 are significantly different and therefore not fully culturally adapted.

The Mann Whitney Test was performed to compare the remaining items with three separate characteristics of item 1. This test was performed using the MINITAB statistical package. The following results were obtained: if at least one characteristic is significantly different from Item 1, that item is considered not culturally adapted, and therefore not suitable to use for collection of information. As a result, Items 4, 8, 14, 15, 17, 18, 20, 24, and 32 were considered significantly different and not culturally adapted.

e. Validation Discussion

Using both non-parametric test results, Items 4, 8, 14, 21, and 23 were significantly different than their original version. This result is based on the comparison done between the comparability of language, similarity of interpretability, and degree of understandability of translated version. Comparability of the language was significantly different in Items 15, 24 and 32, but other two characteristics were not considered significant at 5% significance level. Items 20 and 18 were significantly different compared to the similarity of interpretability, but not when compared for the other two characteristics. Item 17 was not significant for similarity of interpretability but was significant when compared to the other two characteristics.

Translated items (Items 4, 8, 14, 15, 17, 18, 20, 21, 23, 24, and 32), with a 5% significance level (95% confidence), cannot be validated for use in the survey. This is relatively consistent with the results. However, the results do permit the final questionnaire to contain items with several "poor" evaluations in all characteristics, with a 5% significance level (i.e. Items 11 and 13). This was believed to have the potential of skewing the results. Therefore, the Mann Whitney test was repeated with a 10% significance level.

Other than rejected items at a 5% significance level, the following items are significantly different at a 10% significance level compared to Item 1. Those are Items 9, 11, 12, 13, and 29.

When 10% significance level (90% confidence) was considered for statistical comparison, the items that were found to be not culturally adapted could not be validated for use in the survey.

5. Conclusion

During the translation and validation stages, the major take away was the importance of an in-depth translation and validation. The simple back translation does not work in the Chinese culture, which is so different from Western cultures. Beyond cultural differences, language poses multiple difficulties, as the culture reflects a different system of thinking that exists in China. For instance, the rigidity of the Chinese structure made literal translations almost impossible. The visual ways of the Chinese language found "expensive homes" translated as "luxurious" and expensive clothes as "fine" or "designer clothes." Moreover, verbs can express different meanings, such as the verb "羡慕[xiànmù]," which means both "admire" and "envy."

Furthermore, buying can mean shopping 购物 [gòuwù] or paying 买东西 [mǎi dōng xī], and pleasure can be expressed as "fun," "delight" 乐趣 [lèqù], or "happiness" 快乐 [kuàilè]. Each of these terms has uses in certain contexts, and missing such nuances can result in a poorly designed question or survey and, ultimately, in poor findings and conclusions.

A different type of confusion related to questions that were contradicting or asking obvious things from a Chinese cultural perspective. For example, a question asked to what extent "acquiring material possessions" is important. This acquisition is key in the Chinese culture, and so everyone will say it is important. This response does not necessarily reflect materialism, but rather cultural aspects of family life and face. Another example, the item "I have all the things I need to enjoy life," is interpreted this way: Things one really needs are not used for enjoying life but for satisfying basic needs. When this point was discussed during the English-to-Chinese-language translation stage, very few translators seemed to believe that having all the things one needs to enjoy life is ever possible.

An important recommendation emerging from this study regards the practices of back translation for research instruments, which is insufficient when used in cross-cultural contexts when the cultural distance is significant, as it is between the Western cultures for which instruments were first developed and the Chinese culture. Most research instruments must be culturally adapted for China in a process of translation and cultural adaptation as portrayed here. Subtle cultural details cannot be uncovered by a simple back translation, but a problem exists with this method in that it is a costly, time consuming process usually avoided in everyday research. Because of this obstacle, many researchers consider that simple back translations are "often successful enough for the purpose of

the research at hand" (Rode, 2005, p. 17), which is contradicted by the present study. Rode's (2005) suggestion that a rigorous manner and direction must be preserved in research is correct and supported by the current study, and indeed, special care should be taken to check the quality of the translated instrument.

Studies by Wallendorf and Arnould (1988) and Rudmin (1988) advocated that Belk's (1984) materialism scale is more suitable for the United States' culture than for other cultures, particularly more suitable than for those of the so-called "Third World" (Rudmin, 1988; Wallendorf & Arnould, 1988). Ger (1990) attempted to remedy this situation, but the conclusion was that, "It is clear that despite the attempt to construct a cross-culturally reliable materialism scale, the resulting scale is more reliable in the United States and Europe than in Turkey" (Ger, 1990). He argued that this does not mean the scale can't be used, as he thought the materialistic consumer culture is being increasingly emulated around the world. The present study contradicts this theory and describes that the cultural adaptation process of a Western instrument can result in significant modifications of the original instrument, requiring revalidation and reliability analysis.

If the concepts are significantly differently interpreted and viewed across different cultures, there is a need to improve the research instrument. Creating equivalent items might be impossible, but researchers can follow the indications of Sumathipala and Murray (2000) and remove those items that are inappropriate after translation. Elimination of items can also be based on item format or duplication, criteria used by Farmer and Sundberg (1986).

Another possible solution could be creating research instruments that are designed specifically for use in China. These instruments could be shortened versions of Western instruments, where inadequate statements would be removed. The idiosyncrasies from the source instruments must also be removed, as well as the outdated items, as many research instruments were created in the 80s and 90s, and new settings due to changes in society over time have to be taken into consideration (Gjersing et al., 2010).

Elimination of items could be performed based on feedback, literature review, and ultimately by expert committee decision-making. Certainly, previous research has employed significant modifications, such as adding or removing items. As this study has shown, the newly resulted instruments have to be evaluated and checked for validity and reliability, but in some previous research this step has been overlooked; as well, no criteria were defined for reaching item elimination or addition decisions, and every decision was based solely on the subjective judgment of the researcher and group discussions (Gjersing et al., 2010). This did not occur in the present study, where all guidelines were strictly followed and additional steps were created for the sake of reliably adapting the two materialism scales. Furthermore, the in-depth information provided in this study will act as support for further efforts in this direction of adapting and using better research instruments.

Very little has been written regarding how a scale could be shortened. From Richins' (2004) examination of forty-four studies, thirteen of those were found to have used shortened versions of the materialism (MVS) scale, and ten of these studies used an ad-hoc way to shorten the scale (Richins, 2004). Most of the studies analyzed only used six or seven items, chosen entirely at the latitude and decision of the researcher. Therefore, Richins identified a need for a shorter materialism scale and evaluated the original scale of materialism, identifying best and worst performing items. The worst items were 6, 7, 10 and 13, and the present study also identified items 7, 10 and 13 from the Richins MVS scale as being difficult to culturally adapt. Items 1 and 5 were found to perform best in terms of validity, and these items were successfully validated in the present study. Comparing the results of Richins MVS item analysis with the present study's evaluation of translations, in most cases poorer items were identified during the present study's translation and validation stages, and items that were kept were also strong in Richins' study. In addition, the shorter versions of 15, 9, 6 and 3 items were all found reliable and valid, as was the result in the Richins study's statistical analysis, and all original and shortened scales were found to correlate at analogous scale levels with the Belk scale (Richins, 2004). Therefore, the combination of all the translated and culturally adapted items from both scales could provide useful results, but further investigation is required before this conclusion can be drawn with confidence.

There is a positive relationship between the two materialism instruments, as the Richins MVS scale was found to correlate with Belk's instrument of materialism (Richins, 2004). Therefore, the two scales could be expected to measure materialism from complementary perspectives. In addition, there are simplified versions of each instrument, each of these versions measuring materialism with certain validity and reliability. Therefore, it can be assumed that the two scales overlap, and it might make sense to merge the culturally adapted items into one instrument and attempt to identify whether it can provide any useful result.

References

Andriesen, S. (2008). Benefiting from back translations. Retrieved from http://www.medilingua.com/pdf/BackTranslationsICTSummer%202008.pdf

Beaton, D., Bombardier, C., Guillemin, F., & Bosi Ferraz, M. (1998). *Recommendations for the cross-cultural adaptation of health status measures.* Rosemont, IL: American Academy of Orthopaedic Surgeons Institute for Work & Health. Retrieved from http://www.ortho.umn.edu/prod/groups/med/@pub/@med/@ortho/documents/asset/med_asset_360072.pdf

Behling, O., & Law, K. (2000). *Translating questionnaires and other research instruments: Problems and solutions.* Thousand Oaks, CA: Sage.

Belk, R. (1984). Three scales to measure constructs related to materialism: Reliability, validity, and relationships to measures of happiness. (1984). *Advances in Consumer Research, 11*(1), 291-297. Retrieved from http://www.acrwebsite.org/search/view-conference-proceedings.aspx?Id=6260

Borsa, J., Damasio, B., & Bandeira, D. (2012). Cross-cultural adaptation and validation of psychological instruments: Some considerations. *Paidéia (Ribeirão Preto),* 22(53), n.p. Retrieved from http://www.scielo.br/scielo.php?pid=S0103-863X2012000300014&script=sci_arttext&tlng=en

Bravo, M., Woodbury-Fariña, M., Canino, G. J., & Rubio-Stipec, M. (1993). The Spanish translation and cultural adaptation of the Diagnostic Interview Schedule for Children (DISC) in Puerto Rico. *Culture and Medical Psychiatry, 17*(3), 329-344. https://doi.org/10.1007/BF01380008

Bullinger, M. (2003). International comparability of health interview surveys: An overview of methods and approaches. In A. Nosikov & C. Gudex (Eds.), *EUROHIS: Developing Common Instruments for Health Surveys* (pp. 1-12). Amsterdam: IOS Press.

Canino, G., Bird, H., Rubio-Stipec, M., & Bravo, M. (1997). The epidemiology of mental disorders in adult population in Puerto Rico. *Puerto Rico Health Science Journal, 16*(2), 117-124.

Cavusgil, S.T. (1997). Measuring the potential of emerging markets: An indexing approach. *Business Horizons*, *40*, 87-91.

Chan, K., & Prendergast, G. (2007). Materialism and social comparison among adolescents. *Social Behavior and Personality*, *35,* 213-228. https://doi.org/10.2224/sbp.2007.35.2.213

Doctoroff, T. (2005). *Billions: Selling to the new Chinese consumer.* New York, NY: Palgrave Macmillan.

Drennan, G., Levett, A., & Swarts, L. (1991). Hidden dimension of power and resistance in the translation process: A South African study. *Culture and Medical Psychiatry, 15*(3), 361-381. Retrieved from http://link.springer.com/journal/11013/15/3/page/1

Farmer, R., & Sundberg, N. (1986). Boredom proneness: The development and correlates of a new scale. *Journal of Personality Assessment*, *50*(1), 4-17. https://doi.org/10.1207/s15327752jpa5001_2

Gaviria, F. M., Pathak, D. S., Flaherty, J., Pacheco, C. G., Martinez, H., Wintrob, R., & Mitchell, T. (1984). *Designing and adapting instruments for a cross cultural study on immigration and mental health in Peru.* Washington, DC: American Psychiatric Association.

Ger, G. (1990). Measuring and comparing materialism cross-culturally. *Advances in Consumer Research, 17*, 186-192. Retrieved from http://www.acrwebsite.org/search/view-conference-proceedings.aspx?Id=7019

Gjersing, L., Caplehorn, J., & Clausen, T. (2010). Cross-cultural adaptation of research instruments: Language, setting, time and statistical considerations. *BMC Medical Research Methodology*, *10*(13), n.p. https://doi.org/10.1186/1471-2288-10-13

Hanzaee, K.H., & Adibifard, S. (2012). Development and validation of native material values scale for Iran (NMVS-I). *Research Journal of Applied Sciences, Engineering and Technology, 4*(14), 2140-2146. Retrieved from http://maxwellsci.com/print/rjaset/v4-2140-2146.pdf

Harkness, J. (2003). Questionnaire translation. In J. Harkness, F. Van de Vijer, & P. Moher (Eds.), *Cross cultural survey methods* (pp. 35-56). Hoboken, NJ: John Wiley & Sons.

Harkness, J., & Schoua-Glusberg, A. (1998). Questionnaires in translation. *ZUMA-Nachrichten Spezial.* 87-126. Retrieved from http://isites.harvard.edu/fs/docs/icb.topic506406.files/znspez3_04_Harkness_Glusberg.pdf

Harkness, J., Pennell, B., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Presser, J. Rothgeb, & M. Couper (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 453-473). Hoboken, N.J.: John Wiley & Sons.

Harkness, J., van der Vijver, F., & Johnson, T. (2003). Questionnaire design in comparative research. In J. Harkness, F. Van de Vijer, & P. Moher (Eds.), *Cross cultural survey methods*. Hoboken, NJ: John Wiley & Sons.

Hunt, S., & Bhopal, R. (2004). Self report in clinical and epidemiological studies with non-English speakers: The challenge of language and culture. *Journal of Epidemiology and Community Health*, *58*(7), 618-622. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1732826/

Hunter, D. J., McKee, C. M., Sanderson, C. F., & Black, N. A. (1994). Appropriate indications for prostatectomy in the UK: Results of a consensus panel. *Journal of Epidemiology and Community Health*, *48*(1), 58-64. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1059896/

Jones, J., & Hunter, D. (1995). Consensus methods for medical and health services research. *BMJ*, *311*, 376-380. https://doi.org/10.1136/bmj.311.7001.376

McCarthy, S., Jafaar, J., Kamal, A., & Zubair, A. (Eds.). (2013). *Psychology at work in Asia: Proceeds of the 3rd and 4th Asian Psychological Association conferences and the 4th international conference on organizational psychology*. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Montoya, A., Llopis, N. & Gilaberte, I. (2011). Validation of the translation of an instrument to measure reliability of written information on treatment choices: a study on attention deficit/hyperactivity disorder (ADHD). *US National Library of Medicine National Institutes of Health*, *24*(3), 577, PMID: 22267353. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22267353

Müller, A., Smits, D.J.M., Claes, L., Gefeller, O., Hinz, A., & de Zwan, M. (2013). The German version of the Material Values Scale. *Psychosoc Med, 10*(5). https://doi.org/10.3205/psm000095

Nisbett, R. (2003). *The geography of thought : how Asians and Westerners think differently...and why*. New York, NY: Free Press.

Opree, S.J., Buijzen, M., van Reijmersdal, E.A., Valkenburg, P.M. (2011, June 23). *Development and validation of the Material Values Scale for Children*. Paper presented at the 10th International Conference on Research in Advertising, Berlin. Retrieved from http://dare.uva.nl/document/2/122159

Ozolins, U. (2008). *Issues of back translation methodology in medical translations*. Shanghai: FIT, International Federation of Translators.

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

Podoshen, J.S., Li, L., & Zhang, J. (2011). Materialism and conspicuous consumption in China: a cross-cultural examination. *International Journal of Consumer Studies, 35,* 17-25. https://doi.org/10.1111/j.1470-6431.2010.00930.x

Reinecke Flynn, L., Goldsmith, R.E., & Kim, W-M. (2013). A cross-cultural study of materialism and brand engagement. *Journal of Multidisciplinary Research, 5*(3), 49-69.

Richins, M., & Dawson, S. (1992). A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of Consumer Research*, *19*(3), 303-316. https://doi.org/10.1086/209304

Richins, M. L. (2004). The Material Values Scale: Measurement properties and development of a short form. *Journal of Consumer Research, 31*(1), 209-219. https://doi.org/10.1086/383436

Rode, N. (2005). Translation of measurement instruments and their reliability: An example of job-related affective well-being scale. *Advances in Methodology and Statistics, 2*(1), 15-26. Retrieved from http://ams.sisplet.org/

Rosdolsky, M. (2010). Back translation of medical documents with English-German examples. *Caduceus,* 12-14. Retrieved from http://www.ata-divisions.org/MD/caduceus/spring-summer-2010-part1.pdf

Rudmin, F. (1988). *Ownership as interpersonal dominance: a history and three studies of the social psychology of property* (Doctoral thesis). Queen's University, Ontario, Canada.

Smith, T. (2003). Developing comparable questions in cross-national surveys. In J. Harkness, F. van der Vijver, & P. Mohler (Eds.), *Cross-cultural survey methods*. Hoboken, NJ: John Wiley & Sons.

Solano-Flores, W., & Hurtado, M.. (2005). CAHPS Guidelines for Assessing and Selecting Translators and Reviewers. Retrieved from https://cahps.ahrq.gov/surveys-guidance/helpful-resources/resources/CAHPS-Translatorselection-guidelines-106051.pdf

Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology, 126*(1), 124-128. https://doi.org/10.1053/j.gastro.2003.10.016

Sumathipala, A., & Murray, J. (2000). New approach to translating instruments for cross-cultural research: a combined qualitative and quantitative approach for translation and consensus generation. *International Journal of Methods in Psychiatric Research, 9(*2), 87-95. https://doi.org/10.1002/mpr.83

Tabachnick, B.G. & Fidell, L.S. (2015). *Using multivariate statistics* (6th ed.). Boston: Pearson Education.

Trinh, V.D., & Phau, I. (2012, August 30). *A new set of measurements for the materialism scale.* Paper presented at 2012 Australian and New Zealand Marketing Academy (ANZMAC). Annual Conference, University of South Australia, Australia. Retrieved from http://www.anzmac.org/conference_archive/2012/papers/403ANZMACFINAL.pdf

United States Census Bureau. (2015). *American community survey 5 year data (2010-2014).* Powell, L.T. (author). Retrieved from http://www.census.gov/data/developers/data-sets/acs-survey-5-year-data.html

Van der Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research.* London: Sage.

Wallendorf, M., & Arnould, E. (1988). My favorite things: A cross-cultural inquiry into object attachment, possessiveness and social linkage. *Journal of Consumer Research*, *14*(4), 531-547. https://doi.org/10.1086/209134

Watchravesringkan, K. (2012). Cross-cultural equivalence of materialistic values scale (MVS): An exploratory study between the United States and Thailand. *Journal of Targeting, Measurement and Analysis for Marketing*, *20*, 235-253. https://doi.org/10.1057/jt.2012.19

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient reported outcomes (pro) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*, *8*(2), 94-104.

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? Case of the Material Values Scale. *Journal of Consumer Research, 30*, 72-91.

Zarco, T.H. (2014). The influence of materialism on consumer preferences: A conjoint analysis approach. *Philippine Management Review, 21*, 39-52.

Zhao, B. (1997). Consumerism, Confucianism, communism: Making sense of China today. *New Left Review, 222*, 43-59. Retrieved from http://newleftreview.org/I/222/bin-zhao-consumerism-confucianism-communism-making-sense-of-china-today

Copyrights