Psychological Assessment
2016, Vol. 28, No. 6, 750–764

© 2015 American Psychological Association
1040-3590/16/$12.00 http://dx.doi.org/10.1037/pas0000213

Text Mining a Self-Report Back-Translation

Angel Blanch and Anton Aluja
University of Lleida and Institute of Biomedical Research (IRB) Lleida, Spain

There are several recommendations about the routine to undertake when back translating self-report instruments in cross-cultural research. However, text mining methods have been generally ignored within this field. This work describes a text mining innovative application useful to adapt a personality questionnaire to 12 different languages. The method is divided in 3 different stages, a descriptive analysis of the available back-translated instrument versions, a dissimilarity assessment between the source language instrument and the 12 back-translations, and an item assessment of item meaning equivalence. The suggested method contributes to improve the back-translation process of self-report instruments for cross-cultural research in 2 significant intertwined ways. First, it defines a systematic approach to the back translation issue, allowing for a more orderly and informed evaluation concerning the equivalence of different versions of the same instrument in different languages. Second, it provides more accurate instrument back-translations, which has direct implications for the reliability and validity of the instrument's test scores when used in different cultures/languages. In addition, this procedure can be extended to the back-translation of self-reports measuring psychological constructs in clinical assessment. Future research works could refine the suggested methodology and use additional available text mining tools.

Keywords: text mining, back-translation, cross-cultural psychology

Supplemental materials: http://dx.doi.org/10.1037/pas0000213.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The adaptation of self-report assessment instruments for cross-cultural research has been applied to a variety of psychological constructs within fields such as learning (Bornman, Sevcik, Romski, & Pae, 2010), emotions (Kuppens, Realo, & Diener, 2008; Moscoso & Spielberger, 2011), or well-being (Blanch & Aluja, 2009a; Fischer & Boer, 2011). However, human personality is by far the most studied psychological construct between different cultures. The study of personality across cultures has been addressed from several approaches, even though the quest of universal traits as mental qualities shared by members of different cultures has probably guided the most considerable and significant research effort (Allik & McCrae, 2004; Barrett, Petrides, Eysenck, & Eysenck, 1998; Benet-Martínez & John, 1998; McCrae & Costa, 1997; Schmitt, Realo, Voracek, & Allik, 2008). Most, if not all, cross-cultural studies about personality make use of questionnaire self-report measures designed to gauge analogue personality dimensions across different cultures. However, it has been suggested that relying solely on self-reports bears several methodological shortcomings when measuring personality or akin psychological dimensions (Church, 2001; Heine, Lehman, Peng, & Greenholtz, 2002; Norenzayan & Heine, 2005; Peng, Nisbett, & Wong, 1997). These methodological challenges stem from the adaptation and translation of personality questionnaires into different languages, because written language conveys diverse concepts and meanings subject to substantial variations across cultures (Geisinger, 1994; Spielberger, 2006).

The availability of reliable and valid psychological instruments is a crucial issue in cross-cultural research. Indeed several works propose a number of strategies dealing with the adaptation of assessment instruments to different languages/cultures to obtain a commensurate form of the target instrument. To the best of our knowledge, however, none of these studies has considered an intensive use of text mining methodology when addressing this endeavor. The current work aims to fill this gap by suggesting a feasible text mining procedure during the adaptation of a self-report instrument in English language to different foreign languages.

Personality Measurement Across Different Cultures

There are four different perspectives to addressing the study of personality in different cultures: evolutionary, cross-cultural, indigenous, and cultural psychology perspectives. Indigenous and cultural psychology perspectives investigate constructs thought to reflecting socially built circumstances by the adoption of a culture relativistic view. On the other hand, evolutionary and cross-cultural perspectives investigate human nature, adaptive mechanisms, and how heritable traits (e.g., Big Five) explain variations in personality across cultures, which implies exporting some sort of self-report instrument that is adapted to each specific culture or

This article was published Online First August 24, 2015.

Angel Blanch and Anton Aluja, Department of Pedagogy and Psychology, University of Lleida, and Institute of Biomedical Research (IRB) Lleida, Spain.

This research was supported by the Spanish Ministry of Science and Innovation (PSI2008-00924/PSIC). This research was performed within the framework of DURSI' Consolidated Group 2014 SGR 111.

Correspondence concerning this article should be addressed to Angel Blanch, Department of Pedagogy and Psychology, Faculty of Education Science, University of Lleida, Avda de l'Estudi General, 4, 25001 Lleida (Catalonia), Spain. E-mail: ablanch@pip.udl.cat

language, generally by applying a back-translation process (Church, 2010; Poortinga & Van Hemert, 2001).

Nevertheless, because cultural differences influence the meaning of words used to describe personality traits (Geisinger, 1994; Spielberger, 2006), there are several mechanisms and effects that may violate key validity and reliability assumptions (Church, 2001; Heine et al., 2002; Norenzayan & Heine, 2005; Peng et al., 1997). Two of such closely intertwined mechanisms are bias and equivalence (Church, 2001; van de Vijver & Leung, 2001; van de Vijver & Tanzer, 2004). In general, bias tends to lower equivalence. Thus, an optimal adaptation of an instrument to different cultures seeks to reduce bias while increasing equivalence, because the biases arising in the back-translation (i.e., construct, method, and item), may produce differences in scores between groups while being interpreted as true cultural differences. Likewise, the equivalence concept implies that the same psychological dimensions emerge across all the studied cultural groups supporting its underlying universality (*etic* or imported position), whereas the inequivalence concept implies that different psychological dimensions emerge across all the studied cultural groups supporting different behaviors across groups (*emic* or indigenous position). Biases and equivalence are major concerns in cross-cultural research implementing the adaptation of an imported self-report instrument. Thus, several recommendations have been made for the adaptation of self-reports to different cultures/languages addressed to reduce and/or control biases while increasing equivalence.

Adaptation and Translation of Psychological Questionnaires

The adaptation of an instrument to a different language does not presuppose a mere literal translation from the original instrument to the target language. The process is rather of a higher complexity, implying the inclusion of idiomatic expressions, the consideration of words that may have several valid translations, or bearing in mind that different concepts may be translated with the same single word depending on the specific target language (Moscoso & Spielberger, 2011). For instance, detailed recommendations regarding the translation of self-reports to improve the compliance with a given language contemplate using shorter and simpler sentences, repeating nouns instead of using pronouns, avoiding colloquialisms, or using specific instead of general terms (Brislin, 1986). Moreover, the entanglement of translating and adapting a self-report instrument grows exponentially because of three different factors. The amount of languages to which the instrument is to be translated, the need to be undertaken by teams rather than individually, and because it should be addressed through a number of predefined stages (Geisinger, 1994; Ozolins, 2009; Su & Parham, 2002; Sumathipala & Murray, 2000).

The International Test Commission (ITC) suggested for instance 22 very general guidelines within four comprehensive domains (context, development, administration, and documentation and score interpretation) for adapting and translating tests to establish the construct equivalence across languages (Hambleton, 2001; Hambleton, Yu, & Slater, 1999; ITC, 2010; Tanzer & Sim, 1999; van de Vijver & Hambleton, 1996). Table 1 includes several studies that provide an in depth description of adaptation and translation issues of specific self-report instruments. All of them, however, report a key method that characterizes the actual translation, review, and adaptation of the instrument, the back-translation process (Geisinger, 1994). There are three basic steps in a self-report back-translation. First, the source (SO) language instrument (SO) is translated to the target language instrument (TA). Second, the target language instrument is back-translated

Table 1
Studies Reporting Procedures to Implement the Translation/Adaptation of Self-Report Instruments to Different Languages/Cultures

Study	Stages	Instrument(s)
Al Jabery & Arabiat, 2011	5	Revised Children's Manifest Anxiety Scale
Araya-Vargas, Gapper-Morrow, Moncada-Jiménez, & Buckworth, 2009	5	Mindful Awareness Attention Scale
Banville, Desrosiers, & Genet-Volet, 2000	7	Value Orientation Inventory-2
Bornman et al., 2010	4	Ages and Stages Questionnaires Mullen Scales of Early Learning
D'Ath, 2005	4	Mental Health Inventory
Geisinger, 1994	10	—
Hambleton et al., 1999	22	Mathematics test (Grade 8)
Jeanrie & Bertrand, 1999	4	California Personality Inventory-434
Ketterer, Han, Hur, & Moon, 2010	4 ~ 6	Minnesota Multiphasic Personality Inventory-2
Moscoso & Spielberger, 2011	—	State Trait Anger Expression Inventory-LAM
Ozolins, 2009	9	Multi-Attribute Arthritis Prioritisation Tool
Sireci et al., 2006	6	Employee engagement
Steele & Edwards, 2008	5	Beck Depression Inventory-II Beck Hopeless Scale Beck Anxiety Inventory
Su & Parham, 2002	3	Evaluation of Sensory Processing
Suleiman, Yates, Berger, Pozehl, & Meza, 2010	2	Pittsburgh Sleep Quality Index
Sumathipala & Murray, 2000	3	Bradford Somatic Inventory

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

(BT) to the SO language (TA → SO = BT). Third, the assessment of equivalence is performed between the original SO language instrument and the BT instrument (SO~BT).

The Present Study

The amount of cross-cultural studies has grown considerably in the latter years (see Figure 1), together with specific methods to address remarkable problems within this field such as 3-level meta-analyses (Cheung, 2014), item bifactor analyses (Cai, Yang, & Hansen, 2011), or integrative data analyses (Shrout, 2009). The use of text mining during the back-translation, however, has been scanty, with just a single study reporting the application of a lexicon to compare synonyms and/or item restatements (Sireci, Yang, Harter, & Ehrlich, 2006). Text mining includes an assortment of especial techniques within the broader data mining and machine learning technologies that are widely used in linguistics, natural language processing, computational statistics, and computer science. Whereas data mining searches for patterns in numerical data, text mining searches for patterns in text data (Feinerer, Hornik, & Meyer, 2008; Hornik, Feinerer, Kober, & Buchta, 2012). A detailed description of text mining is beyond the scope of the current study. Interested researchers can review introductory works (Manning & Schütze, 1999; Weiss, Indurkhya, Zhang, & Damerau, 2004; Witten & Frank, 2005), or specific text mining applications in psychological research (Bantum & Owen, 2009; Clore, Ortony, & Foss, 1987; Cretchley, Rooney, & Gallois, 2010).

Because of its capabilities to examine text data, text mining methodology is a particularly compelling and flexible approach to confront the back-translation problem when adapting and translating self-report instruments from one SO language to a single foreign language or to several foreign languages. In this study, we show with a working example how to treat a considerable amount of text data to examine systematically the back-translation forms of a personality questionnaire generated within a cross-cultural study. More specifically, we demonstrate how text mining can be used to facilitate the analyses of BT forms, to provide a more informed assessment about the quality of each back-translation, and eventually, to improve the attributes of the items within the instrument BT form. Throughout the example, *R* syntax snippets illustrate the application of the suggested text mining procedure. The method was recently implemented during the adaptation and back-translation of the Zuckerman-Kuhlman-Aluja Personality Questionnaire (ZPA-PQ) to 12 different cultures/languages in a cross-cultural research project.

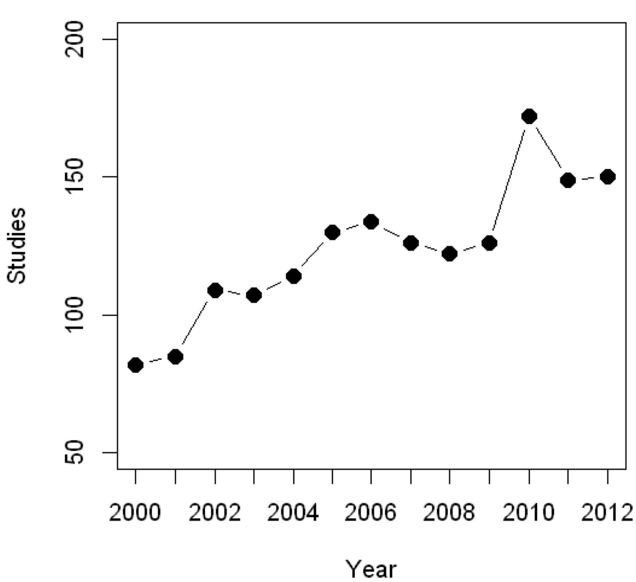

Figure 1. Number of studies found with the keywords "*cross-cultural research*" (source: PsycINFO, 2014).

Method

Measure

The ZKA-PQ (Aluja, Kuhlman, & Zuckerman, 2010) is a 200-item personality questionnaire derived from the alternative Big Five personality model, a facet structured instrument derived from earlier forms used in past research (Aluja & Blanch, 2007; Blanch & Aluja, 2009b; Blanch, Aluja, & Gallart, 2013). This self-report has a 4-point Likert-type response format (1 = *disagree strongly*, 2 = *disagree somewhat*, 3 = *agree somewhat*, and 4 = *agree strongly*). There are four facets per each five broader personality factor: extraversion (positive emotions, social warmth, exhibitionism, and sociability), sensation seeking (thrill and adventure seeking, experience seeking, disinhibition, and boredom susceptibility/impulsivity), neuroticism (anxiety, depression, dependency, and low self-esteem), aggressiveness (physical aggression, verbal aggression, anger, hostility), and activity (work compulsion, general activity, restlessness, and work energy).

Procedure

In 2013, psychologists from several countries and cultures were invited to conduct a cross-cultural study aimed at evaluating the generalization of the ZKA-PQ personality factors and facets. There were 12 participating countries in the project, for which a translation and back-translation of the instrument into each native language was needed. The participating countries were Bosnia, Brazil, Germany, Greece, Iran, Israel, Poland, Portugal, Taiwan, Thailand, Tunis, and Turkey. Figure 2 shows the translation and back-translation procedure proceeding from the English language ZKA-PQ version as the SO language instrument, to the target language instrument (TA) to which the SO language instrument was translated (SO → TA). A psychologist fluent in English and who did not participate in the previous SO → TA translation, BT the translated version (TA) again into English language (TA → SO = BT), giving rise to a BT version of the instrument. The equivalence of the original SO version with the BT version (SO~BT) was assessed by comparing the 200 items in the SO version with their 200 matching items in the BT version. Items conveying different meanings were detected and evaluated. Then, feedback was provided to the team involved in the specific back-translation in the target culture concerning how a particular item could be amended to transmit the original intended meaning formulated in the original SO instrument. This iterative process went on until the 200 items in the BT version expressed reasonably well an analogue meaning as the 200

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

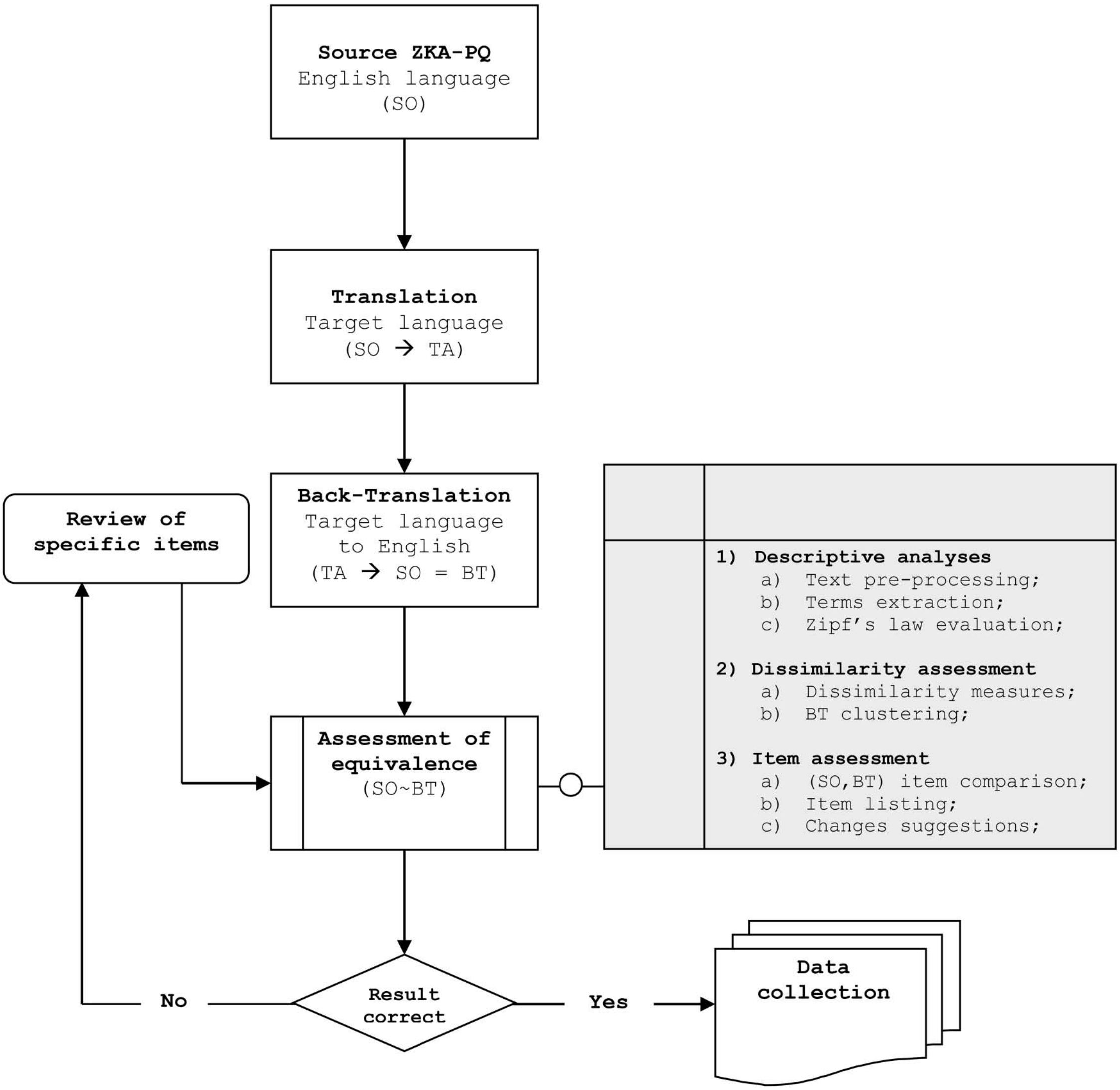

Figure 2. Translation and back-translation procedure. The assessment of the instruments equivalence was performed in accordance with the methodology shown in the gray shadowed box.

parallel items in the SO language instrument. The procedure ended with the data collection in each country.

Text Input Data and Analyses

Input data were the ZKA-PQ 200-item BT versions plus the SO language version embedded within separated plain text files from the 12 participating countries. These text files are available on request to the corresponding author of the current study. Thus, there were 2,600 items included in the text analyses. The 13 ZKA-PQ questionnaires in English language were analyzed with the tm and stringr *R* packages, designed for text analyses and text mining (Feinerer et al., 2008; R Development Core Team, 2014). The R code to perform these analyses is available as supplementary material. The assessment of equivalence was carried out between the SO language and the BT instrument versions (SO∼BT). The gray shadowed block in Figure 2 depicts the text mining application described in the present study. There were three main stages to analyze the text data: descriptive analyses, dissimilarity assessment, and item assessment.

Descriptive analyses. The main purpose of the descriptive analyses stage consisted in gaining a basic understanding of the whole available text data, and studying the most frequent used terms and expressions that might be most relevant for the description of human personality and ultimately aid in the design of more adequate target language instruments (TA). The text data within the text files were uploaded into a Corpus object, a conceptual entity suitable to hold and manage text documents (Feinerer et al., 2008). This Corpus was subsequently instantiated into a raw Document by Term Matrix (V), with the documents (i) as rows and the terms (j) as columns. The entries in this matrix represented the frequencies ($V = [v_{ij}]$) of the terms within each of the 13 documents, the number of times word j appeared in document i. This

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

matrix was analyzed to detect and eventually reduce its sparsity (s), which indicates the percentage of nonpopulated or zero cells within V ($v_{ij} = 0$). Text corpora involving diverse documents usually show a high sparsity because the amount of different words in a single document is markedly lower than the size of the whole vocabulary in all the available documents, therefore, specific algorithms have been designed to address this phenomenon in natural language processing (Barman, Miah, & Singh, 2011; Dhillon & Modha, 2001). For the descriptive aims of the present analysis stage, however, we considered that a text matrix with a low sparsity should be a better choice in contrast with prediction problems where sparsity may be more desirable (Eisenstein, Smith, & Xing, 2011). Besides, a high sparsity might hamper the aforementioned aims of the present analysis stage and the subsequent dissimilarity assessment stage by eliciting poor discrimination distances effects, and the presence of irrelevant and redundant attributes (Houle, Kriegel, Kröger, Schubert, & Zimek, 2010).

Two cleaned Document by Term Matrices with a low amount of empty cells ($s < 10\%$) were generated including terms appearing frequently in the SO and BT instrument forms, and further inspected to detect the most common words and their interrelationships. The sparseness regulating the maximal allowed sparsity was set to a low value to obtain a tractable amount of words concerning the aims of the descriptive analysis stage, which was considered around 50 terms. The two matrices were obtained in accordance with two word weighting schemes (Manning & Schütze, 1999), term frequency (*tf*) and term frequency—inverse document frequency (*tf–idf*), in accordance with the expressions shown in (Equation 1) and (Equation 2), respectively, where tf_{ij} is the number of occurrences of term i in document j, df_i is the number of documents including the term i, and N is the total number of documents in the text corpus.

$$tf_{ij} = 1 + \log(tf_{ij}), \qquad for \quad tf_{ij} > 0 \tag{1}$$

$$tf_{ij} - idf_i = \begin{cases} (1 + \log(tf_{ij})) \cdot \log\dfrac{N}{df_i} & if \quad tf_{ij} \geq 1 \\ 0 & if \quad tf_{ij} = 0 \end{cases} \tag{2}$$

Higher *tf* weightings indicate the importance of a word in a document, whether it is a good descriptor of the content of the document, whereas higher *df* values indicate whether a given word is semantically focused in some field of interest. The *tf–idf* weighting is intended to reflect the relative occurrences of words, together with their semantic specificities surrounding the documents within the given analysis. To examine short sequences of words forming intelligible phrases, bigrams (two words) and trigrams (three words) were also extracted and their frequency examined.

The raw Document by Term matrix with a high sparsity ($s >$ 10%) and the cleaned Document by Term matrix with a low sparsity ($s < 10\%$) were contrasted concerning its adequacy to the Zipf's law. The Zipf's law in statistical linguistics addresses the frequency distribution of words within a language or a collection of words (Zipf, 1945). The N unique words of a vocabulary (V) are ranked in descending order by their frequency (f) such as the most frequent word has rank $r = 1$, the second most frequent word has rank $r = 2$, and so on. The frequency of a word is a power law of its rank with $\alpha \approx 1$, determining an inverse proportional relationship as indicated in (Equation 3).

$$f(r) \propto r^{-\alpha} \tag{3}$$

The Zipf's law is meaningful for evaluating the behavior of natural language and texts and pertinent for the main purpose of the present study (Ferrer i Cancho & Solé, 2002; Piantadosi, 2014). The current text corpus and the two cleaned text matrices obtained with the two weighting schemes (*tf* and *tf-idf*) were therefore assessed concerning its adequacy to the Zipf's law by plotting the log(r) against the log($f(r)$) of the available words. The adequacy of these data to the Zipf's law should be denoted by a descending curve with a slope ≈ -1, and higher departures from the curve for the most frequent and least frequent words (Li, 1992).

Dissimilarity assessment. In the dissimilarity assessment stage, the Document by Term Matrix with $s = 0$ was used to find numerical measures about how different each of the BTs were among them and when compared with the SO form. Measures of dissimilarity (D) determine the difference between two data objects (i, j), being lower when both data objects are more alike and higher when are more dissimilar. Thus, dissimilarity values closer to zero would be indicative of a higher SO~BT similarity. Several dissimilarity measures hold the properties of symmetry ($D[i, j] = D[j, i]$), non-negativity ($D[i, j] \geq 0$), identification mark ($D[i, i] = 0$), definiteness ($D[i, j] = 0 \leftrightarrow i = j$), and triangle inequality ($D[i, j] \leq D[i, k] + D[k, j]$; Mardia, Kent, & Bibby, 1988). For comparative purposes, we determined three dissimilarity measures: euclidean, cosine, and extended Jaccard. The euclidean measure is one of the most common, whereas the cosine and extended Jaccard measures are particularly useful when comparing text data (Hastie, Tibshirani, & Friedman, 2008). For instance, euclidean dissimilarity measures (D_{ij}) were found with the expression shown in (4), where x is the frequency of a given term, i and j are two different instrument versions, and m is the number of terms shared by the two versions.

$$D_{ij} = \sqrt{\sum_{k=1}^{m} (x_{i,k} - x_{j,k})^2} \tag{4}$$

The dissimilarity measures allowed building a dendrogram to detect clusters of different BT versions in accordance with their similitude to the SO form. In addition, intercountry BT distances were obtained and used to build heat maps to depict the hierarchical clustering structure of the self-report versions (Wilkinson & Friendly, 2009). These analyses provided an initial estimation of the quality of each back-translation, suggesting that back-translations with higher dissimilarities should be screened in greater detail in the item assessment stage.

Item assessment. In the item assessment stage, there were three differentiated steps. First, the items in the SO form were compared with their matching items in the BT forms to determine whether these paired matching items were exactly equal in both versions (i.e., Item 5 in SO and BT). Each pair of matching items was treated as equal when the items in each pair had the same words organized in the same order. Second, only unequal paired matching items were listed together to facilitate the identification of items in the BT form conveying an antagonistic meaning to that intended by the matching item in the SO form. Particular care was taken at this stage to avoid any confounding with *emic* item content that might be unique to the country/culture at hand. Third, specific changes were suggested for the problematic items detected in the previous step, aiming to reformulate the item in accordance with the meaning denoted in the original SO form prior to the data collection process.

Results

Descriptive Analyses

The first halve in Table 2 presents the *R* syntax of the text mining library (tm), the uploading of the text data into the Corpus object (*zka*), the preprocessing of the text data by eliminating numbers, punctuation symbols, blank spaces, and the transformation of the words to lower case. Stop words such as prepositions, articles, and pronouns, which typically have a low informative value, were also removed. Moreover, the raw Document Term Matrix (*dtm*) was instantiated and the sparse factor (*dtmS*) reduced.

The second halve in Table 2 compares basic data for the raw text and the cleaned text with the two weighting schemes (*tf* and *tf-idf*). For the *tf* weighting, the raw text Document by Term Matrix had 1,378 different terms, which for 13 documents made 17,914 entries. Besides, there were 5,030 nonsparse entries or words used at least once in any of the 13 available documents, while leaving 12,884 of zero cells in the text matrix with a sparse factor of 72% (12884/17914 = 0.7192). The maximal term length indicates the number of terms in the longest document within the Corpus text object. In contrast, the cleaned Document by Term Matrix with the sparseness set at .07, yielded a null sparse factor ($s = 0$) for 51 different terms, 663 nonsparse entries throughout the whole text matrix, and a maximal term length of just 10. For the *tf-idf* weighting, the raw text Document by Term Matrix had also 1,378 different terms and 17,914 entries. There were 4,367 nonsparse entries, 13,547 cells at zero, and a sparse factor of 76% (13,547/17,914 = 0.7562). The cleaned Document by Term Matrix with the sparseness set at .15 resulted in a sparse factor of 8% (39/507 = 0.0769) for 39 terms, and 468 nonsparse entries in the text matrix. The cleaned matrices (*dtmS*) with low sparsity including just 51 and 39 terms in accordance with both weightings (*tf* and *tf-idf*), were much simpler than the raw matrices (*dtm*), more informative for grasping the most common words and their associations, and more helpful for evaluating the meaning of the paired matching items in the item assessment stage.

Figure 3 shows the most common words under the two weighting schemes that were used across all versions of the instrument, 51 terms (*tf* weighting) and 39 terms (*tf-if* weighting). Lighter boxes indicate that terms were used less frequently in the respective versions, while darker boxes indicate that terms were used more frequently in the respective versions. Moreover, the first two halves in Table 3 show a sample of the six most common words found under both weightings throughout the respective text matrices ordered in descending frequency of use. For each term within this sample, the three terms with the higher correlations are shown. For instance, *people* was the most used term in the *tf* weighting (first halve in Table 3), and was most strongly associated with the terms *stay* (.73), *job* (.70), and *prefer* (.62). Whereas *enjoy* was the most used term in the *tf-idf* weighting (second halve in Table 3), and was most strongly associated with the terms *unrest* (.80), *unfortunate* (.79), and *spending* (.78). Further, the third halve in Table 3 displays the most frequent bigrams (two-word sentences) and trigrams (three-word sentences) ordered in descending frequency of use, *some-*

Table 2
Text Preprocessing and Description of Raw (dtm) and Cleaned Text (dtmS) With Two Weighting Schemes (tf and tf-idf)

```
#Library loading
library(tm)
#Creates the Corpus containing the 13 documents;
#"/directory/" is the directory address containing the 12 plain text #files #corresponding to the
  performed back-translations;
zka<-Corpus(DirSource("/directory/"), readerControl = list(language="en"))
#Remove numbers, punctuation, whitespace, stop words and convert to lower case;
z<-tm_map(zka, removeNumbers)
z<-tm_map(z, removePunctuation)
z<-tm_map(z, stripWhitespace)
z<-tm_map(z, tolower)
stopw<-c("the", "what", "when", "with","'","'")
z<-tm_map(z, removeWords, stopw)
z<-tm_map(z, removeWords, stopwords("english"))
#Creates the Document Term Matrix (dtm) and removes sparse terms (dtmS)
dtm<-DocumentTermMatrix(z, list(stemming = F, stopwords = T)
dtmS<-removeSparseTerms(dtm, 0.07)
```

	tf Weighting		*tf-idf* Weighting	
	Raw text (*dtm*)	Cleaned text (*dtmS*)	Raw text (*dtm*)	Cleaned text (*dtmS*)
Documents	13	13	13	13
Terms	1378	51	1378	39
Nonsparse entries	5030/12884	663/0	4367/13547	468/39
Sparseness	—	.07	—	.15
Sparse factor (*s*)	72%	0%	76%	8%
Maxim term length	28	10	28	11

Note. The "ZKA" corpus object contains the source and the twelve back-translated forms.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

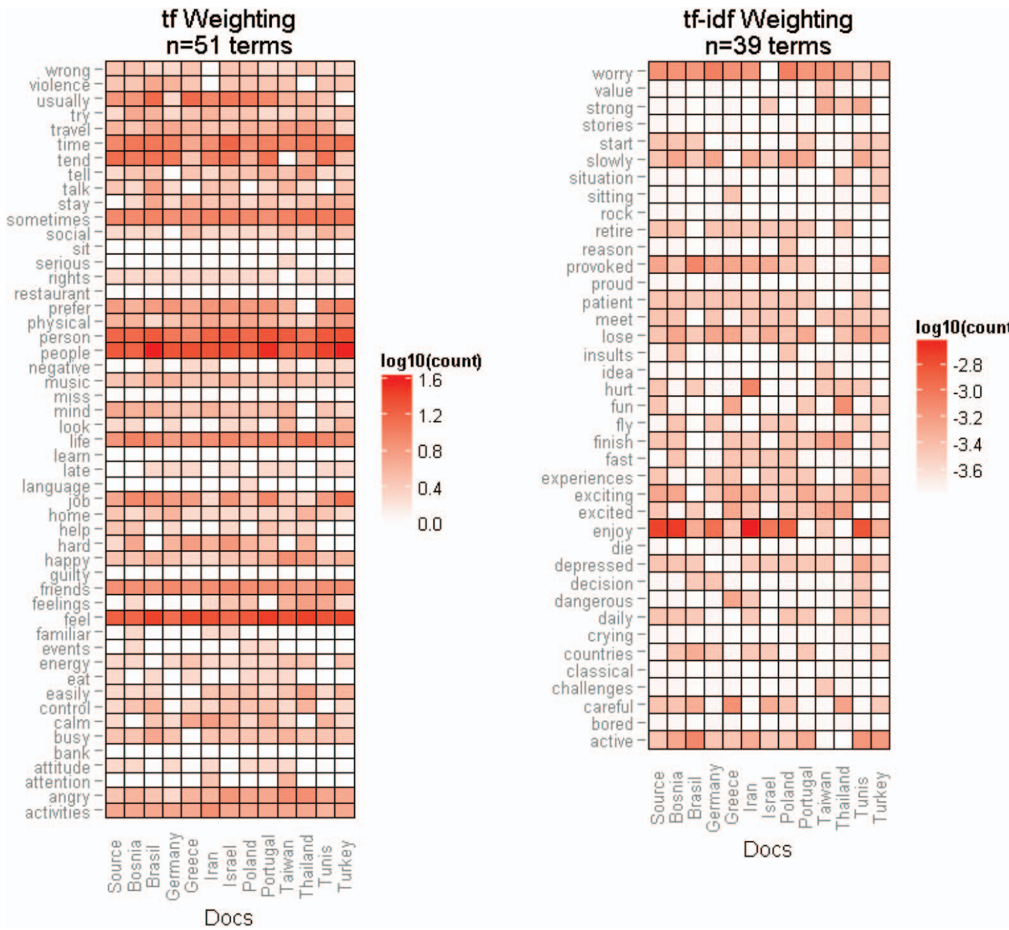

Figure 3. Most common words under two weighting schemes across the source and back-translated versions of the instrument: 51 terms with *tf* weighting, and 39 terms with *tf-if* weighting. Values indicate the log_{10} terms frequencies. See the online article for the color version of this figure.

times feel (34) and *sometimes feel depressed* (11) were the most used n-grams within the present data. This summary provides an overview of how the most frequent terms and combination of two and three terms may appear together to describe the meanings that are to be transmitted by the self-report items in the available back-translations.

Figure 4 shows the evaluation of three Document by Term Matrices (raw text, cleaned text with *tf* weighting, and cleaned text with *tf-idf* weighing), concerning their adequacy to the Zipf law. These findings support the consistence with the Zipf law of the text matrices at $s = 72\%$ ($R^2 = .93$), $s = 0\%$ ($R^2 = .95$), and $s = 8\%$ ($R^2 = .96$), with higher discrepancies from the curve for the most frequent and least frequent words. Also, the slopes for the raw text and *tf* weighting matrices are very close to -1, even though it differs markedly from -1 for the *tf-idf* weighing matrix (Li, 1992; Zipf, 1945).

Dissimilarity Evaluation

Table 4 shows information about the dissimilarity evaluation stage of the analysis plan. The first two rows in Table 4 show the amount of BT items with a markedly different meaning in respect to the SO form and for which some feedback concerning the modification of its meaning was provided. The second row indicates the percentage over the 200 instrument items. Basic spelling mistakes were disregarded here, albeit they were highlighted in the item assessment stage. Thus, for the back translation received from Portugal, there was only one item (0.5%) with a different meaning for which feedback on how to amend it was provided, whereas for the back translation from Taiwan, there were 16 items (8%) of this kind. It should be remarked that this information was obtained in fact from the item assessment stage, and that it is included here just for post hoc comparison with the obtained dissimilarity indices.

The next three sections in Table 4 show the euclidean, cosine, and extended Jaccard dissimilarity measures among the 12 BT versions and the SO form. Lower values in each of these measures, indicated that the respective BT was more alike with the SO form. The three dissimilarity measures were highly correlated (.60∼.97, $p < .05$). Thus, in accordance with the cosine and extended Jaccard dissimilarity measures, the BTs from Bosnia, Germany, Iran, and Tunis were the more similar to the SO form, whereas the BTs from Thailand, Taiwan, Greece, and Turkey were the most dissimilar to the SO form.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
The 12 Most Frequent (F) Terms in the Document by Term Matrix With Their Three Most Correlated Terms (First Six Terms With tf Weighting, Second Six Terms With tf-idf Weighting), and the Most Frequent Bigrams and Trigrams

Term	*F*	Associated terms (correlation, frequency)		
people	314	stay (.73, 38)	job (.70, 78)	prefer (.62, 78)
feel	279	tell (.61, 37)	stay (.55, 38)	people (.52, 314)
person	215	people (.58, 314)	prefer (.58, 78)	busy (.56, 40)
time	132	busy (.47, 40)	person (.45, 215)	job (.44, 78)
sometimes	125	easily (.80, 33)	feelings (.77, 32)	angry (.58, 63)
tend	112	friends (.50, 101)	rights (.48, 25)	mind (.45, 42)
enjoy	75	unrest (.80, 6)	unfortunate (.79, 3)	spending (.78, 9)
worry	46	sports (.77, 29)	hostile (.71, 16)	main (.65, 13)
active	33	people (.82, 314)	tasks (.78, 5)	aggressive (.76, 25)
exciting	31	conversations (.73, 19)	criticize (.71, 11)	efforts (.71, 11)
lose	31	rights (.82, 25)	rock (.82, 12)	experiences (.74, 21)
provoked	31	parties (.77, 50)	criticized (.73, 7)	countries (.71, 20)

Bigram	*F*	Trigram	*F*
sometimes feel	34	sometimes feel depressed	11
defend rights	20	people speak language	10
physical activities	20	violence defend rights	10
busy time	19	control tone voice	9
serious person	13	force defend rights	9
stay home	13	parties meet exciting	9
feel crying	12	people walk slowly	9
feel depressed	12	activities organized friends	8
time familiar	12	foreign countries people	8
feel guilty	11	insults tend aggressive	8
rock music	11	physical activities sports	8
social life	11	rich social life	8
speak language	11	travel foreign countries	8

The dendrograms obtained with the four dissimilarities matrices (see Figure 5), show how the back-translations from the 12 countries assembled among them. For instance, in the Jaccard dendrogram there were two main clusters from a height around .30, the first of them including the BTs from Brazil, Portugal, Tunis, and Turkey, and the second including the rest of BTs. Furthermore, the heat maps obtained for each kind of measure show the arrangement of the BTs in the plot's rows and columns so that similar rows and similar columns (i.e., BTs) are grouped together, with separate dendrograms for rows and columns. Darker colors indicate lower distances among the countries BTs, lighter colors indicate higher distances among the countries BTs. For instance, in the euclidean distances it can be seen that the BTs from Turkey, Brazil, and Portugal, are closer (i.e., more similar) than the BTs from the SO version, Bosnia, and Germany, with these three latter versions being in turn more similar among them. These data indicate that extra care should be taken in the next item assessment stage particularly concerning the BTs showing higher dissimilarity indices and/or clustering more far apart from the SO version.

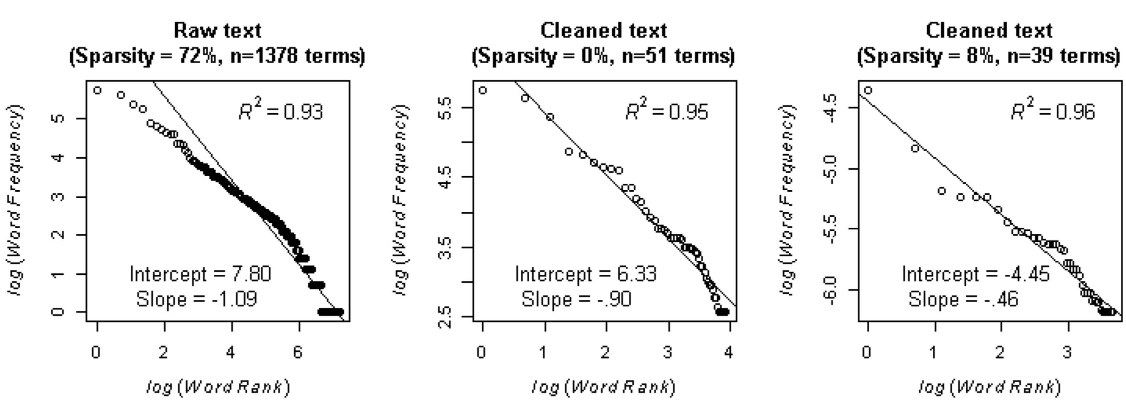

Figure 4. Zipf law evaluation in three Document by Term matrices at three sparse-factors $s = 72\%$, $s = 0\%$, and $s = 8\%$. The left and central plots correspond to the *tf* weighting scheme, the right plot corresponds to *tf-idf* weighting scheme.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 4
Euclidean, Cosine, and Jaccard Distance Coefficients Between the ZKA-PQ Source Form With 12 Back-Translated Forms

	0	1	2	3	4	5	6	7	8	9	10	11	12
Feedback*	—	1	2	5	5	6	6	6	7	9	10	15	16
% of 200 items	—	0.5	1	2.5	2.5	3	3	3	3.5	4.5	5	7.5	8
Euclidean													
0. Source	0												
1. Portugal	21.24	0											
2. Turkey	26.87	17.35	0										
3. Germany	9.85	25.46	26.89	0									
4. Israel	13.11	22.87	27.31	16.09	0								
5. Bosnia	9.59	24.76	28.53	11.70	10.95	0							
6. Iran	10.91	19.18	25.08	15.68	14.53	14.52	0						
7. Poland	12.97	24.80	28.43	15.91	15.30	12.88	11.53	0					
8. Greece	18.71	23.47	28.64	20.32	20.40	19.18	14.53	12.65	0				
9. Thailand	21.12	27.51	31.37	22.91	23.45	19.85	19.47	16.37	21.17	0			
10. Tunis	13.42	15.33	18.22	15.91	16.85	17.38	13.75	18.97	22.49	22.89	0		
11. Brazil	26.06	12.49	20.81	29.70	26.02	29.55	24.21	29.44	26.44	33.21	21.28	0	
12. Taiwan	18.33	17.35	29.43	20.02	21.40	18.11	17.97	13.42	22.18	15.30	21.49	34.31	0
Cosine													
0. Source	0												
1. Portugal	.03	0											
2. Turkey	.09	.04	0										
3. Germany	.03	.05	.07	0									
4. Israel	.04	.07	.11	.05	0								
5. Bosnia	.03	.06	.10	.04	.02	0							
6. Iran	.03	.03	.08	.06	.05	.05	0						
7. Poland	.05	.05	.10	.08	.05	.05	.03	0					
8. Greece	.09	.06	.11	.12	.10	.10	.06	.04	0				
9. Thailand	.11	.10	.15	.14	.13	.10	.09	.07	.11	0			
10. Tunis	.03	.03	.04	.03	.06	.05	.04	.06	.10	.11	0		
11. Brazil	.05	.02	.06	.06	.07	.08	.05	.07	.06	.14	.05	0	
12. Taiwan	.10	.11	.11	.12	.11	.10	.09	.05	.14	.06	.09	.14	0
Jaccard													
0. Source	0												
1. Portugal	.16	0											
2. Turkey	.24	.08	0										
3. Germany	.06	.23	.25	0									
4. Israel	.08	.17	.23	.13	0								
5. Bosnia	.05	.21	.27	.08	.06	0							
6. Iran	.06	.12	.20	.13	.09	.11	0						
7. Poland	.09	.21	.27	.15	.11	.10	.07	0					
8. Greece	.17	.18	.26	.22	.18	.19	.11	.09	0				
9. Thailand	.21	.24	.30	.26	.23	.19	.17	.14	.20	0			
10. Tunis	.08	.07	.10	.12	.11	.13	.08	.16	.21	.20	0		
11. Brazil	.21	.04	.11	.28	.20	.27	.18	.27	.21	.31	.13	0	
12. Taiwan	.17	.27	.28	.22	.21	.17	.16	.10	.24	.12	.20	.34	0

Note. BT forms list in increasing number of items with feedback.
* Amount of items for which feedback was provided.

Item Assessment

Table 5 shows the necessary R commands from the stringer package to finding equal and unequal items. In the example, the SO form saved within *x*, was compared with the BT form Brazil (ZKA_Brasil.txt), saved within *y*. Both objects, *x* and *y*, were subsequently equated, yielding either a 0 indicating that matching items in both forms were unequal, or a 1, indicating that matching items in both forms were equal. For example, the six lines with 0 and 1 values show that the items numbered 1 to 12 yielded 0[rquote]s, thus these items differed in both forms, SO and BT, concerning its words and/or the words order. On the other hand, Item 13 yielded 1, thus this item was exactly the same in the SO and BT forms, concerning its words and/or the words order. The in depth meaning assessment of matching items was therefore made only for the items yielding 0, that is, unequal in both forms. Thus, only different matching items were printed and subsequently inspected to assess whether the original item meaning conveyed in the SO form was maintained in the item meaning in the BT form. The list of paired items was obtained with the loop shown at the bottom part of Table 5.

Table 6 shows examples of matching items in the three situations that arose at this stage, equal items, unequal items conveying the same meaning, and unequal items conveying different mean-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

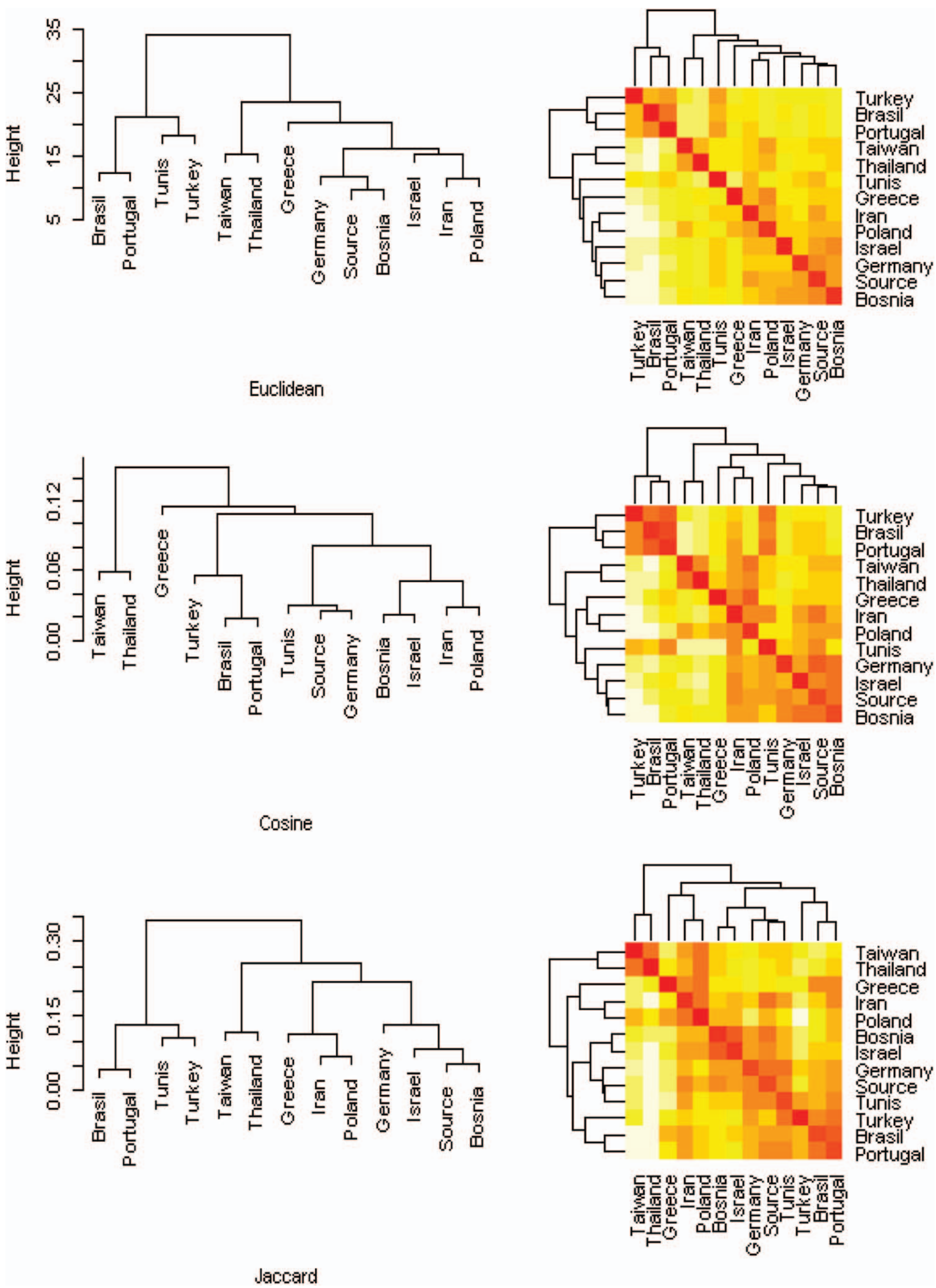

Figure 5. Dendrograms and heat maps for 13 versions of the instrument (euclidean, cosine, and Jaccard dissimilarities). See the online article for the color version of this figure.

ings. All these items were chosen randomly across the 12 BT forms of the instrument. The first item in each pair corresponds to the SO version of the instrument, the second item in each pair correspond to the matching item in the BT version of the instrument. The first part in Table 6 shows that matching Items 5, 90, and 159 were identical in both forms (SO and BT), thus, items of this kind were not printed in the previous step or further reviewed. The second part in Table 6 shows the matching items numbered 21, 64, 89, 127, 166, and 175. These items had different words or the same words used in a different ordering in both, the SO and BT items. However, these items channeled primarily the same meaning as that intended by the matching item in the SO form. Items of this type were examined by the BT review team although no feedback regarding its modification was provided to the back-translating team. The third part in Table 6 shows examples of the most problematic kind of items. It should be remarked that this type of items mostly belonged to BT versions with higher dissimilarities as found in the previous analysis stage. As it can be seen,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 5
Item Assessment of Source and Back-Translated Versions

```
#Library loading
library(stringr)
#Item comparison
x<-read.table("ZKA.txt", header = F, sep="\t", quote="")
y<-read.table("ZKA_Brasil.txt", header = F, sep="\t", quote="")
str_count(as.matrix(x),as.matrix(y))
#Item comparison outcomes
#[Item number]
[1] 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0
[35] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[69] 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[103] 0 0 1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
[137] 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[171] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
#This loop prints the items at 0. Items at 1 are not listed;
for(i in 1:200) {
if(str_detect(as.matrix(x[i,]),as.matrix(y[i,]))== FALSE) {
print(as.matrix(x[i,]))
print(as.matrix(y[i,]))
}
}
```

these BT items express contradictory meanings to that intended by their matching SO items. The discrepancy is evident. Consider for instance the case for item number 166. This is a complex worded item formed by two clauses and including a negation in the second clause. This item is correctly BT in the second halve of Table 6, with both BT items transmitting an analogue meaning as the SO matching item. However, this same item number 166 is incorrectly BT in the third halve of Table 6, with both BT items transmitting precisely a completely antagonistic meaning to that intended by the SO matching item. Item 175 also shows the same behavior in both kinds of items. For the problematic items, a recommendation was provided to the back-translating team in order to readapt them to a more appropriate translation before actually realizing the data collection with the instrument.

Discussion

This is the first study addressing the back-translation of a self-report instrument with text mining methods, and showing how they can be used to improve both, the back-translation process in itself, and the eventual version(s) of an imported self-report instrument in cross-cultural research. Several works have proposed a variety of stages to adapt a self-report instrument to a different culture/language (see Table 1). Nevertheless and to the best of our knowledge, text mining applications have been rather scarce to analyze the text data typically produced during this process. The current study aimed to fill this gap by suggesting a text mining method to be applied in the back-translation of a personality questionnaire when adapted to different languages. The procedure describes a real application that was carried out in a cross-cultural research project focused in replicating the factors and facets of the ZKA personality questionnaire in 12 different cultures/languages. The method considered a three-stage assessment of the equivalence of the SO language instrument with each of the 12 available back-translations: a descriptive analysis of the available text data, a dissimilarity assessment between the BT forms and the SO language instrument, and an item assessment where original and BT items were contrasted concerning its meaning.

This text mining method can contribute to the back translation process of psychological instruments within cross-cultural research in two meaningful ways. First, the procedure contrives a systematic and three-stage orderly approach to accomplishing a more informed evaluation about the equivalence of alternative versions of the self-report in different languages. Second, the procedure aids to refining the BT forms before being actually used in pilot and/or field studies, which is likely to improving in turn the reliability and validity instrument's scores.

Systematization of the Back-Translation Process

Text data per se are complex. Let alone the evaluation of sentences listed in the form of paired items that are supposed to transmit an analogue meaning underlying an abstract psychological construct. The three-stage approach suggested here aims to ease this demanding task.

The descriptive analysis stage constitutes an initial exploratory scenario, useful to have an initial look at the available text Corpus, and to familiarize with the main characteristics of the text at hand. The simplification of the text Corpus by reducing the sparse terms allows to determining the most used words encountered at latter meaning assessment stages, and more importantly, how these words associate with other words. Moreover, assessing the Zipf law (Zipf, 1945) in both, the raw and simplified text matrices explores whether major abnormalities exist in the text data. The material obtained at this stage (relative terms frequencies and terms semantic specificities, terms associations, bigrams and trigrams) can be useful to shed light on how the words used in the BTs combined to describe the psychological constructs aimed to be measured with the instrument. For instance, back-translation teams can use this information as an empirically derived starting point to consider whether the detected terms and associations are worth further evaluation while bearing in mind the linguistic

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 6
Three Types of Items in the Source and Back-Translated Versions

(1) EQUAL ITEMS	
5.	I often feel restless for no apparent reason.
5.	I often feel restless for no apparent reason.
90.	Sometimes I find it hard to concentrate.
90.	Sometimes I find it hard to concentrate.
159.	Other people consider me to be a loner.
159.	Other people consider me to be a loner.
(2) UNEQUAL ITEMS CONVEYING THE SAME MEANING	
21.	If I am pushed far enough, I might hit another person.
21.	When provoked strongly enough, I could hit another person.
64.	I enjoy my daily activities.
64.	I have fun with my daily activities.
89.	I am rather a cold person with others.
89.	I am probably cold towards others.
127.	I would not like a job involving a lot of travel.
127.	I would not like to have a job that demanded traveling a lot.
166.	When I think someone is wrong, I cannot help telling them.
166.	If I think that somebody is mistaken, I can't stop myself from saying so.
166.	When I think that someone is wrong, I cannot help saying so.
175.	My confidence in myself is lost when someone I love is critical of me.
175.	I lose my confidence when someone I love is judgmental with me.
(3) UNEQUAL ITEMS CONVEYING A DIFFERENT MEANING	
89.	I am rather a cold person with others.
89.	I tend to be rather cool toward others.
91.	I have a strong temperament.
91.	I am a generous person.
95.	I feel helpless if there is no one to advise me.
95.	I feel like I have done the right thing even without the advice of others.
166.	When I think someone is wrong, I cannot help telling them.
166.	I find it difficult to tell people when I think they are wrong about something.
166.	If I believe someone is wrong I can't tell them.
175.	My confidence in myself is lost when someone I love is critical of me.
175.	I can no longer trust someone I love after they have criticized me.
198.	I am not very motivated in my job and I do it out of necessity.
198.	I do not feel very motivated at work I do not need to work.

Note. (1) Items that are equal; (2) Items that are unequal but convey the same semantic meaning; and (3) Items that are unequal but convey a different semantic meaning. The first item of each pair corresponds to the source (SO) version, the second item of each pair correfirst item of each pair corresponds to the sourcesponds to the back-translated (BT) version. For Item 166, three examples are shown in (2) and (3), the first item in each triplet corresponds to SO, the other two items to BT versions.

system of a single culture, and particularly, the words difficulty, their familiarity, and their specific connotations to the members of this culture (Geisinger, 1994).

This descriptive analysis stage can be adapted to other problems to addressing different aims of those involved in self-report back-translation, particularly by stemming the available terms or by defining specific stop words. Stemming consists in erasing word suffixes while retaining word radicals. The terms in the current study were not stemmed because of different reasons. First, stemmer algorithms are mainly used in fields such as linguistic morphology and information retrieval. Second, stemming appears to be more useful when dealing with huge amounts of data (several megabytes and even gigabytes), which is not the case of the BTs text data in the present study. Third, stemming algorithms are also subject to problems such as over stemming and under stemming, which might hinder the subsequent analyses carried out in a back-translation process of the kind described in the present study. For instance, these problems are likely to arise when dealing with words with the same stem that convey unrelated idiomatic meanings in different languages. In addition and to the best of our knowledge, this is the first in depth description of a text mining application to address a very specific problem such as the back-translation of a self-report instrument. Therefore, we attempted to provide a rather simplified methodology. Depending on the specific text mining problem at hand, researchers can define their own stop words. In the present application, care was taken concerning the selection of stop words, these were mainly determinants (a, another, the . . .), conjunctions (and, but, or . . .), prepositions (before, between, under . . .), pronouns (it, he, she . . .), or auxiliary verbs (be, can, have . . .), which are likely to be of little use in attempting to examine the description of human personality in different cultures, as this was the goal of this study.

The dissimilarity assessment stage is particularly useful in situations where there are several back-translations from different languages (Ozolins, 2009). However, it can also be valuable to detect marked departures concerning the two emerging forms of a self-report instrument in bilingual translations (Benet-Martínez & John, 1998; Schwartz et al., 2014). This analysis allows foreshadowing potential problematic BT forms unusually dissimilar to the original language form prior to the item assessment stage, by providing some indications about the back-translations where problematic items are more likely to arise. However, it should also be remarked that there were back translations with high dissimilarity indices although with a reduced percentage of problematic items.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

After these initial analyses, researchers involved in the review and evaluation of a back-translation version are already acquainted with the most frequently used words and expressions, their most robust interrelationships, the adequacy of its basic text structure concerning the relation of word rank with word frequency (Zipf law), and its degree of similarity with the self-report original language version. From a practical standpoint, researchers can be therefore much more informed about specific linguistic peculiarities of the available back-translations, and of particular back-translations that might be especially problematic because of significant departures in their similarity in respect to the original version. Furthermore, apart from saving time and effort in the back-translation process, the proposed method is helpful to systematize some tedious parts of the review process, such as item listing and reviewing. For instance, looking for items that are already equal in the SO and BT forms eases the whole process by reducing the amount of items to review. Besides, the method suggested here intends to define a common procedure amenable to be shared and contrasted across different research teams and/or research projects, while flexible enough to allowing for further modifications and/or improvements.

Quality of Back-Translations

The main purpose of the back translation process is to contribute to obtain target language instruments capable to tapping equivalent psychological constructs across cultures. Written language expresses a diversity of complex concepts and meanings subject to considerable variations across different cultures (Geisinger, 1994; Spielberger, 2006). Thus, the items inefficiently revised and adapted can contribute to increase biases while reducing equivalence (van de Vijver & Leung, 2001; van de Vijver & Tanzer, 2004), with direct implications for the reliability and validity of the instrument's scores (Church, 2001, 2010; Peng et al., 1997). In the present study, the amount of items for which meaning changes were suggested was rather low for both, a single back translation with a maximum of 16 items out of 200 (8%) and the whole set of 12 back translations with 88 items out of 2,400 (4%). Moreover, it has been argued that item translation deficiencies might not increment biases at mean level comparisons because averaging out when aggregated into scales (McCrae, 2001).

Nonetheless, the implemented method was inexpensive in accordance with economic or time standards, relatively easy to apply, and contributed to detect and amend the meaning of the aforementioned problematic items. This is likely to have contributed to improving the reliability and validity of the instrument's scores across the different participating countries. Indeed, subsequent data analyses from the cross-cultural research project described in this work, indicate that the respective adapted instruments calibrated well the intended psychological construct (alternative five-factor personality model), supporting its usefulness for research across the studied cultures.

Prospects and Suggestions for Future Research

Despite the large literature in personality and cross-cultural research, no study has reported so far a detailed text mining application to address the back translation problem in cross-cultural research. The described methodology should be taken, however, as an initial basis subject to improvement, modification, or accommodation to other related problems. For example, although the present study focused in a personality questionnaire, the suggested text mining method is extensible to study self-reports of a variety of psychological constructs measured across cultures. In addition, the text mining methodology is not restricted to cross-cultural research, but also pertinent to other problems involving the adaptation or modification of a psychological self-report instrument, such as in research works striving for extensions of a construct nomological network to a different age group (Lynam, Derefinko, Caspi, Loeber, & Stouthamer-Loeber, 2007).

Whereas the current text mining method was conducted from a cross-cultural perspective, it could equally be adapted to the study of personality or other psychological constructs from indigenous or cultural psychology perspectives. Text mining the instruments designed in multiple languages can be worthwhile for the organization of a taxonomy of universal human personality (Saucier & Goldberg, 2001). Moreover, text mining methods could also be easily applied to prominent research areas in the field, such as investigation of response biases or reference group effects, the association of indigenous constructs to universal traits, or the development of more contextualized measures of personality (Church, 2010). There are other available text mining tools that could furnish a more in depth text analyses in these lines of research, such as the specialized lexicon *Amplified English Source*, reported to specify acceptable or unacceptable synonyms in item restatements (Sireci et al., 2006). Another available text mining device is WordNet, a large database of English language that groups nouns, verbs, adjectives and adverbs into sets of cognitive synonyms, and addressed to fields related to computational linguistics and natural language processing (Fellbaum, 1998; Miller, 1995).

Conclusion

The present study suggested the usefulness of text mining methodology in addressing a specific problem such as that faced when translating, back translating, and adapting a self-report psychological questionnaire to different cultures/languages in cross-cultural research. Through an example, we described a real application conducted during the adaptation of a personality questionnaire to 12 different cultures. Text mining is a versatile methodology, easily adjusted to a variety of research topics ranging from cross-cultural studies to general psychology research involving the adaptation of self-report psychological instruments to a variety of problems and fields.

References

Al Jabery, M. A., & Arabiat, D. H. (2011). Psychometric properties of the Arabic translated version of the RCMAS: Preliminary indicators from a Jordanian sample. *Journal of International Counselor Education, 3,* 13–24.

Allik, J., & McCrae, R. R. (2004). Toward a geography of personality traits: Patterns of profiles across 36 cultures. *Journal of Cross-Cultural Psychology, 35,* 13–28. http://dx.doi.org/10.1177/0022022103260382

Aluja, A., & Blanch, A. (2007). Comparison of impulsiveness, venturesomeness and empathy (I7) structure in English and Spanish samples: Analysis of different structural equation models. *Personality and Individual Differences, 43,* 2294–2305. http://dx.doi.org/10.1016/j.paid.2007.07.019

Aluja, A., Kuhlman, M., & Zuckerman, M. (2010). Development of the Zuckerman-Kuhlman-Aluja Personality Questionnaire (ZKA-PQ): A factor/facet version of the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ). *Journal of Personality Assessment, 92,* 416–431. http://dx.doi.org/10.1080/00223891.2010.497406

Araya-Vargas, G. A., Gapper-Morrow, S., Moncada-Jiménez, J., & Buckworth, J. (2009). Translation and cross-cultural validation of the Spanish version of the Mindful Awareness Attention Scale (MAAS): An exploratory analysis and potential applications to exercise psychology, sport and health. *International Journal of Applied Sports Science, 21,* 94–114.

Bantum, E. O., & Owen, J. E. (2009). Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychological Assessment, 21,* 79–88. http://dx.doi.org/10.1037/a0014643

Banville, D., Desrosiers, P., & Genet-Volet, Y. (2000). Translating questionnaires and inventories using a cross-cultural translation technique. *Journal of Teaching in Physical Education, 19,* 374–387.

Barman, C., Miah, S., & Singh, B. C. (2011). Feature extraction clustering in text mining using NMF basis probability. *UALB Journal of Science and Engineering, 2,* 50–55.

Barrett, P. T., Petrides, K. V., Eysenck, S. B. G., & Eysenck, H. J. (1998). The Eysenck Personality Questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences, 25,* 805–819. http://dx.doi.org/10.1016/S0191-8869(98)00026-9

Benet-Martínez, V., & John, O. P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75,* 729–750. http://dx.doi.org/10.1037/0022-3514.75.3.729

Blanch, A., & Aluja, A. (2009a). Validation study of the Spanish version of the Work-Family Conflict Questionnaire (CCTF). *The Spanish Journal of Psychology, 12,* 746–755. http://dx.doi.org/10.1017/S1138741600002110

Blanch, A., & Aluja, A. (2009b). Work, family and personality: A study of work-family conflict. *Personality and Individual Differences, 46,* 520–524. http://dx.doi.org/10.1016/j.paid.2008.12.004

Blanch, A., Aluja, A., & Gallart, S. (2013). Personality assessment through internet: Factor analyses by age groups of the ZKA personality questionnaire. *Psychologica Belgica, 53,* 101–119. http://dx.doi.org/10.5334/pb-53-4-101

Bornman, J., Sevcik, R. A., Romski, M. A., & Pae, H. K. (2010). Successfully translating language and culture when adapting assessment measures. *Journal of Policy and Practice in Intellectual Disabilities, 7,* 111–118. http://dx.doi.org/10.1111/j.1741-1130.2010.00254.x

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). Newbury Park, CA: Sage.

Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods, 16,* 221–248. http://dx.doi.org/10.1037/a0023350

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19,* 211–229. http://dx.doi.org/10.1037/a0032968

Church, A. T. (2001). Personality measurement in cross-cultural perspective. *Journal of Personality, 69,* 979–1006. http://dx.doi.org/10.1111/1467-6494.696172

Church, A. T. (2010). Current perspectives in the study of personality across cultures. *Perspectives on Psychological Science, 5,* 441–449. http://dx.doi.org/10.1177/1745691610375559

Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology, 53,* 751–766. http://dx.doi.org/10.1037/0022-3514.53.4.751

Cretchley, J., Rooney, D., & Gallois, C. (2010). Mapping a 40-year history with Leximancer: Themes and concepts in the Journal of Cross-cultural psychology. *Journal of Cross-Cultural Psychology, 41,* 318–328. http://dx.doi.org/10.1177/0022022110366105

D'Ath, P. (2005). *Translation and cultural contextualisation of the Mental Health Inventory into 13 community languages*. West End, Australia: Queensland Transcultural Mental Health Centre.

Dhillon, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine Learning, 42,* 143–175. http://dx.doi.org/10.1023/A:1007612920971

Eisenstein, J., Smith, N. A., & Xing, E. P. (2011, June). *Discovering sociolinguistic associations with structured sparsity*. Paper presented at the 49th Annual Meeting of the Association for Computational Linguistics, Portland, Oregon.

Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software, 25,* 1–54.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Ferrer i Cancho, R., & Solé, R. V. (2002). Zipf's law and random texts. *Advances in Complex Systems, 5,* 1–6. http://dx.doi.org/10.1142/S0219525902000468

Fischer, R., & Boer, D. (2011). What is more important for national well-being: Money or autonomy? A meta-analysis of well-being, burnout, and anxiety across 63 societies. *Journal of Personality and Social Psychology, 101,* 164–184. http://dx.doi.org/10.1037/a0023663

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6,* 304–312. http://dx.doi.org/10.1037/1040-3590.6.4.304

Hambleton, R. K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17,* 164–172. http://dx.doi.org/10.1027//1015-5759.17.3.164

Hambleton, R. K., Yu, J., & Slater, S. C. (1999). Field-test of the ITC guidelines for adapting psychological tests. *European Journal of Psychological Assessment, 15,* 270–276. http://dx.doi.org/10.1027//1015-5759.15.3.270

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning. Data mining, inference, and prediction* (2nd ed.). New York, NY: Springer-Verlag.

Heine, S. J., Lehman, D. R., Peng, K., & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of Personality and Social Psychology, 82,* 903–918. http://dx.doi.org/10.1037/0022-3514.82.6.903

Hornik, K., Feinerer, I., Kober, M., & Buchta, C. (2012). Spherical k-Means clustering. *Journal of Statistical Software, 50,* 1–22.

Houle, M. E., Kriegel, H. P., Kröger, P., Schubert, E., & Zimek, A. (2010, June–July). *Can shared-neighbor distances defeat the curse of dimensionality?* Paper presented at the Scientific and Statistical Database Management International Conference, Heidelberg, Germany.

ITC. (2010). *International test comission guidelines for translating and adapting tests*. Retrieved from http://www.intestcom.org

Jeanrie, C., & Bertrand, R. (1999). Translating tests with the international test commission's guidelines: Keeping validity in mind. *European Journal of Psychological Assessment, 15,* 277–283. http://dx.doi.org/10.1027//1015-5759.15.3.277

Ketterer, H. L., Han, K., Hur, J., & Moon, K. (2010). Development and validation of culture-specific variable response inconsistency and true response inconsistency scales for use with the Korean MMPI-2. *Psychological Assessment, 22,* 504–519. http://dx.doi.org/10.1037/a0019511

Kuppens, P., Realo, A., & Diener, E. (2008). The role of positive and negative emotions in life satisfaction judgment across nations. *Journal of Personality and Social Psychology, 95,* 66–75. http://dx.doi.org/10.1037/0022-3514.95.1.66

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Li, W. (1992). Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory, 38,* 1842–1845. http://dx.doi.org/10.1109/18.165464

Lynam, D. R., Derefinko, K. J., Caspi, A., Loeber, R., & Stouthamer-Loeber, M. (2007). The content validity of juvenile psychopathy: An empirical examination. *Psychological Assessment, 19,* 363–367. http://dx.doi.org/10.1037/1040-3590.19.3.363

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. London, UK: MIT Press.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1988). *Multivariate analysis*. London, UK: Academic Press.

McCrae, R. R. (2001). Trait psychology and culture: Exploring intercultural comparisons. *Journal of Personality, 69,* 819–846. http://dx.doi.org/10.1111/1467-6494.696166

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509–516. http://dx.doi.org/10.1037/0003-066X.52.5.509

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM, 38,* 39–41. http://dx.doi.org/10.1145/219717.219748

Moscoso, M. S., & Spielberger, C. D. (2011). Cross-cultural assessment of emotions: The expression of anger. *Revista de Psicología, 29,* 343–360.

Norenzayan, A., & Heine, S. J. (2005). Psychological universals: What are they and how can we know? *Psychological Bulletin, 131,* 763–784. http://dx.doi.org/10.1037/0033-2909.131.5.763

Ozolins, U. (2009). Back translation as a means of giving Trans. a voice. *The International Journal for Translation & Interpreting Research, 1,* 1–13.

Peng, K., Nisbett, R. E., & Wong, N. Y. C. (1997). Validity problems comparing values across cultures and possible solutions. *Psychological Methods, 2,* 329–344. http://dx.doi.org/10.1037/1082-989X.2.4.329

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review, 21,* 1112–1130. http://dx.doi.org/10.3758/s13423-014-0585-6

Poortinga, Y. H., & Van Hemert, D. A. (2001). Personality and culture: Demarcating between the common and the unique. *Journal of Personality, 69,* 1033–1060. http://dx.doi.org/10.1111/1467-6494.696174

R Development Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Saucier, G., & Goldberg, L. R. (2001). Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of Personality, 69,* 847–879. http://dx.doi.org/10.1111/1467-6494.696167

Schmitt, D. P., Realo, A., Voracek, M., & Allik, J. (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *Journal of Personality and Social Psychology, 94,* 168–182. http://dx.doi.org/10.1037/0022-3514.94.1.168

Schwartz, S. J., Benet-Martínez, V., Knight, G. P., Unger, J. B., Zamboanga, B. L., Des Rosiers, S. E., . . . Szapocznik, J. (2014). Effects of language of assessment on the measurement of acculturation: Measurement equivalence and cultural frame switching. *Psychological Assessment, 26,* 100–114. http://dx.doi.org/10.1037/a0034717

Shrout, P. E. (2009). Short and long views of integrative data analysis: Comments on contributions to the special issue. *Psychological Methods, 14,* 177–181. http://dx.doi.org/10.1037/a0015953

Sireci, S. G., Yang, Y., Harter, J., & Ehrlich, E. J. (2006). Evaluating guidelines for test adaptations. A methodological analysis of translation quality. *Journal of Cross-Cultural Psychology, 37,* 557–567. http://dx.doi.org/10.1177/0022022106290478

Spielberger, C. D. (2006). Cross-cultural assessment of emotional states and personality traits. *European Psychologist, 11,* 297–303. http://dx.doi.org/10.1027/1016-9040.11.4.297

Steele, G. I., & Edwards, D. J. A. (2008). Development and validation of the Xhosa translations of the Beck Inventories: 1. Challenges of the translation process. *Journal of Psychology in Africa, 18,* 207–216.

Su, C.-T., & Parham, L. D. (2002). Generating a valid questionnaire translation for cross-cultural use. *American Journal of Occupational Therapy, 56,* 581–585. http://dx.doi.org/10.5014/ajot.56.5.581

Suleiman, K. H., Yates, B. C., Berger, A. M., Pozehl, B., & Meza, J. (2010). Translating the Pittsburgh Sleep Quality Index into Arabic. *Western Journal of Nursing Research, 32,* 250–268. http://dx.doi.org/10.1177/0193945909348230

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,* 87–95. http://dx.doi.org/10.1002/mpr.83

Tanzer, N. K., & Sim, C. Q. E. (1999). Adapting instruments for use in multiple languages and cultures: A review of the ITC guidelines for test adaptations. *European Journal of Psychological Assessment, 15,* 258–269. http://dx.doi.org/10.1027//1015-5759.15.3.258

Van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1,* 89–99. http://dx.doi.org/10.1027/1016-9040.1.2.89

Van de Vijver, F. J. R., & Leung, K. (2001). Personality in cultural context: Methodological issues. *Journal of Personality, 69,* 1007–1031. http://dx.doi.org/10.1111/1467-6494.696173

Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 54,* 119–135. http://dx.doi.org/10.1016/j.erap.2003.12.004

Weiss, S., Indurkhya, N., Zhang, T., & Damerau, F. (2004). *Text mining: Predictive methods for analyzing unstructured information*. Berlin: Springer-Verlag.

Wilkinson, L., & Friendly, M. (2009). The history of the cluster heat map. *The American Statistician, 63,* 179–184. http://dx.doi.org/10.1198/tas.2009.0033

Witten, I. H., & Frank, E. (2005). *Data mining. Practical machine learning tools and techniques* (2nd ed.). San Francisco, CA: Elsevier.

Zipf, G. K. (1945). The repetition of words, time-perspective, and semantic balance. *Journal of General Psychology, 32,* 127–148. http://dx.doi.org/10.1080/00221309.1945.10544486

Received April 10, 2015
Revision received June 26, 2015
Accepted July 7, 2015 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.