

Developing Multiple Language Versions of Instruments for Intercultural Research

Sumru Erkut

Wellesley Centers for Women

ABSTRACT—*This article examines the strengths and weaknesses of several translation techniques currently in use through the lens of emerging opinions on the science and ethics of intercultural research. Broad scientific and ethical dimensions relevant to translating instruments and a distinction between generating multiple language forms of two kinds of instruments are introduced: those in which wording in the source language cannot be altered and those in which constraints of the target language can lead to changes in the original instrument's wording. Developmental psychologists engaged in intercultural research can consider techniques for minimizing the influence of Western perspectives while pursuing conceptual equivalence in order to satisfy science's concern for internal validity of translated instruments.*

KEYWORDS—*translation; instruments; equivalence; Western bias; measures; intercultural research*

I situate the examination of different techniques for developing multiple language versions of instruments within the broader scientific and ethical concerns of intercultural research (Gergen, Gülerce, Lock, & Misra, 1996; Marshall & Batten, 2003; Moghaddam, 1987). The social sciences have long understood the need for intercultural researchers to be concerned with the potential for undue influence of Western perspectives on their research (see, e.g., Geertz, 1983; Kağıtçıbaşı, 1984). However, techniques for minimizing the influence of Western perspectives are still not widespread among developmental psychologists. I

review translation techniques for intercultural research within this broader concern.

SCIENTIFIC AND ETHICAL CONCERNS

The scientific issue in translating instruments is avoiding one of the threats to internal validity, commonly referred to as “instrumentation” (Campbell & Stanley, 1963; Shadish, Cook, & Campbell, 2002). The instrumentation threat can occur if different respondents receive a different version of the measure, making it invalid to infer that differences in answers are owing to the respondents’ characteristics because the differences in the versions of the instrument are a viable alternative explanation. Attempts to achieve equivalence in different language versions of a measure address this instrumentation threat. For example, if a vocabulary test for young children were to use the word *hair* in the English version and translate it as *cabello* into Spanish for use with Puerto Rican children, the two language versions would have different levels of difficulty, thereby introducing an instrumentation threat. This is because whereas *hair* is widely used in English, the average Puerto Rican refers to hair as *pelo*. *Cabello* is more common among well-educated Puerto Ricans, which would raise concerns for a social class bias as well, an ethical concern if the vocabulary test were to be used to place children in academic tracks.

Taking on the scientific challenge translation poses to validity, Peña (2007) has framed the ethical issue in terms of fairness. The American Educational Research Association’s (1999) definitions of fairness, articulated in *Standards for Educational and Psychological Testing*, include the notion of equal treatment in context and purpose of testing and comparable opportunity for all undergoing testing to demonstrate their abilities on the construct the test is intended to measure. If the translated version is different, scores on the test are not comparable across different language versions, and using such scores to make educational decisions violates the principles of equal treatment and comparable opportunity. Arguing that it is possible to apply these

I thank my colleagues, Odette Alarcón, Cynthia García Coll, Linda R. Tropp, and Heide A. Vázquez García, for their past and future collaboration.

Correspondence concerning this article should be addressed to Sumru Erkut, Wellesley Centers for Women, 106 Central Street, Wellesley, MA 02481; e-mail: serkut@wellesley.edu.

© 2010, Copyright the Author(s)

Journal Compilation © 2010, Society for Research in Child Development

principles to intercultural research, Peña has made an important contribution to the developmental literature on translating instruments into languages other than English. She draws attention to the need to go beyond linguistic equivalence to include functional equivalence, cultural equivalence, and metric equivalence to improve internal validity.

I will not review Peña's (2007) work but invite readers to read the original. Rather, I will expand on the fairness framework to encompass an ethical concern for avoiding a cultural bias when producing multiple language versions of instruments for intercultural research.

Potential for Western Bias

There is a potential for bias when researchers from one language or culture group wish to measure some aspect of the psychological development of the members of a different group by using a translation of an instrument developed in the researchers' culture. Foucault (1975) commented on the "clinical gaze" to draw attention to power differences between the observer and observed, whereby the clinician views the observed through a lens that reflects the history, norms, and economic circumstances of the observer's culture. Gergen et al. (1996) have argued that when Western concepts and methods guide research, the resulting product can be of little relevance to other cultures and may disregard and undermine alternate cultural traditions. Greenfield (1994) explained this phenomenon as the product of psychologists' familiarity with their own culture when she argued that psychologists tend to base their intercultural research on an implicit understanding of the culture in which they grew up. Rogler (1999) has suggested that these unexamined insiders' perspectives often become the basis for norms, in that they can set the standard for what is studied in other culture groups and how it is studied.

I contend that to minimize the influence of Western perspectives, generating multiple language versions of an instrument can begin with examining the motivation for the research. Researchers need to be able to answer *why* they are pursuing their research goals. It may be easy to answer the "why" question with, "We want to compare . . .," but I urge caution with research questions and hypotheses that can lead to invidious comparisons, which have been the fodder of the much criticized deficiency model (e.g., García Coll & Magnuson, 1997; Kağıtçıbaşı, 2007). The deficiency model refers to studies whose results "explain" why members of less powerful culture groups (such as Third World societies, indigenous populations of First World societies, immigrants, and minorities) are deficient in some aspect of growth and development. Paraphrasing Foucault (1975), when we privilege one culture or language as the *source* and the other as the *target*, we give primacy to the source culture's history, norms, and economic circumstances. One useful heuristic for not falling into an unintended deficiency paradigm is to ask what aspects of development would members of the "other" culture group deem important to

study? Psychologists from other cultures rarely if ever "study" North Americans. Consider an example from the practice of age mixing in education, which is more widespread in Russia than in the United States. I believe there would be resistance if Russian psychologists came to the United States with an English translation of their instruments for a comparative study of the impact of younger students learning from older students. Similar to the faults of the deficiency model, people might feel this is a setup to highlight the superiority of a Russian pedagogical practice.

Horizontal Collaboration

What is a conscientious researcher to do? It is not a trivial matter that financial resources for research reside mostly in the West (Moghaddam, 1987), and within Western societies they are more available to members of the educated elite of the dominant White culture group. The answer can be found in nonhierarchical "horizontal collaboration," which Sinha (1984) proposed to manage one culture's domination of the others.

Horizontal collaboration requires researchers from each culture and language group who come together to jointly decide on what constructs to research. Indigenous coleaders, who are full members of the team, can provide a safeguard against the unexamined exportation of ideas and methods because people from the cultures under study take a leading role in defining the goals and methods of the study. It is important for the collaborative research team to examine the constructs underlying the instruments to be translated. Although most discussions of translating instruments focus on item wording to achieve equivalence, a more fundamental concern is whether their conceptual foundations have comparable relevance with development in the cultures under study. Substantive problems can occur as a result of an unexamined transfer of constructs and concepts from one culture and language system to another. For example, the Japanese and Western constructions of the "self" are not strictly comparable (DeVos, 1985). This need to focus on constructs is reinforced by the long tradition in psychometrics that gives constructs a primary role in validity studies (Campbell & Fiske, 1959). If a serious examination of the constructs in the cultures to be studied reveals that the underlying concepts are not equivalent, the research need not be abandoned. Rather, the research questions can be revised. In such cases, a worthy research question may be what social, cultural, and physical environmental conditions have given rise to different conceptualizations in the different language groups.

A COMPARISON OF METHODS FOR GENERATING MULTIPLE LANGUAGE VERSIONS OF INSTRUMENTS

Direct, one-way translation is the most basic approach, but it is not recommended as a technique for translating instruments; I do not include it in Table 1, which presents the characteristics of alternative translation techniques.

Table 1
Characteristics of Translation Methods

	Who “translates”	Attention to minimizing cultural bias	Target language(s) can alter source language	Underlying constructs examined	Suitable for what uses?
Single methods					
Back translation	Professional translators	No	No	No	Instruments with established history in source language
Multiple-forward translation	Professional translators and experts in the subject	No	No	No	Instruments with established history in source language
Combined methods					
Back translation with decentering	Professional translators; can include experts in the subject	No	Yes	Sometimes	New instruments
Back translation with decentering and multiple-forward translation	Professional translators and experts in the subject	No	Yes	Sometimes	New and established instruments
Dual focus: New instruments	Experts in the subject	Yes	Yes	Yes	New instruments
Dual focus: Existing instruments	Experts in the subject	No	No	Yes	Instruments with established history in source language

Back Translation

When researchers want to go beyond direct translations, back translation (Brislin, 1970, 1986) is currently the most widely used technique. The back translation method works as follows: To create a Portuguese version of a measure originally developed in English, one person (or a team of translators) translates from English into Portuguese, and a different person (or a team of translators) translates from Portuguese back into English. It is recommended to use several iterations of back translation until the last back translation matches the source language. Because the translation centers on the source language, which remains unchanged, this approach is most appropriate for translating established instruments that have a long history of use in the source language. Back translation has had its detractors (see Bontempo, 1993; Olmedo, 1981). Maxwell (1996) provides a compelling example of the potential pitfalls of relying solely on back translation in the following item on a science test.

In the question, “What does a carnivore eat?” the word “carnivore” would read “meat-eater” in many translations, making the questions very much easier. But if in the back translation “meat-eater” was translated back to “carnivore,” one would not know about the flawed original translation. (p. 6)

Back translation’s main weaknesses include the absence in the process of input from researchers knowledgeable about the subject matter, lack of provisions for examining whether underlying constructs are equivalent in the cultures being studied, and failure to consider the interface of the potential for bias and scientific issues.

Back Translation With Decentering

This technique begins with back translation from the source to the target language and back. Discrepancies between the source

and back-translated versions are dealt with through “decentering.” The instrument is decentered or moved away from the idiosyncrasies of the source language by subjecting both the source and target language versions to modification through a process of several iterations (Werner & Campbell, 1970). One example of decentering is the hypothetical item, “able to meet deadlines,” from a hypothetical measure of attentional processes. In Turkish, a literal translation would be “ölüm çizgisi ile buluşma yeteneğine sahiptir.” This can be back translated as “has the ability to get together with the line of death,” indicating a serious need for decentering. The appropriate rendering in Turkish requires specifying what the deadline is for. Is it homework, a job, or a task? If it is homework, “able to meet deadlines” can be approximated in Turkish with a phrase that back translates as “finishes homework on time.” At this point in the decentering process, the translators become aware that the Turkish version has dropped “able to.” They debate whether “finishes on time” has the same meaning as “able to meet deadlines.” The next iteration might be to add words to the Turkish version to recapture “able to.” They can try, “Ödevini vaktinde bitirebilme yeteneğine sahiptir.” This rendition back translates into “has the ability to finish homework on time.” Although grammatically correct, the Turkish version is awkward. Translators may experiment with a different wording that back translates into “always finishes homework on time,” and the iterations will continue until the translators are satisfied. When they are satisfied, we have a case where the idiosyncrasies of both the source and target languages have led to changes in the other.

Compared with back translation, decentering is more likely to yield functionally equivalent instruments. It is better suited to translate new instruments because decentering makes this technique unsuitable for translating established measures when researchers feel compelled to preserve the original wording in

the source language. It shares with the back translation method the absence of provisions for input from bilingual experts knowledgeable on the topic, the failure to examine underlying constructs, and the lack of consideration for the interface of cultural bias and scientific issues.

Multiple-Forward Translation

This technique, also called the committee method (Nasser, 2005), involves several bilingual individuals who work independently to translate an instrument from the source into the target language. A committee consisting of translators and researchers deals with discrepancies. The committee method has the advantage of bringing together individuals with language expertise and researchers with expertise in the topic. They work together to make informed decisions about whether the chosen words in the target language have the same connotations as the words in the source language. Larkin, Dierckx de Casterlé, and Schotsmans (2007) recommend a similar collaborative approach for qualitative research to generate multiple language versions of interview questions.

Joint Use of Multiple Methods

State-of-the-art approaches to translation tend to incorporate several techniques. For example, for the Third International Mathematics and Science Study, Maxwell (1996) reported that their tests have been translated into 31 languages using at least two and often three of the following techniques: multiple-forward translation, back translation, translation review by bilingual judges, and item response theory procedures. The World Health Organization (n.d.) also recommends an approach that makes use of forward and back translations, input from experts, pretesting with target populations, and interviewing members of the target group about alternative wordings and use of expressions. Other examples of the use of multiple methods include the linguistic validation method (Mapi Research Institute, n.d.) and the International Quality of Life Assessment Project method (Bullinger et al., 1998). However, none of these rigorous approaches explicitly pays attention to the interrelated concerns of scientific merit and cultural bias.

An Alternative Approach for Developing Multiple Language Versions of Instruments

In the foregoing discussion, I have argued that translation from the source to the target language is unlikely to produce valid translations of instruments for intercultural research, even when using multiple techniques, because collaborations, research questions, and study design are relevant to the process of generating multiple language versions of measures. My colleagues and I have developed an alternative method, the dual-focus approach (Erkut, Alarcón, García Coll, Tropp, & Vázquez García, 1999), which can be used to generate two or more language versions of a new instrument simultaneously.

In the dual-focus approach, experts in the subject matter—including bilingual and bicultural native speakers of the culture

and language systems under study—work as a team of horizontal collaborators. Team members jointly decide on the research questions, the constructs to be measured, and how best to measure the constructs in different languages. A distinguishing feature of this approach is the involvement of members of the research team in all aspects of generating different language versions of an instrument, including choosing the appropriate wording in each language. Full participation of bilingual and bicultural experts who are native speakers of the culture and language systems to be studied decreases not only the instrumentation threat but also the possibility of undue influence of Western perspectives. This approach is an alternative to translation techniques in that the different language versions of an instrument are not translated from a source to target languages. Rather, the focus is on all language versions under development simultaneously (hence the name *dual-focus approach* when there are only two languages involved). Decentering comes closest to the dual-focus approach but fails to call for the horizontal collaboration of bilingual and bicultural subject experts in all aspects of the research leading up to and including the wording of the instruments.

The steps in the implementation of the dual-focus approach are as follows:

- Step 1. Formation of a research team that includes bilingual and bicultural professionals from the culture and language systems to be studied who have expertise on the research topic.
- Step 2. Team members reach consensus on conceptual equivalence of constructs, a process that can inform reformulation of the original constructs to be studied.
- Step 3. Team members jointly generate items to measure the constructs.
- Step 4. The team obtains external input from monolingual and bilingual members of the communities for whom the measure is intended in an iterative process of revision and more external input.
- Step 5. Final drafts of the measure are piloted with members of the intended research participants.
- Step 6. Different language versions of the instrument's psychometric properties are evaluated.

In the first step, inclusion of bilingual and bicultural topic experts as peers on the research team brings the advantages of guarding against an undue Western influence in deciding on the research questions. Bilingual and bicultural experts on the topic area have the cultural and linguistic background necessary for examining equivalence in constructs and are able to judge whether the chosen words adequately reflect the constructs. Lay translators can and do bring valuable language expertise to translations but are limited by their lack of familiarity with the theories and constructs under study.

In the third step, when team members jointly decide on the wording of items, say, for an instrument to be used with Puerto Rican immigrants on the U.S. mainland, "How would we say it

in Spanish?” and “How would we say it in English?” initiate the discussion. Team members determine wording for each item simultaneously in both languages and then examine the wording of each item to see if it has the same level of *difficulty*, the same *affect*, and the same *clarity* of meaning in both languages. An example of nonequivalence because of differences in affect can be seen in the words *pain* in English and *dolor* in Spanish. Whereas pain refers to hurting primarily because of a physical and secondarily to an emotional injury, *dolor* indicates emotional and physical pain equally and encompasses sorrow and the sadness of regret. Awareness of this difference forces the researchers to examine whether emotional and/or physical pain is the intended meaning, leading to changes in either language version to convey the intended meaning. Therefore, much like in the process of decentering, the constraints of one language are just as likely to influence choice of final wording as the constraints of the other language. In effect, both Spanish and English become target languages although the conceptual base serves as the source.

In the fourth step, the team seeks feedback from both monolingual and bilingual members of the community under study, especially if the targeted sample will comprise monolingual speakers. Whereas the usefulness of feedback from bilinguals for evaluating language equivalence is widely accepted (see Streiner & Norman, 1995), the importance of feedback from monolingual informants is less well recognized (see Hulin, 1987). Monolinguals' input is important because their speech has not been influenced by the mastery of a second language.

A weakness of the dual-focus method is that in its original formulation it is best suited for generating multiple language versions of a new instrument. Another concern with this approach can be the difficulty of finding bilingual and bicultural experts in the subject matter. Bilingual and bicultural graduate students in the department from which the research emanates can approximate the input of experts. Employing graduate students is preferable to working with professional translators because of the importance of examining conceptual equivalence; with professional translators, the collegial debate among peers on what to study and how to study it will be lost. If researchers lack access to either bilingual and bicultural experts or graduate students in their university or the possibility of collaborations with researchers from another university, Larkin et al.'s (2007) approach of researchers collaborating with translators can be an alternative. In this approach, researchers and professional translators work together to arrive at wordings that best capture the researchers' intended meanings.

Dual-Focus Approach to Translating Existing Instruments

Although it was designed for generating multiple language versions of new instruments, the dual-focus method can be adapted to translate existing measures whose language cannot be altered. In these cases, a team of researchers, some of whom are bilingual and bicultural, come together to examine the instrument in the

source language. Researchers scrutinize what each item in the source instrument is trying to assess in light of the operational definition of the construct it is measuring. These conversations delve into the original instrument developer's intentions in the choice of wording of the items in the source language. Informed by these discussions, one of the bilingual and bicultural members of the research team produces a first draft in the target language. The research team then scrutinizes each item to see if the translated version expresses the same idea as the original in terms of clarity, difficulty, and affect. Members of the research team who are not bilingual or bicultural can help this process by asking questions about multiple meanings of the words to get at clarity, whether people from different educational levels can understand the wording to get at difficulty, and whether the affect associated with the words have similar meaning or, if not, whether the words need to be qualified to make the meaning more precise. These discussions result in producing a second draft, which is vetted in a focus group made up of members of the target population. The research team reviews suggested revisions and produces a third draft. Focus group input and research team's revisions continue until no more changes are suggested. At that point, the translated instrument is ready for psychometric testing.

In conclusion, I presented the specific steps of the dual-focus approach not as an exemplar of translation techniques for developing multiple language versions of instruments, but rather as a method that pays particular attention to minimize the influence of Western perspectives while maintaining a scientific concern for internal validity. Table 1 provides a summary of the different features of the methods I have mentioned. Guided by their research needs, researchers can mix and match different approaches to suit their purposes.

REFERENCES

- American Educational Research Association. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bontempo, R. (1993). Translation fidelity of psychological scales: An item-response theory analysis of an individualism-collectivism scale. *Journal of Cross-Cultural Psychology, 24*, 149–166.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 185–216.
- Brislin, R. W. (1986). Back-translation methods: The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Cross-cultural research methodology series: Vol. 8. Field methods in cross-cultural psychology* (pp. 137–164). Beverly Hills, CA: Sage.
- Bullinger, M., Alonso, J., Apolone, G., Leplege, A., Sullivan, M., Wood-Dauphinee, S., et al. (1998). Translating health status questionnaires and evaluating their quality: The IQOLA project approach. *Journal of Clinical Epidemiology, 51*, 913–923.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81–105.

- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.
- DeVos, G. (1985). Dimensions of the self in Japanese culture. In A. J. Marcella, G. DeVos, & F. L. K. Hsu (Eds.), *Culture and self: Asian and Western perspectives* (pp. 141–184). New York: Tavistock.
- Erkut, S., Alarcón, O., García Coll, C., Tropp, L. R., & Vázquez García, H. A. (1999). The dual-focus approach to creating bilingual measures. *Journal of Cross-Cultural Psychology*, 30, 206–218.
- Foucault, M. (1975). *The birth of the clinic: An archeology of medical perception* (A. M. Sheridan Smith, Trans.). New York: Vintage.
- García Coll, C. T., & Magnuson, K. (1997). The psychological experience of immigration: A developmental perspective. In A. Booth (Ed.), *Immigration and the family: Research and policy on US immigrants* (pp. 91–131). Hillsdale, NJ: Erlbaum.
- Geertz, C. (1983). *Local knowledge: Further essays in interpretive anthropology*. New York: Basic Books.
- Gergen, K. J., Gülerce, A., Lock, A., & Misra, G. (1996). Psychological science in cultural context. *American Psychologist*, 51, 496–503.
- Greenfield, P. M. (1994). Independence and interdependence as developmental scripts: Implications for theory, research, and practice. In P. M. Greenfield & R. R. Cocking (Eds.), *Cross-cultural roots of minority child development* (pp. 1–37). Hillsdale, NJ: Erlbaum.
- Hulin, C. L. (1987). Psychometric theory of evaluations of item and scale translations—Fidelity across languages. *Journal of Cross-Cultural Research*, 18, 115–142.
- Kağıtçıbaşı, Ç. (1984). Socialization in a traditional society: A challenge to psychology. *International Journal of Psychology*, 19, 145–157.
- Kağıtçıbaşı, Ç. (2007). *Family, self, and human development across cultures: Theory and applications*. New York: Routledge.
- Larkin, P. J., Dierckx de Casterlé, B., & Schotsmans, P. (2007). Multilingual translation issues in qualitative research. *Qualitative Health Research*, 17, 468–476.
- Mapi Institute. (n.d.). *Linguistic validation methodology*. Retrieved January 8, 2010, from <http://www.mapi-institute.com/linguistic-validation/methodology>
- Marshall, A., & Batten, S. (2003). Ethical issues in cross-cultural research. In W.-M. Roth (Ed.), *Connections '03* (pp. 139–151). Victoria, Canada: University of British Columbia.
- Maxwell, B. (1996). Translation and cultural adaptation of the survey instruments. In M. O. Martin & D. L. Kelly (Eds.), *Third International Mathematics and Science Study (TIMSS) Tech. Rep.: Vol. I. Design and development* (pp. 1–9). Chestnut Hill, MA: Boston College.
- Moghaddam, F. M. (1987). Psychology in the three worlds: As reflected by the crisis in social psychology and the move toward indigenous third-world psychology. *American Psychologist*, 42, 912–920.
- Nasser, R. (2005). A method for social scientists to adapt instruments from one culture to another: The case of the Job Description Index. *Journal of Social Sciences*, 1, 232–237.
- Olmedo, E. L. (1981). Testing linguistic minorities. *American Psychologist*, 36, 1078–1085.
- Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, 78, 1255–1264.
- Rogler, L. H. (1999). Methodological sources of cultural insensitivity in mental health research. *American Psychologist*, 54, 424–433.
- Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sinha, J. (1984). Toward partnership for relevant research to the Third World. *Indian Journal of Psychology*, 19, 169–178.
- Streiner, D. L., & Norman, G. R. (1995). *Health measurement scales* (2nd ed.). Oxford, UK: Oxford University Press.
- Werner, O., & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A handbook of method in cultural anthropology* (pp. 398–420). New York: The Natural History Press.
- World Health Organization. (n.d.). *Process of translation and adaptation of instruments*. Retrieved June 21, 2008, from http://www.who.int/substance_abuse/research_tools/translation/en/index.html

Copyright of Child Development Perspectives is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.