Cross-cultural influences affect people’s perceptions and health practices and are areas of concern and study for nurses and other health professionals. Given the cultural diversity of many countries, culturally sensitive assessment methods are needed, but many challenges exist in obtaining valid and reliable measurement. Translating questionnaires for cross-cultural research is fraught with methodological pitfalls related to colloquial phrases, jargon, idiomatic expressions, word clarity, and word meanings. It cannot be assumed that a particular concept has the same relevance across cultures. Simply translating an English version word-for-word into another language is not adequate to account for linguistic and cultural differences. Ideally, the perspectives of people from the culture about the concept of interest should be studied first, but often a practical alternative is to find and translate a tool developed in another culture. The purpose of this article is to describe important considerations in conducting translation for equivalence, types of equivalence, and strategies to translate instruments that promote equivalence and how to test the translated version for equivalence. These concepts and strategies are illustrated by describing the translation process of Hilton’s Uncertainty Stress Scale into French and the use and testing of the French version with a French Canadian sample in Skrutkowski’s study of perceived uncertainty in adult survivors of cancer.
wording, and word meanings commonly used in English affect validity. Some flaws are difficult to detect, leading to erroneous conclusions that cultural differences are substantive when, in fact, they stem from inconsistencies in meaning. It cannot be assumed that a particular concept has the same relevance across cultures. Developing a culturally equivalent translated instrument requires familiarity with basic problems of linguistic adaptation, cultural constructs, and psychometric changes inherent in the translation process. Simply translating an English version word-for-word into another language is not adequate to account for linguistic and cultural differences.

Ideally, measurement of a concept should be done from the perspective of the cultures under investigation, with a preliminary qualitative study with people from that culture. Another method uses a dual-focus approach. A bilingual/bicultural team, with researchers indigenous to the cultures, collaboratively decides on concepts that provide equally valid definitions and jointly generate items. Items must be directly and easily understood with parallel wording in the other language. Wording is examined to ensure similar difficulty, affect, and meaning clarity. Revisions are based on feedback from monolingual and bilingual community members. Unfortunately, the resources and time required may render these approaches as unfeasible. A practical alternative is to find and translate a tool developed in another culture. The intent is to develop another version of the scale with equivalence to the original instrument. The critical need is to consider meaning and not literal word-for-word translation.

Skrutkowski investigated uncertainty in long-term survivors of cancer. She examined how physical status and beliefs about control related to uncertainty; how uncertainty was associated with depression, illness, worry, and sense of control; and the relationships among personal factors, uncertainty, and psychologic responses. The study involved English-speaking and French-speaking Canadian people and required a French version of the Uncertainty Stress Scale (USS) that preserved its meaning. Skrutkowski worked with Hilton in the translation and testing of this scale. The purpose of this paper is therefore to describe important considerations in conducting translation and testing for equivalence, and to illustrate these processes by describing the translation and testing of the USS into French.

### Translation and Equivalence

Werner and Campbell defined 2 categories of translation: symmetrical and asymmetrical. Symmetrical translation requires the original and translated instruments to be equally familiar and to have loyalty of meaning and colloquialness. Items cannot be specific to only one culture. Asymmetrical translation emphasizes loyalty to one language, usually the original language. The translated version therefore is often unnatural in the new language, and problems can arise if the original instrument has a history of use and the developer resists altering the content as items are translated. The original reliabilities may be compromised if items need to be changed to achieve conceptual equivalence.

The 5 types of cross-cultural equivalence are described as content, semantic, technical, criterion, and conceptual. Content equivalence means that each item’s content is relevant in each culture; however, some constructs cannot be insinuated into instruments for other cultures. For example, items referring to government programs need to reflect the social structure and healthcare delivery system of the nation where the instrument is employed. Failure to do this makes interpretation of results questionable. Semantic equivalence emphasizes similarity of meaning of each item in each culture after translation. Even with appropriate processes, concerns may relate to regional or national idiomatic differences. Differences due to ethnic variation may arise between same-language versions from different regions. Technical equivalence means that the data collection method (eg, pencil and paper, interview) is comparable. Criterion equivalence means that the interpretation remains the same when compared with the norm for each culture. It refers to the instrument’s relationship to established independent criteria for the same phenomena. Conceptual equivalence means that the instrument measures the same theoretical construct in each culture. The usual method for examining conceptual equivalence is to assess the relationship between constructs and to compare it with known and predicted relationships.

### Strategies for Enhancing Equivalence of a Translated Version

The aim of translation is to achieve equivalence. Guidelines for enhancing equivalence when developing instruments that may later be translated (Figure 1) and guidelines for translating existing instruments using one-way translation, translation by committee, or double or back translation have been created.

When using one-way translation, which is the most frequently used method because of its simplicity and cost, a bilingual person translates the original version into the other language. Total dependence on the translator’s skill and knowledge often results in low validity and reliability, and even with skilled and experienced translators, results may still be poor. Using the committee approach, 2 or more bilingual people work separately or collaboratively on the translation. Serious limitations arise if committee members have common views or if pressure is felt to form a consensus, something that might be evident when using bilingual graduate students or faculty colleagues.

The preferred back-translation approach requires at least 2 independent translators. Multistage procedures are recommended. The first translator works independently to produce a translated version. A second translator translates the translated version back to the original language. Both interpreters are consulted to identify discrepancies, and adjustments are made for inconsistencies. If the original and back-translated versions are identical, the translated version is likely equivalent in meaning. Giving instructions to the translators regarding inference, wording, and phrasing and emphasizing translation for meaning promotes conceptual equivalence. Translators should identify words that reflect different connotations or phrases that are awkward when translated back into
Translating Instruments

English. Although preferred, this approach is time-consuming and impractical for multilingual studies, and even with care and diligence, differences may exist between the original and back-translated versions. Keeping grammatical forms intact may result in a confusing and awkwardly phrased translation. Variation may include interpretations of items that may be culture-specific, and cultural differences in demographics may influence score variability.

Therefore, the translators influence the translation quality, regardless of method. Experts and interpreters should include professional interpreters, lay people who are monolingual and representatives of the populations under study, people who are bilingual with the source language as their first language, and people who are bilingual with the target language as their first language. Some content areas can be challenging, and translators are not always knowledgeable about the content. If 2 translators share a common world view but a different view from those targeted for instrument use, they may develop identical versions but still miss the original intent. Experienced translators may infer the original meaning even though the translation is poor. Problems can also arise if translators are not sufficiently aware of the rigorous requirements of cross-cultural translation.

Differences in instrument formatting and administration can produce variations in response even if meticulous care is taken. For example, the cultural and linguistic differences between interviewer and respondents can bias results, and self-report instruments may wrongly assume literacy. The format may be foreign to some cultures. Some groups may be more prone to respond in a moderate way, whereas others may be prone to extreme responses. Answering questions by choosing from arbitrary categories may also be foreign in many countries.

Field testing and refining the instrument with members of the target culture is then needed. This serves as a check for translation quality and the practical aspects of test administration. People are asked to read or listen to the items, item-by-item, and asked to paraphrase their understanding. Their responses should closely resemble the original version. The instrument should then be pretested with a small sample.

**Figure 1** Guidelines for developing new instruments that may later be translated.

- Use (or revise) questionnaire written to an easy level of difficulty (third-grade English)
- Keep sentences simple and short (16 or fewer words)
- Repeat nouns rather than using pronouns
- Avoid possessive forms when possible
- Avoid metaphors and idiomatic phrases
- Use specific rather than general terms (eg, measles rather than general term diseases)
- Avoid sentences with two different verbs if the verbs suggest different actions
- Avoid words open to wide interpretation (eg, “often,” “probably,” or “beyond”)
- Avoid the subjunctive mode (eg, verbs using “would” or “could”)

**Strategies for Reflecting and Testing Equivalence**

Several methods support equivalence: the more rigorous and inclusive the process, the more likely the translation supports equivalence. A full description of the process is needed, including consideration of format, administration, and translator selection. Other strategies include expert review, response comparisons, and psychometric examination. Experts are asked to review the translation for clarity and linguistic appropriateness.

Comparing item and total scores by 1 or 2 monolingual groups who complete the original English version and the back-translated version is recommended as is comparing responses by bilinguals on the English and translated versions. Items yielding different scores should be examined and reformulated. High correlation between individual responses on both instruments is desirable. Item analysis should be performed because it may indicate that specific items perform differently on the two forms and total scores may hide inconsistency.

Using bilingual respondents can be problematic because they have likely adopted concepts, values, attitudes, and role expectations of the culture of the second language they mastered and may therefore differ from the monolingual population. Explanations have also been made to explain discrepancies when bilinguals respond to an instrument in their mother tongue and one in a foreign language. Presenting an instrument in a language different from the primary tongue arouses awareness of the person’s ethnicity and causes a more extreme (more ethnic) response than answering the instrument in the primary language. Giving a response that is appropriate in the other culture might also explain discrepancies: this is referred to as cross-cultural accommodation.

Claims for translation adequacy are also based on similarity in reliability and validity to the original version. Internal consistency coefficients can be compared. According to Nunnally, translated versions can be treated like new instruments when an alpha of = 0.70 may be considered adequate.

Determining that a translated measure covers the same dimensions and in the same quantities is critical to interpretation and use of scores. Some authors recommend comparing factor structures if there are sufficient subjects or cluster analysis if numbers are insufficient. Gorsuch describes factor-analytic techniques to make these comparisons. Others argue against using traditional exploratory factor-analytic techniques and emphasize that cross-cultural samples must be large and representative of the population. Still others say that factorial structure replication across cultures is no guarantee that item measurement and theoretical structures are invariant across groups. Considering that original measures are not often developed using factor analytic methods, it is not reasonable to compare factor structures across cultures for evidence of equivalence.

**Strategies for Dealing With Open-ended Responses**

The previous strategies primarily relate to force-choice response formats. When instruments are administered verbally
or require open-ended responses, additional considerations are warranted. If responses are in written form, 2 bilingual translators should independently translate the response. If translated versions are not identical, the investigator should confer with both translators to determine reasons for the differences. When responses are verbal, it is ideal that they be audiotaped and the tapes transcribed. Two translators can then proceed as described. If taping cannot be done, the interviewer should write the exact words immediately after the interview and not attempt a rapid field translation. The language of the informant’s choice should be used when carrying out open-ended interviews, which makes translation even more challenging because of the volume and complexity of data.30

To illustrate the above, a researcher examined the influence of translation on reliability and validity by comparing themes that emerged from a qualitative study exploring women’s perceptions of factors influencing getting a Pap smear.30 In-depth interviews were done in Cantonese and then translated and transcribed into English independently by 2 translators. Another researcher transcribed them into Chinese. Grammatical style and lack of no-word equivalents added complexity. The categories and themes generated from the Chinese and English data sets were compared. No significant differences were identified.

Examples of Translated Instruments Pertinent to Oncology

Examples of instruments pertinent to cancer nursing that have been translated and described have been compiled for the reader’s interest as follows: the Ferrans and Powers Quality of Life Index—Cancer Version into Spanish, the Thai version of the State-Trait Anxiety Inventory for Children and the Child Medical Fear Scale, the Spanish version of the Health Screening Questionnaire related to breast cancer, the Chinese version of a scale to measure breast self-examination variables, and the Sexual Dimensions Instrument for Hispanic Women used with women being treated for gynecologic cancer.

Using the Uncertainty Stress Scale as an Example

The USS is a 54-item instrument designed to measure uncertainty in illness-related situations and the stress, threat, and positive feelings generated from the uncertain state. Based on a phenomenologic study, uncertainty was defined as a cognitive perceptual state that ranges from a feeling of just less than surety to vagueness that changes over time and is accompanied by threatening and/or positive emotions. Uncertainty is not being able to foretell the future; lacking clarity about the present; being in doubt; being undecided because things are not definite, clear-cut, or determined; not being able to rely, count, or depend on someone or something; and having a sense of vagueness about what to do, expect, know, and ask. Uncertainty generates feelings of anxiety, fear, anger, wonder, frustration, helplessness, curiosity, hope, and depression.

The scale has 3 parts: Part A consists of 54 items. Participants are asked to rank their level of uncertainty in a number of areas about their health condition and their coping. Each item is ranked on a 5-point scale from no uncertainty (0) to a great deal of uncertainty (4). A “not applicable” response is also permitted. Part B asks participants to rank the degree of stress related to their uncertainty for the 54 items on a 3-point scale from no stress to very high stress. Part C consists of four 10-cm-long visual analog scales that measure global uncertainty, global stress, global threat, and positive feelings. The latter visual analog asked respondents if they perceived any positive feelings because of their uncertain state and if so, the degree of positive feelings they have. The USS was developed and has been validated on several samples. For more information on the development and psychometric properties of the USS, the interested reader is referred to Hilton.38

The French version of the USS was developed using iterative back-translation in a study that examined uncertainty in long-term survivors of cancer.6 Translating the conceptual rather than literal meaning was the goal. Two bilingual persons from the community worked together to translate the items and instructions into French: one of French-Canadian background and the other of English-Canadian and European background. When the first draft of the French version was completed, it was given to a bilingual nurse with a French-Canadian background to back-translate into English. She had not seen the English version. The results revealed that most items were clear, but some items needed further revision. For example, item 9 read “I am uncertain about the effectiveness of my treatments.” This was back-translated and read “I am uncertain if my treatments were successful.” Item 19 read “I am uncertain how to manage my symptoms,” but the back-translation read, “I am uncertain how to manipulate my symptoms.” A couple of items were difficult to translate, such as, “I am uncertain whether my cancer situation will be involved in my death.” Items that were difficult were reviewed; most difficulties related to inference and wording.

Skrutkowski identified the items that had not translated or back-translated well and another bilingual person of English-Canadian and European background, who had taught mathematics in a French school for 20 years, was asked to translate those items into French. This person revised the translations, and those items were then back-translated by the same nurse who had done the initial back-translation work. The back-translations improved considerably. For example, the first back-translation of “about the effectiveness of my treatments” was “si mes traitements ont été effectives,” (si mes traitements ont été effectives), clearly much closer to the original intent. Another item, “whether my medical condition will cause me to have symptoms,” was back-translated initially as “si mon état medical causera des symptômes.” After a clearer translation and back-translation was done, it read, “if my medical state will cause symptoms” (si mon état medical causera des symptômes). Another item read, “how to manage my medical care.” The initial translation meant, “how to manipulate my medical care” and was revised...
to mean, “how to manage my medical care (comment gérer mes soins médicaux). The translated scale was then pilot-tested with a French-speaking person who found it to be workable and the content appropriate.

In discussions with the translators, most challenges related to translating words such as coping, managing, and threat. There is no real translation in French for the word coping or to cope. In fact, French-speaking people often use the word coping in their French language to give meaning to the word or else choose among a variety of expressions to describe it, such as “faire face à,” “vivre,” and “composer avec.” Finding a word in French with similar meaning was a challenge because to manage something implies something a little different from coping. The word manage can have more than one verb, “gercer” means to manage in a business or commerce sense, manner means to handle, and “diriger” means to manage a project or plan. Because the word threat was problematic, it was first translated as “angoisse” and finally as “menace.” In the province of Quebec, Canada, there are many dialects in which words can be said differently or imply slightly different meanings. The French language is also spoken in different ways depending on the region of the province or country the person is from. For example, “Parisian” French is distinctly different from “Québécois” French. “Québécois” adopts different expressions that might not be used or known in France.

Skrutkowski administered the English scale to 59 English-speaking monolinguals and the French scale to 13 French-speaking monolingual individuals who had participated in the Nucare study or were diagnosed with cancer between 8 and 12 years earlier, were not in a terminal state of the disease, and were able to speak English or French. The Nucare project was a program aimed at fostering active coping to improve a sense of personal control by using individual strengths. Results presented here focus on the secondary analysis pertaining to the equivalence testing of the scale.

Total uncertainty, total stress, and visual analog scale responses for uncertainty, stress, threat, and positive feelings about having uncertainty, as well as individual uncertainty item scores, were examined for the total group, and English-speaking respondent scores were compared to French-speaking respondent scores (Table 1). If the English and French versions were equivalent, there should have been no differences in each of those components across the groups. For the total group (English- and French-speaking respondents combined), total uncertainty was low as was the stress and threat from uncertainty. Most people (41 of the 72) found no positive connotation to their uncertainty. In comparing French-speaking and English-speaking responses, there were no significant differences in total uncertainty, total stress, or any of the visual analog scales that measured uncertainty, stress, threat, and the level of positive feelings related to the uncertainty. In addition, similar items reflected the highest uncertainty for both groups. Highest uncertainty related to whether a check-up might find something wrong, the cause of their condition, whether changes in their condition would be detected early, about the stability of their cancer, whether the cancer would be the same in 5 years, and whether the cancer would return. The English- and French-speaking groups differed on only 2 items: uncertainty related to choosing the treatments they should have and whether they were being told the truth about their cancer. The similarity in response further supported the content equivalence of the translated scale.

Testing for conceptual equivalence involved comparing responses of those with recurrence to those with no recurrence. If the French version was equivalent to the English version, a difference in uncertainty would be expected between those with and those without recurrence. Similarly, those differences should hold across language groups. This was generally the case. All responses were in the direction expected. Total uncertainty and the visual analog responses for uncertainty, stress, and threat differed significantly between the recurrence and the nonreurrence groups (Table 2). Because of the small number of French-speaking respondents, the nonparametric procedure of Mann-Whitney U was used and the level of significance was set at P = .10 to allow reflection of trends. The French- and English-speaking respondents demonstrated similar trends, although not as closely as one might expect, likely influenced by the small sample (Table 3).

Examination of items also reflected similar differences between the recurrence and nonrecurrence groups. The items related to their chances to be well and whether their cancer was under control differed significantly between those with and without recurrence for both language groups. Other items that discriminated quite well included whether their cancer situation would be the same in 5 years. The patterns were therefore similar for the English- and French-speaking respondents, which provided additional evidence for conceptual equivalence.

### Table 1 • Distribution of Uncertainty Responses Between English- and French-speaking People

<table>
<thead>
<tr>
<th>Item</th>
<th>All Respondents (n = 72)</th>
<th>English (n = 59)</th>
<th>French (n = 13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total uncertainty*</td>
<td>13.9 (18.0)</td>
<td>14.6 (19.2)</td>
<td>11.0 (10.8)</td>
</tr>
<tr>
<td>Total stress†</td>
<td>3.8 (8.8)</td>
<td>4.2 (9.5)</td>
<td>2.2 (4.2)</td>
</tr>
<tr>
<td>Visual analogue scales‡</td>
<td>15.3 (23.2)</td>
<td>15.4 (23.7)</td>
<td>14.9 (21.6)</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>10.0 (18.6)</td>
<td>9.6 (20.0)</td>
<td>11.7 (20.3)</td>
</tr>
<tr>
<td>Threat</td>
<td>9.7 (18.6)</td>
<td>8.9 (18.4)</td>
<td>13.5 (19.7)</td>
</tr>
<tr>
<td>Positives</td>
<td>28.0 (36.1)</td>
<td>29.6 (37.0)</td>
<td>20.7 (32.2)</td>
</tr>
</tbody>
</table>

*Possible range 0–216.
†Possible range 0–108.
‡Possible range of all visual analogue scales is 0–100 mm.
Instruments and procedures need to be developed that are relevant and meaningful for people from many cultures and languages; however, for those instruments to be useful, appropriate procedures need to be followed in their development and testing. Let us not assume that by simply translating an instrument from one language to another that it will be valid and reliable and have equivalence to the original version. The process was not without challenge, but the outcome was worth it. The resulting French version has evidence of content, technical, and conceptual equivalence. Although further work is needed to refine the semantic equivalence, the current French version provides an excellent tool for clinicians and researchers.

## Conclusions and Implications

Instruments and procedures need to be developed that are relevant and meaningful for people from many cultures and languages; however, for those instruments to be useful, appropriate procedures need to be followed in their development and testing. Let us not assume that by simply translating an instrument from one language to another that it will be valid and reliable and have equivalence to the original version. To do the process well requires a thoughtful and methodical approach incorporating guidelines developed and suggested by people experienced in cross-cultural work and research. The resultant product is an adaptation of the original version. The term implies more than a simple translation process.

The process of adapting the USS into French using back-translation and the procedures for testing equivalence have been described. The process was not without challenge, but the outcome was worth it. The resulting French version has evidence of content, technical, and conceptual equivalence. Although further work is needed to refine the semantic equivalence, the current French version provides an excellent tool for clinicians and researchers.

### References


### Table 2 • Differences in Uncertainty Responses Based on Recurrence Status Using Mann-Whitney U Nonparametric Test

<table>
<thead>
<tr>
<th>Item</th>
<th>All Respondents (n = 72)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recurrence (n = 21)</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>23.6 (27.9)</td>
</tr>
<tr>
<td>Total stress</td>
<td>8.1 (14.8)</td>
</tr>
<tr>
<td>Visual analogue scales</td>
<td></td>
</tr>
<tr>
<td>Uncertainty</td>
<td>30.8 (30.9)</td>
</tr>
<tr>
<td>Stress</td>
<td>20.7 (30.2)</td>
</tr>
<tr>
<td>Threat</td>
<td>17.9 (25.1)</td>
</tr>
<tr>
<td>Positives</td>
<td>40.5 (36.0)</td>
</tr>
</tbody>
</table>

*P = .05.
†P = .01.
‡P = .001.

Further evidence of equivalence can be shown if there is similarity in scale reliability data. In terms of psychometric testing, the original scale used by the English-speaking respondents had a Cronbach alpha of 0.96. The translated scale had a Cronbach alpha level of 0.88, indicating that internal consistency was quite good. Both alpha levels were similar and therefore provided further support of equivalence.

These tests for equivalence therefore demonstrated that the French version of the USS has overall equivalence to the English version.

### Table 3 • Mean and Standard Deviation of Uncertainty by Language Spoken and Recurrence Status Using Mann-Whitney U Nonparametric Test

<table>
<thead>
<tr>
<th>Item</th>
<th>English Speaking</th>
<th>French Speaking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Recurrence (n = 17)</td>
<td>Recurrence (n = 4)</td>
</tr>
<tr>
<td></td>
<td>Nonrecurrence (n = 42)</td>
<td>Nonrecurrence (n = 9)</td>
</tr>
<tr>
<td>Total uncertainty</td>
<td>25.7 (30.4)</td>
<td>14.8 (11.7)</td>
</tr>
<tr>
<td>Total stress</td>
<td>8.9 (16.2)</td>
<td>4.3 (6.1)</td>
</tr>
<tr>
<td>Visual analogue scales</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncertainty</td>
<td>30.7 (31.7)</td>
<td>31.3 (31.6)</td>
</tr>
<tr>
<td>Stress</td>
<td>20.2 (30.2)</td>
<td>22.5 (34.4)</td>
</tr>
<tr>
<td>Threat</td>
<td>15.1 (24.3)</td>
<td>30.0 (28.6)</td>
</tr>
<tr>
<td>Positives</td>
<td>39.8 (38.0)</td>
<td>43.8 (30.1)</td>
</tr>
</tbody>
</table>

*P = .10.
†P = .01.