Published in final edited form as:

J Adv Nurs. 2011 May ; 67(5): 1012–1021. doi:10.1111/j.1365-2648.2010.05569.x.

Translating Patient Satisfaction into Chinese: A Pilot Study of a Systematic Translation Method for Health Services Research Survey Instruments

Ke Liu, MSN, RN[Associate Professor],

School of Nursing, Sun Yat-sen University Guangzhou, China

Allison Squires, PhD, RN[Assistant Professor], and

New York University College of Nursing aps6@nyu.edu

Li-Ming You, MSN, RN[Professor]

School of Nursing, Sun Yat-sen University Guangzhou, China

Abstract

Aims—This paper is a report of a descriptive comparative pilot study of use of a method that simultaneously tests the content validity and quality of translation of English-to-Chinese translations of two patient satisfaction questionnaires: the La Monica Oberst Patient Satisfaction Scale and Hospital Consumer Assessment of Healthcare Providers and Systems.

Background—Patient satisfaction is an important indicator of the quality of healthcare services. In China, however, few good translations of patient satisfaction instruments sensitive to nursing services exist.

Methods—The descriptive pilot study took place in 2009 and used Content Validity Indexing techniques to evaluate the content, context, and criterion relevance of a survey question. The expert raters were ten nursing faculty and ten patients who evaluated the two patient satisfaction questionnaires. The experts evaluated the relevance each item on a scale of 1 to 4 and the research team compared their responses to choose the most appropriate. Only the nurse faculty experts, who were bilingual, evaluated the quality of the translation using a binary rating.

Results—The “Nurse Rater” relevance scores of the LaMonica-Oberst Patient Satisfaction Scale and the Hospital Consumer Assessment of Healthcare Providers and Systems were .96 and .95 respectively while the patient’s overall relevance scores were .89 and .95. A Mann-Whitney U test demonstrated that results between the two groups were significantly different ($p=.0135$).

Conclusions—Using content validity indexing simultaneously with translation processes was valuable for selecting and evaluating survey instruments in different contexts.

Author Contributions: AS, KL & LY were responsible for the study conception and design

AS, KL & LY performed the data collection

AS, KL & LY performed the data analysis.

AS, KL & LY were responsible for the drafting of the manuscript.

AS, KL & LY made critical revisions to the paper for important intellectual content.

AS, KL & LY provided statistical expertise.

AS, KL & LY obtained funding

AS, KL & LY provided administrative, technical or material support.

AS, KL & LY supervised the study

Conflict of interest: No conflict of interest has been declared by the authors.

Keywords

nurses; nursing; patient satisfaction; translation; content validity indexing; pilot study; China

INTRODUCTION

In China and other places around the world, policymakers, healthcare providers, and health system administrators are asking a simple question to consumers: Are you satisfied with the care you received while hospitalized? Despite the increased attention it now receives on the global scene, patient satisfaction is not a new issue. Only recently has patient satisfaction gained acceptance as an important outcome of health services provision. Often, researchers find patient satisfaction associated with patient outcomes in hospitalized patients (Aiken et al. 1997, McDonald et al. 1998, Doran 2003), but they also find it varies widely by geographic region as recent studies by the CareChex group discovered (2010).

As more countries want to explore and compare patient satisfaction in different contexts, the appropriate selection and translation of patient satisfaction survey instruments becomes an important methodological challenge for researchers undertaking these studies. The purpose of this study is to describe how we pilot tested the use of a Content Validity Indexing (CVI) based approach to choose between two translated patient satisfaction instruments developed in the United States (US), the LaMonica-Oberst Patient Satisfaction Scale (LOPSS) and the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS), for use in the Chinese context.

BACKGROUND

Patient Satisfaction Measurement in Western Countries

Patient satisfaction instruments vary in their content, reliability, validity, and the aspects of patient satisfaction that they measure. Many hospitals in the US use the Press-Ganey patient satisfaction evaluation service to gauge unit level outcomes related to patient satisfaction (Platonova et al 2006, Clark et al., 2007, Donahue et al 2008). In nursing, a few researchers in the US developed patient satisfaction with nursing care instruments in the late 1980s (Risser 1975, La Monica et al 1986, Eriksen 1987, Nash et al. 1994). Other patient satisfaction instruments appear in the literature, but few were used in large-scale studies (Hinshaw & Atwood 1982, McColl et al. 1996).

Among these instruments, nurse researchers often utilized the LOPSS and validated it by factor analysis (La Monica, et al. 1986, Aiken et al. 1997, Aiken et al. 1999, Vahey et al. 2004). The LOPSS includes 40 items and 3 subscales, with Chronbach Alphas at .91, .92, and .89 for the three subscales and .92 (test 1), .95 (test 2) for the total instrument. The construct validity and content validity also were tested (La Monica et al. 1986). Another researcher translated LOPSS into Spanish and used a similar validation process (Lange 2002).

Recently, however, researchers and policymakers in many countries have developed multiple new instruments to measure patient satisfaction with hospital care, but these instruments are not widely available for comparison. In the last five years, the US is no exception as researchers developed the HCAHPS to measure Medicare patient satisfaction with hospital care. In addition to English, the HCAHPS is available in Spanish and Chinese for use in US hospitals with Medicare patients.

Both the LOPSS and HCAHPS state that they generally measure patient satisfaction. Upon closer examination, however, the survey questions in the LOPSS center specifically on aspects of nursing services delivery, including interpersonal interactions between patients and nurses and timing of services as the way to measure satisfaction with nursing care. In contrast, the HCAHPS broadly addresses patient satisfaction with hospital care and has many questions that are sensitive to the quality of nursing services, such as the quality of discharge teaching and assistance with toileting and activities of daily living.

Measuring Patient Satisfaction in China

China began evaluating patient satisfaction with health services in the 1990s as part of a new hospital credentialing process. During that process, the healthcare system surveyed more than 30 million patients. The Ministry of Health now uses patient satisfaction as a major indicator for credentialing government-run institutions (Ren 2003).

In the first reported patient satisfaction study in the Chinese research literature, Xu and Mou (1993) conducted a survey about mainland Chinese patient satisfaction with a self-developed instrument and tested it in one hospital in Shandong province in 1993. Chinese health services researchers developed and tested with several patient populations a set of patient satisfaction instruments. Those instruments underwent a rigorous validation process with multiple types of patients that used test-retest and internal reliability for the reliability, content validity, and construct validity (Chen et al. 1999, Cai & Chen 2002, Zhang et al. 2006), but did not specifically evaluate that quality of nursing services or examine its relationship to patient satisfaction. Yet closer examination of the Chinese literature reveals that most of the patient satisfaction surveys developed by Chinese researchers failed to employ sufficiently rigorous methods for testing the instrument's reliability and validity, were institutionally or regionally specific, and were not usable for comparative international studies (Zhang 2001, Liu et al 2007).

Translation of instruments from other languages into Chinese is also inconsistent. For example, in 2007 in China, one national survey of patient satisfaction was conducted, using a modified questionnaire from Taiwan. That survey involved 4,260 patients in 71 hospitals cross the Mainland of China (Guo et al. 2007). The adapted survey from Taiwan presented some conceptual issues with translation that affected the results because health system administrative hierarchies did not translate well between countries, even though the language was similar.

Instrument Translation: Challenges for Multi-Country Studies

When a researcher needs to translate any instrument, the systematic management of the translation process and evaluation of contextual applicability issues becomes an inevitable methodological challenge (Im et al 2004). The World Health Organization (WHO 2010) advocates for a lengthy multi-step process for instrument translation that can take several years and require more resources in both time and money than most researchers possess. With the urgent need for better instruments for cross-national comparisons of common healthcare quality indicators, researchers need a consistent, reliable, and valid method for validating the quality of translation and the applicability of the items in a new context that can fit within limited resources of researchers.

The literature offers several suggestions for researchers. For a foundation, McDermott and Palchanes (1994) authored another classic article about instrument translation, which discussed the importance of the translation process and its potential influence on quantitative results. The need for a faster, more reliable translation process spurred Perneger, Leplege, and Etter (1999) to develop a rapid translation method with multiple translators that, while

very systematic, did not include expert feedback in their process and increased the risk of obtaining invalid results in the new context. The time factor highlighted by the aforementioned research team is one reason why rigorous translation methods for instruments often receive inadequate attention by researchers, as an integrative review by Maneersriwongul and Dixon discovered (2004). A challenge, then, for researchers who wish to use established instruments and translate them for their own studies is how to do so in a way that produces quantifiable results that can reinforce the reliability and validity of their translation processes.

Flaherty et al.'s (1988) collaborative study outlined possibly one of the most rigorous approaches for cross-cultural validation of instruments. Labeled as cross-cultural validity or equivalence steps, they require the researcher to evaluate content, semantic, technical, criterion, and conceptual equivalence. Table 1 provides Flaherty and colleagues' definitions for the five areas. Each criterion ensures that the researcher accounts for the contextual differences found between different cultures or, as would be the case of patient satisfaction research, different health systems. Of note, researchers have had difficulty over the years finding ways to quantify various aspects of their methodological process (Mallinckrodt & Wang 2004). Others observe that the definition of conceptual equivalence remains ill defined, even to this day (Johnson 2006). Content Validity Indexing (CVI), however, offers a potential way to address the criterion outlined by Flaherty et al (1988), enhance the rigor of the conceptual equivalence process, and provide the quantification aspect that previous studies lacked.

THE STUDY

Aim

The aim of this study was to describe use of a method that simultaneously tests the content validity and quality of translation of English-to-Chinese translations of two patient satisfaction questionnaires: the La Monica Oberst Patient Satisfaction Scale (LOPSS) and Hospital Consumer Assessment of Healthcare Providers and Systems.

Design

This was a comparative, descriptive study that used CVI testing with two sets of expert raters: nurses and patients. CVI is a method for instrument validation that relies on the feedback received from expert raters about the survey's items. An evaluation of the scale's content validity measures the overall quality of the item's construct and its applicability in different contexts (Lynn 1986, Polit et al. 2007). The process evaluates each item individually, known as the Item-CVI (I-CVI) and then calculates and overall average of the I-CVI scores for the instrument, known as the Scale-CVI (S-CVI).

Five raters are the minimum number required to participate in a CVI study and produce statistically valid results (Lynn 1986, Beck & Gable 2001, Polit et al. 2007). Ten respondents are ideal because it decreases variance in responses and the likelihood that the agreements occurred due to chance (Polit et al. 2007). The process occurs before an instrument is pilot tested or used in a full research study with subjects. The feedback obtained from expert raters improves the odds of obtaining "good" data from subjects. Experts recommend an I-CVI score of .78 or above on each item to include it in an instrument (Lynn 1986, Beck & Gable 2001, Polit et al. 2007). In this study, the I-CVI score of .78 would serve as a quantifiable item level measure of Flaherty et al's (1988) five criteria for cross-cultural instrument validation. The S-CVI scores represent the cumulative average of all relevance scores and the summative evaluation of the entire scale. The minimum S-CVI score is .8, with .9 or higher considered ideal.

A common concern with CVI rating process, however, is that only ten experts rate the survey items, posing a risk that all the raters rate every item the same way. Polit, Beck, and Owen (2007) tackled this methodological issue with CVI by examining how chance might affect expert raters' responses on survey items (see the formula in Figure 1). To account for chance agreement between raters, the analyst calculates the number by factoring in the probability of chance agreement and the proportion of chance agreement between the raters, also known as a modified kappa (k) statistic (Polit et al. 2007). By factoring in chance agreement (pc), the strength of the instrument's validity increases and the possibility for errors in data analysis decreases. With translation processes integrated into the CVI evaluation, the method also reduces the likelihood for errors occurring during the translation, thus helping to ensure conceptual, technical, and semantic equivalence.

With the method for instrument translation and validation solidified, we selected the LOPSS and HCAHPS instruments to test the translation validation method because both instruments had adequate validation studies in English, thereby providing the research team with reliable instruments in the original language to work with for the translation validation process. While there is a Chinese language version of the HCAHPS developed for use in the US, the mainland Chinese team members, upon review of the translation, determined that the American Chinese translation would not work well when applied in China due to geographically induced linguistic drifts in the evolution of the language.

Thus, the first step in the instrument translation process was to adapt the items to the Chinese context. The Chinese researchers modified the personal information questions in both scales to conform with Chinese cultural norms. The Chinese research team deleted the question about the discharge plan in HCAHPS because it was not applicable the Chinese healthcare system practices. Once the Chinese team completed the adaptation process in English, the initial translation processes occurred. The Chinese team used a doctorally prepared, bilingual nurse researcher with previous translation experience to translate the original instruments from English to mainland Mandarin Chinese. Upon completion, they emailed the translation back to the US team, who conducted the back translation process with a bilingual, native Chinese doctoral student who also had experience conducting translations. The back translation process identified several conceptual issues with the translation that members of both teams clarified through discussion via the Internet. The study team leaders approved the final translated versions.

Sample

With the final approval on the translations complete, the study team then sought "expert" raters using purposive sampling techniques. Within the team, a discussion occurred over the concept of who was the best "expert" to rate the survey, nurses or patients since the survey aimed to measure *patient* satisfaction with *nursing* care. The team made the methodological choice to include both nurses and patients as raters and conduct a comparative analysis between the rating patterns of each group.

Data Collection

During early 2009, ten bilingual, graduate-level educated nursing professors from nine nursing schools across China served as raters for the simultaneous translation evaluation process. They were chosen according to their educational level, work experience, and English competence. Each nurse rater worked in one of eight government-designated economic development zones. This provided a geographically representative sample for the entire country and did not rely on raters from one geographic area. The CVI rating template was then sent to each rater by email and they were asked to return the survey within 30 days to the study team.

To rate the instrument and the translation using the CVI process, the original item and the translated items were placed side-by-side on a rating form (see Table 2 for an example and the rating scale). The raters scored each item using the CVI process. Nursing faculty raters also completed a “Yes/No” rating about the conceptual equivalence of the translated item. The binary choice forced the nursing faculty experts to make a choice about conceptual, semantic, and technical equivalence of the items instead of allowing a neutral or less clear decision that would prolong the translation process. The CVI template also allowed the expert raters to complete comments about each individual item or the instrument as a whole, thus adding rigor to the process.

Once faculty completed their side-by-side ratings of the translation, the study team recruited ten patients from a single site who evaluated just the Chinese translation. Finding bilingual patients proved difficult, so the research team made the decision to only have the Chinese patients perform a single language CVI. For practical reasons, the research team in Guangzhou selected ten patients hospitalized in a university-affiliated hospital as the patients selected to evaluate the two scales. Purposive sampling of patients based on age, gender, and diagnosis guided the selection process of the patient raters. For the patient’s age, the team targeted patients in different decile ranges (ages 18-39, 40-49, 50-59, 60-69, 70+) in order to account for possible age related differences in patient satisfaction with nursing care. Patients also had to be mentally competent and not impaired by any physical illness that might hinder their ability to complete the CVI process.

All patients completed the anonymous rating process in the same day. A study team member then approached a patient at the time of discharge from the hospital and then described the rating process to the patient and asked them to participate in the project. Only the study team member collecting the data knew the identity of the patient. Patients could refuse to participate. If the patient provided oral informed consent to participate, the study team member provided the patient with the CVI rating template and left the patient alone to complete the rating process. After 30 minutes, the study team member came back to confirm the template had been completed and asked if there were any further suggestions or comments about each item.

Ethical Considerations

Institutional Review Board (IRB) approval for this study occurred at the data collection site through the Ethical Committee of School of Nursing, Sun Yat-sen University, the University of Pennsylvania’s IRB, and hospital ethics committee where patients were approached. Oral consent to participate, a method culturally appropriate in China, was obtained from all participants.

Data Analysis

Once the raters completed the scoring process, the US team first analyzed the two sets of scores using the formulas developed by Polit, Beck, and Owen (2007). In theory, if the instruments are appropriate for the context, there should be no difference between raters’ evaluation scores either at the item level or at the scale. Nonetheless, in case the scores appeared significantly different, the researchers planned to conduct a Mann-Whitney U. STATA 11 served as the data analysis software for the Mann-Whitney U test while Microsoft Excel 2003 was used to complete the CVI calculations.

Polit, Beck, and Owen’s (2007) CVI with chance correction method also recommends that researchers integrate modified kappa scores as additional evaluation criteria. A modified kappa score calculates to .74 or higher for an “excellent” item and helps evaluate the quality of translation. The team also took Polit, Beck, and Owen’s recommendation that S-CVI

scores equal .9 or higher for a truly excellent instrument. In terms of translation, this would mean that the entire instrument translated well into the new context with minimal conceptual or technical errors.

RESULTS

A total of 20 raters individually evaluated the relevance of the 40 LOPSS items and the 24 HCAHPS items. This produced a total of 1,280 relevance scores by the raters for the study. Relevance ratings were normally distributed among both the NR and Patient raters. Bilingual faculty raters (NRs) had between 11 and 29 years of nursing experience, and 60% had doctoral degrees. Patient raters were mostly in their mid-30s or older. Five patients were chosen from medical wards and 5 from surgical wards, including 6 female and 4 male patients. Most had surgical or oncology diagnoses due to the sampling site choice of the researchers. Complete rater demographics are found in Table 3.

Table 4 summarizes the S-CVI scores from each instrument and highlights problem items. As the table illustrates, the bilingual NRs rated both instruments very similarly, with S-CVI scores of .96 for the LOPSS and .95 for the HCAHPS. Patient ratings of the scales, however, were different than the nurses' ratings. Patient S-CVI scores of the LOPSS and HCAHPS were .89 and .95 respectively. To see if the difference was significant, the team opted to use the Mann-Whitney U test. The test revealed significant differences between the patient and nurses ratings of the instruments ($p=.0135$).

Further examination of the results illustrated where these differences occurred. The NRs modified kappa scores identified three problematic items in the LOPSS and three in the HCAHPS. The LOPSS items with modified kappa scores below .74 and I-CVI scores below .78 related to nursing concepts that did not translate well between cultures. The HCAHPS items with low modified kappa and I-CVI scores from the NRs all related to patients rating their personal health or the hospital itself. For the patient raters, however, the LOPSS presented many more problems as 10 out of 40 items (25%) did not achieve the required modified kappa score even though they might have an acceptable I-CVI score. Patients frequently reported in the comments section for the LOPSS that they did not understand why certain questions were asked or what certain concepts meant.

Finally, low modified kappa score ratings on HCAHPS occurred in the same three items for both patients and NRs. In these cases, patient comments echoed those of the NR's, as neither saw demographic questions as relevant. Consequently, from the patient perspective, the findings suggested that the HCAHPS instrument would be a better measure of patient satisfaction with nursing and hospital care in China than the LOPSS.

DISCUSSION

The CVI method calculating for chance proved effective for helping researchers choose the right instrument for measuring patient satisfaction with nursing care in the Chinese context and validate item content. It also proved highly effective identifying problems with translation, tackling semantic equivalence problems, and validating the translated versions of the instruments. Yet despite its promise as an approach to evaluate translation and contextual applicability of an instrument, this approach did have its limitations.

To begin, the first limitation occurred in our choice of translator. We opted not to use a professional translator for the initial translation and subsequent back translation for both convenience and cost reasons. Even though our two translators had translation experience and were both highly educated, their lack of professional translation credentials may have affected the overall quality of the translation. We believe, however, that the CVI process

mediated the effects of not having a professional translator by incorporating the additional feedback of other bilingual individuals. The other limitation to the approach is that it is new and requires further testing to ensure its rigor as a way to evaluate translated instruments. Finally, as we only had patient raters from one city, additional testing in other geographic regions may be required to ensure its applicability in a variety of hospital and community-based settings.

Nonetheless, several findings from the study provide interesting methodological points for discussion. First, rater identity appears to be important when implementing CVI. The patient's and NR's comments and rating scores, together, suggest that language that nurses use to describe their interactions with patients may not always translate well to patients themselves. For instrument developers, this is an important lesson as nursing and health services research expands around the globe. It is easy to use our professional language and assume that patients and colleagues will easily understand what an item means. Health services research instrument developers would be well-served to keep the language they use when writing instrument as simple and clear as possible. Over time, this will facilitate translation processes.

Second, the results show that the strength of using CVI for choosing between instruments is that it provides concrete, objective data for comparing instruments to guide the final selection. With the information provided by the CVI process, the research team was able to select the best instrument for the Chinese context and determine the validity of the content prior to data collection. Even if a researcher opted not to CVI test two instruments to determine which one was the better for that context, the CVI process would provide useful data for evaluating the applicability of the instrument in the context.

These two points lead us to Flaherty et al's (1988) criteria for cross-cultural equivalence of instruments because it previously lacked a systematic way to quantify each category. The study's results appear to show that the CVI translation validation method can address all five criteria in a systematic way. A key to the process is that researchers ask the expert raters to evaluate the relevance of the survey questions to their local context while also simultaneously evaluating the quality of translation. The process permits the expert to evaluate the traditional content and criterion equivalence of normal instrument validation processes while simultaneously collecting feedback from them about the issues most likely to affect the quality of translation: Semantic, technical, and conceptual equivalence in language. The aggregated CVI scores of the raters provide a cumulative average that represents the likelihood that the item and the instrument are equivalent cross-culturally.

Another outcome of this work is that the CVI with translation also enables a monolingual researcher to evaluate the quality of translation without having to know the language. Consequently, incorporating CVI testing into the translation evaluation process for survey instruments also demonstrates promise for providing a sound method for evaluating the quality of translated instruments and their applicability in different contexts. The numeric rating score of the item that CVI testing produces provides the researcher the necessary information to evaluate two important factors in instrument development: The quality of the item's translation and its applicability in the new cultural context. The rater's ability to comment on the items and, when applicable, their translations helped reinforce the rater's numeric score, shed light on the rationale for assigning a specific rating, and identified problems with translation or concepts foreign to the context. These kinds of data are important for creating valid translations of any instrument and improve the likelihood that the instrument will produce reliable results when implemented in the field.

On a final methodological note, researchers should remember that when they have instruments developed in one context and language translated for use in another context and language, CVI testing serves as an important pre-data collection step. While researchers can use CVI testing as a part of the instrument development process, we caution researchers who want to translate instruments to only use the CVI translation validation process on previously validated instruments. Validated instruments have shown their worth for measuring certain concepts in one location already, thus providing a scientific foundation for use of the instrument. Therefore, if a researcher develops a new instrument in one language and translates it into another language before evaluating its reliability and validity, it will decrease the rigor of the instrument as a whole, regardless of the language in which it occurs. That approach with instrument translation would be unlikely to produce reliable or valid results at any stage of instrument development.

CONCLUSION

CVI testing of a translated instrument before a pilot study or larger study appears to increase the odds of obtaining reliable and valid results, along with choosing the best instrument to use. The CVI translation method illustrated in this paper shows strong potential as a structured approach for validating the quality of a translation performed on a previously validated instrument, as well as its cross-cultural validity. It provides researchers a quantifiable way to evaluate content, semantic, technical, criterion, and conceptual equivalence in a translated instrument when applied to a different cultural context. It may also prove useful for re-evaluating the quality of translations found in other instruments that have obtained poor results when compared to their original counterparts. The approach also has the potential to determine if the quality of an instrument's translation, despite its use in other studies, is poor but the monolingual researcher may not be aware of that fact. Further replication of the method with translated instruments will help to confirm the validity of this approach for managing issues related to the quality of instrument translation and contextual applicability.

What is already known about the topic

- No comparative studies of satisfaction with nursing care exist and few reliable and valid instruments for measuring patient satisfaction with nursing care in China are few.
- Researchers know that the process of translating established instruments will affect the quality of results when researchers apply the instrument in new cultures and contexts.

What this paper adds

- This paper illustrates a new method for validating translated versions of health services research instruments through the use of Content Validity Indexing with chance correction.
- Nursing services research that use "nursing language" may not be well understood by patients, as evidenced by the patients' scoring of the more nursing services focused instrument.
- Significant differences were found between the Content Validity Indexing scores from nursing experts and patients and allowed for an evidence-based selection of an appropriate instrument to measure patient satisfaction.

Implications for Practice and/or Policy

- The method illustrated here may provide researchers with a faster, quantifiable, and more rigorous translation and cross-cultural validation process for established survey instruments.
- If researchers evaluate the cross-cultural applicability of several instruments that measure a similar concept prior to data collection, they can choose the best instrument for the context.

Acknowledgments

FUNDING Dr. Squires' contributions were funded by a post-doctoral fellowship at the Center for Health Outcomes and Policy Research, University of Pennsylvania from the National Institute for Nursing Research, NIH (T32-NR-007104), Linda Aiken, PI. Funding for the data collection in China came from the China Medical Board, Liming You, PI; Ministry of Health of China, Liming You, PI; the U.S. National Institute of Nursing Research (P30NR05043), Linda Aiken, PI; and the European Commission, RN4CAST, Walter Sermeus and Linda Aiken, Co-PIs.

Funding Statement: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

- Aiken LH, Sloane DM, Lake ET, Sochalski J, Weber AL. Organization and outcomes of inpatient AIDS Care. *Medical Care*. 1999; 37(8):760–772. [PubMed: 10448719]
- Aiken LH, Sloane DM, Lake ET. Satisfaction with inpatient AIDS care: A national comparison of dedicated units and scattered-beds. *Medical Care*. 1997; 35(9):948–962. [PubMed: 9298083]
- Beck CT, Gable RK. Ensuring content validity: An illustration of the process. *Journal of Nursing Measurement*. 2001; 9(2):201–215. [PubMed: 11696942]
- Cai ZY, Chen PY. The analysis of out-patient satisfaction questionnaire in general hospital. *Chinese Hospital Management*. 2002; 8:12–13.
- Carechex. Hospital quality ratings. 2010. Downloaded September 24, 2010 from <http://www.carechex.com/media/maps.aspx>
- Chen PY, Wong CM, Ou YP, Hu J, Zhang C, Liang LP, Cai SY. Pilot study of developing inpatient satisfaction questionnaire as an instrument for measuring quality of medical care in general hospitals. *Chinese Hospital Management*. 1999; 19(2):79–82.
- Cicchetti DV, Sparrow S. Developing criteria for establishing interrater reliability of specific items: Application to assessment of adaptive behavior. *American Journal of Mental Deficiency*. 1981; 86:127–137. [PubMed: 7315877]
- Clark PA, Leddy K, Drain M, Kaldenberg D. State nursing shortages and patient satisfaction: more RNs-better patient experiences. *Journal of Nursing Care Quality*. 2007; 22(2):119–29. [PubMed: 17353747]
- Donahue MO, Piazza IM, Griffin MQ, Dykes PC, Fitzpatrick JJ. The relationship between nurses' perceptions of empowerment and patient satisfaction. *Applied Nursing Research*. 2008; 21(1):2–7. [PubMed: 18226757]
- Doran, DM. *Nursing-sensitive outcomes: state of the science*. Jones & Barlett Publishers; Boston: 2003.
- Eriksen LR. Patient satisfaction: an indicator of nursing care quality? *Nursing Management*. 1987; 18(7):31–35. [PubMed: 3648596]
- Flaherty JA, Gavira FM, Pathak D, Mitchell T, Wintrob R, Richman JA, Birz S. Developing instruments for cross-cultural psychiatric research. *Journal of Nervous and Mental Disease*. 1988; 178(5):257–263. [PubMed: 3367140]
- Fleiss, J. *Statistical methods for rates and proportions*. 2nd ed. John Wiley; New York: 1981.
- Guo YH, Jiao J, Zheng XJ, Liu HP. A survey of inpatient satisfaction with nursing care in 24 districts of China. *Chinese Journal of Nursing*. 2008; 43(4):293–295.

- Hinshaw A, Atwood J. A patient satisfaction instrument: Precision by replication. *Nursing Research*. 1982; 31:170–175. [PubMed: 6918923]
- Im E, Page R, Lin L, Tsai H, Cheng C. Rigor in cross cultural nursing research. *International Journal of Nursing Studies*. 2004; 41:891–899. [PubMed: 15476762]
- Johnson TP. Methods and frameworks for crosscultural measurement. *Medical Care*. 2006; 44(11, S3):S17–S20. [PubMed: 17060823]
- La Monica EL, Oberst MT, Madea AR, Wolf RM. Development of patient satisfaction scale. *Research in Nursing and Health*. 1986; 9:43–50. [PubMed: 3634418]
- Lange JW. Testing equivalence of Spanish and English versions: The LaMonica–Oberst (Revised) patient satisfaction with nursing care scale. *Research in Nursing and Health*. 2002; 25:438–451. [PubMed: 12424781]
- Liu GY, Qin JQ, Zhou CQ, Zhen WH, Zhen YL. Study on patient satisfaction evaluation system. *Chinese Hospital Management*. 2007; 23(3):170–173.
- Lynn MR. Determination and quantification of content validity. *Nursing Research*. 1986; 35(6):382–385. [PubMed: 3640358]
- Mallinckrodt B, Wang C. Quantitative methods for verifying semantic equivalence of translated research instruments: A Chinese version of experiences in close relationships scale. *Journal of Counseling Psychology*. 2004; 51(3):368–379.
- McColl E, Thomas L, Bond S. A study to determine patient satisfaction with nursing care. *Nursing Standard*. 1996; 10:34–38. [PubMed: 8949152]
- McDermott MAN, Palchanes K. A literature review of the critical elements in translation theory. *Image: Journal of Nursing Scholarship*. 1994; 26(2):113–117.
- McDonald R, Free D, Ross F, Mitchell P. Client preferences for HIV inpatient care delivery. *AIDS Care*. 1998; 10(Suppl 2):123–35.
- Nash MG, Blackwood D, Boone EB, Klar R, Lewis E, MacInnis K, McKay J, Okress J, Richer S, Tannas C. Managing expectations between patient and nurse. *Journal of Nursing Administration*. 1994; 24(11):49–55. [PubMed: 7965182]
- Perneger TV, Leplege A, Etter JF. Cross-cultural adaptation of a psychometric instrument: two methods compared. *Journal of Clinical Epidemiology*. 1999; 52(11):1037–1046. [PubMed: 10526997]
- Platonova EA, Hernandez SR, Shewchuk RM, Leddy KM. Study of the relationship between organizational culture and organizational outcomes using hierarchical linear modeling methodology. *Quality Management in Health Care*. 2006; 15(3):200–9. [PubMed: 16849991]
- Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*. 2007; 30:459–467. [PubMed: 17654487]
- Ren ZN. Research on customer satisfaction degrees in modern hospital. *Chinese Journal of Health Administration*. 2003; 19(6):370–372.
- Risser N. Development of an instrument to measure patient satisfaction with nurses and nursing care in primary care settings. *Nursing Research*. 1975; 24:45–52. [PubMed: 1038021]
- Vahey DC, Aiken LH, Sloane DM, Clarke SP, Vargas D. Nurse burnout and patient satisfaction. *Medical Care*. 2004; 42(2):II-57–II-66. suppl. [PubMed: 14734943]
- World Health Organization [WHO]. Process of translation and adaptation of instruments. 2010. Downloaded February 17, 2010 from http://www.who.int/substance_abuse/research_tools/translation/en/
- Xu XL, Mou LF. The importance of medical care evaluation: The patient satisfaction survey of 106 inpatients. *Chinese Hospital Management*. 1993; 13(11):19–20.
- Zhang C, Yang JM, Chen PY. Effective factors on satisfaction of emergency patients in general hospitals. *Chinese Hospital Statistics*. 2006; 13(2):97–99.
- Zhang QH. Establishing comprehensive evaluation system of satisfactory degree in the whole hospital. *Chinese Hospital Statistics*. 2001; 8(2):93–95.

$$\begin{aligned}
 & \text{Number of experts giving a rating of 1 or 0 to the item} \\
 & \text{total number of experts} \\
 & \frac{1}{p} \left(\frac{N}{1 + (N-1)p} \right)^{\frac{1}{p}} \quad \frac{1}{1-p} \left(\frac{N}{1 + (N-1)p} \right)^{\frac{1}{1-p}} \\
 & p = \text{the probability of chance agreement} \\
 & N = \text{number of experts} \\
 & x = \text{number of experts giving the item a rating of 1}
 \end{aligned}$$

Figure 1.
Content Validity Indexing Calculations with Chance Correction (Polit, Beck, & Owen, 2007)

Table 1

Flaherty et al's (1988) Definitions of Cross-Cultural Validity in Instrument Translation

Criteria	Definition
Content Equivalence	The content of each item of the instrument is relevant to the phenomena of each culture being studied.
Semantic Equivalence	The meaning of each item is the same in each culture after translation into the language and idiom (written or oral) of each culture.
Technical Equivalence	The method of assessment is comparable in each culture with respect to the data that it yields.
Criterion Equivalence	The interpretation of the measurement of the variable remains the same when compared with the norm for each culture studied.
Conceptual Equivalence	The instrument is measuring the same theoretical construct in each culture.

Adapted from Flaherty et al (1988), p. 258.

Table 2

Sample of instrument translation & content validity index rating template

Item	Relevance Rating (1 to 4)	Translation	Conceptually Equivalent? Yes or No	Comments About the Item
1. The nurses are not as attentive as they should be.		护士对我不够 关心		
2. The nurses do not seem to do anything with the information I give them.		护士对我提出 的问题/要求没 有采取任何措 施		

Rating Scale:

1 = Not relevant

2 = Somewhat relevant

3 = Very relevant

4 = Highly relevant

Table 3

CVI Rater Demographics

Faculty experts	Role	Education Level	Experience (years)	Level of English	City
1.	Educator	Masters	29	Sociolinguistic competence	Beijing
2.	Educator	Doctorate	11	Sociolinguistic competence	Beijing
3.	Educator	Masters	28	Sociolinguistic competence	ShenYang
4.	Educator	Doctorate	25	Sociolinguistic competence	Xi'an
5.	Educator	Doctorate	21	Sociolinguistic competence	ShangHai
6.	Educator	Doctorate	26	Sociolinguistic competence	ChengDu
7.	Educator	Doctorate	25	Sociolinguistic competence	ChangSha
8.	Educator	Masters	19	Sociolinguistic competence	WuLuMuQi
9.	Educator	Doctorate	15	Sociolinguistic competence	GuangZhou
10	Educator	Masters	15	Sociolinguistic competence	GuangZhou

Patients	Diagnosis	Age	Sex
1.	Peritonitis	74	Female
2.	Left breast tumor	55	Female
3.	Thyroid tumor	46	Female
4.	Cecum tumor	58	Female
5.	Appendicitis	39	Female
6.	Myeloma	68	Female
7.	Leukemia	42	Male
8.	r/o irritable bowel syndrome	83	Male
9.	Stomachache	61	male
10.	Post-op cystostomy	35	male

Table 4

S-CVI and problem item results from expert raters of LOPSS and HCAPS

Instrument	NR S-CVI	PT S-CVI
LOPSS	.9625	.8925
HCAPS	.9541	.95

Problem Item #	I-CVI		Modified Kappa Score *		Item Text in English
LOPSS	NR	PT	NR	PT	
L-1	1.0	.7	.99	.46	The nurses are not as attentive as they should be.
L-7	.8	.6	.71	.18	The nurses appear to enjoy caring for me.
L-11	.7	.8	.46	.88	I feel more like a "case" than an individual with the nurses.
L-14	1.0	.8	.99	.71	The nurses do not answer my call for help promptly enough.
L-15	1.0	.7	.99	.46	The nurses tell me they will return to do something for me and then do not keep their promise.
L-23	.9	.6	.88	.18	The nurses act like I cannot understand the medical explanation of my illness when, in fact, I really can.
L-24	1.0	.6	.88	.18	The nurses fail to consider my opinions and preferences regarding my care.
L-29	1.0	.8	.99	.71	The nurses tell me things which conflict with what my doctor tells me.
L-30	1.0	.8	.99	.71	The nurses seem disorganized and flustered.
L-31	1.0	.8	.99	.71	The nurses seem reluctant to give assistance when I need it.
L-34	.8	1.0	.71	.99	The nurses show me how to follow my treatment program.
HCAPS					
H-22	.8	.9	.71	.88	Would you recommend this hospital to your friends and family?
H-23	.7	.8	.46	.71	In general, how would you rate your overall health?
H-24	.7	.8	.46	.71	What is the highest grade or level of school that you have completed?

L = LOPSS; H = HCAPS; NR = Nurse Rater; PT = Patient Rater

* Revised Kappa Scores (Cicchetti & Sparrow, 1981; Fleiss, 1981): Fair = k of .40 to .59; Good = k of .60–.74; and Excellent = k > .74.