Cross-cultural evaluation of the relevance of the HCAHPS survey in five European countries

ALLISON SQUIRES[1]*, LUK BRUYNEEL[2], LINDA H. AIKEN[3], KOEN VAN DEN HEEDE[2], TOMASZ BRZOSTEK[4], REINHARD BUSSE[5], ANNELI ENSIO[6], MARIA SCHUBERT[7], DIMITRIOS ZIKOS[8] AND WALTER SERMEUS[2]

[1]New York University College of Nursing, 726 Broadway 10th Floor, New York, NY 10003, USA, [2]Katholieke Universiteit Leuven, Belgium, [3]University of Pennsylvania, Philadelphia, PA, USA, [4]Faculty of Health Science, Collegium Medicum, Jagiellonian University, Krakow, Poland, [5]Berlin University of Technology, Berlin, Germany, [6]University of Eastern Finland, Kuopio, Finland, [7]University of Basel, Basel, Switzerland, and [8]National & Kapodistrian University of Athens, Athens, Greece

*Address reprint requests to: Allison Squires, New York University College of Nursing, 726 Broadway 10th Floor, New York, NY 10003, USA. Tel: +212-992-7074; E-mail: aps6@nyu.edu

Accepted for publication 9 June 2012

Abstract

Objective. To describe the systematic language translation and cross-cultural evaluation process that assessed the relevance of the Hospital Consumer Assessment of Healthcare Providers and Systems survey in five European countries prior to national data collection efforts.

Design. An approach involving a systematic translation process, expert review by experienced researchers and a review by 'patient' experts involving the use of content validity indexing techniques with chance correction.

Setting. Five European countries where Dutch, Finnish, French, German, Greek, Italian and Polish are spoken.

Participants. 'Patient' experts who had recently experienced a hospitalization in the participating country.

Main OutcomeMeasure(s). Content validity indexing with chance correction adjustment providing a quantifiable measure that evaluates the conceptual, contextual, content, semantic and technical equivalence of the instrument in relationship to the patient care experience.

Results. All translations except two received 'excellent' ratings and no significant differences existed between scores for languages spoken in more than one country. Patient raters across all countries expressed different concerns about some of the demographic questions and their relevance for evaluating patient satisfaction. Removing demographic questions from the evaluation produced a significant improvement in the scale-level scores ($P = .018$). The cross-cultural evaluation process suggested that translations and content of the patient satisfaction survey were relevant across countries and languages.

Conclusions. The Hospital Consumer Assessment of Healthcare Providers and Systems survey is relevant to some European hospital systems and has the potential to produce internationally comparable patient satisfaction scores.

Keywords: patient satisfaction, measurement, instrument validation, cross-cultural research, health services research, HCAHPS

Across the globe, consumer groups, practitioners and governing agencies (e.g. ministries of health, regulatory boards, etc.) increasingly place patient satisfaction with hospital care as a priority outcome for health system performance [1]. Many researchers have designed instruments to measure patient satisfaction that are specific to a country's health system or individual hospital, with most countries having a standard set of questions on the topic [2–12]. Survey question length and content can vary widely; therefore, comparisons between countries can be challenging [2, 12]. The Picker Institute, for example, conducted some of the first comparative studies of patient satisfaction in Europe and produced some standardized results [13–15]. Interpersonal care processes also influence patients' perceptions of how satisfied they are with their hospital experience, particularly in a country's cultural minorities [4, 6, 10–12]. Other studies have cited factors related to healthcare personnel as key influences in patient satisfaction scores [8, 9, 16–18]. Yet, in order to compare performance across health systems, standardized and comparable measures of patient satisfaction are necessary.

The RN4CAST project is a 12-country (Belgium, England, Finland, Germany, Greece, Ireland, The Netherlands, Norway, Poland, Spain, Sweden and Switzerland) comparative nursing workforce study funded by the Seventh Framework Programme of the European Commission aimed at developing innovative forecasting methods for developing and sustaining the nursing workforce [19]. Researchers from the USA also participated in the study under separate funding mechanisms. One goal of the study was to examine if there was a relationship between patient satisfaction and the nursing workforce. Eight of the 12 countries agreed to collect patient satisfaction data as one of the outcomes sensitive to the performance of the nursing workforce. A previously tested instrument for comparing patient satisfaction in Europe, however, was not available to the study's team.

Therefore, to standardize the measurement of patient satisfaction with the hospitalization experience, the principal investigators proposed that the study use the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey. The survey was originally developed for use in the USA by the Centers for Medicare and Medicaid Services (CMS) in partnership with the Agency for Healthcare Quality Research (AHRQ) [20] and later endorsed by the National Quality Forum (NQF). 'The HCAHPS survey asks discharged patients 27 questions about their recent hospital stay. The survey contains 18 core questions about critical aspects of patients' hospital experiences (communication with nurses and doctors, the responsiveness of hospital staff, the cleanliness and quietness of the hospital environment, pain management, communication about medicines, discharge information, overall rating of hospital, and would they recommend the hospital),' and demographic questions that allow a researcher to adjust for patient mix [21]. The survey's emphasis on communication and interactions between providers and patients also made its potential for cross-cultural sensitivity high. Research by O'Malley et al. [22] found the HCAHPS to be sensitive across differently sized hospitals in the USA. Therefore, for the RN4CAST study, it offered the potential to produce comparable results across health systems among the participating countries and with US data.

The purpose of this study is to describe the systematic translation and cross-cultural evaluation process used by the RN4CAST project in five European countries to determine, prior to data collection, the cross-cultural relevance and applicability of the HCAHPS in the European context. At present, the validated translated versions of the HCAHPS are available in American English, Spanish (Latin American), Mandarin Chinese, Russian and Vietnamese (http://www. hcahpsonline.org/surveyinstrument.aspx). The available translations reflect the dominant non-English speaking immigrant populations in the USA. Eight countries out of 12 involved in the study opted to include patient satisfaction data in their study. The participating countries included Belgium, Finland, Germany, Greece, Ireland, Poland, Spain and Switzerland. No translations of the HCAHPS were available in several of the languages; thus, the RN4CAST team had to translate the instrument into seven additional languages (Dutch, Finnish, French, German, Greek, Italian and

Polish) and cross-culturally evaluate the instrument prior to data collection. In the end, five countries participated in the pre-data collection, cross-cultural evaluation process reported in this study.

Translating an instrument for use in a multi-country, comparative study requires not only translating the instrument from the source language to the target one, but also performing a cross-cultural evaluation of the instrument's applicability to the new context [23–26]. A rigorous review by Maneersriwongul and Dixon [27] concluded that simple forward and back translation of instruments alone, even when researchers conduct factor analyses post-data collection, are insufficient to produce reliable and valid results from a translated instrument. Flaherty et al. [28] recommend that instruments used across cultures and that require translation undergo an evaluation that involves content, context, conceptual, semantic and technical equivalence to ensure that the instrument is appropriate for use in the new location (see Fig. 1 for definitions). Failure to integrate this kind of evaluation can produce significant issues related to contextual and conceptual equivalence [29], especially when administrative language (i.e. managerial roles, terms of reimbursement, etc.) is involved in the translation process.

Prior to embarking on the evaluation study in Europe, Liu et al. [30] undertook a pilot study of the translation method proposed for use in Europe. The pilot study took place in China and attempted to use the US-translated Mandarin Chinese version of the HCAHPS [30]. Initial review of the Chinese translation used in the USA by immigrants found that the translations had some subtle linguistic differences that were deemed sufficient to affect results and resulted in another translation into mainland Mandarin Chinese. The pilot study helped inform the final approach to language translation used in the RN4CAST study for both the nurse and patient surveys [31].

Methods

To translate the HCAHPS survey, with the translation framework developed by Squires et al. [31] serving as a methodological guide for the cross-cultural adaptation process, each country's teams used the following steps. It began with a review of the instrument by 'research experts' comprised of representatives from each participating country's team. For the first 22 items in the survey, the team determined if there were any US health system-specific terms that might pose a problem for translation. The only translation issue that emerged was that some answer descriptors, like the difference between 'fair' and 'poor', proved difficult to conceptually differentiate for translation purposes for most non-native English speakers. For Likert-type responses found in the HCAHPS survey (e.g. never, sometimes, often, always), standardized translations were used to ensure equivalence across languages and cultures.

The demographic questions in the HCAHPS items 23 through 27 posed some problems related to contextual and conceptual relevance. While important for risk adjustment

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

Figure 1 Cross-cultural validity in instrument translation: definitions. Adapted from Flaherty *et al.* [28], p. 258.

purposes, the questions about race and ethnic identity were specific to the USA and not applicable to all countries involved in the study. Issues about educational equivalence also arose since the exact equivalence of primary, secondary and post-secondary education across European countries is not well established. European team members also indicated that these types of questions are not commonly asked in survey research in the region. Thus, the result of the initial review was that the teams opted to keep risk adjustment questions (How would you rate your health overall?) and adapt the educational equivalence criteria found in an additional item. This strategy allowed the main questions of the instrument [1–22] to remain intact and the demographic ones to reflect each country's needs. The final instrument contained 24 questions in total, with the original 22 questions maintained and 2 questions focusing on demographics.

Once the final version was established, the systematic translation process used by the team for the HCAHPS translations involved the use of experienced translators (separate for forward and back translations as is standard practice) and a review of the resulting translations by the country teams, who were all bilingual. These combined steps address all five aspects of Flaherty's criteria. Then, an evaluation by 'expert' raters of the relevance of the survey's questions to the hospital care experience in the country also took place. Expert reviewers have excellent consistency with predicting the relevance of survey questions to the population of interest, as a recent investigation by Olson [32] demonstrated. In the case of this study, recently hospitalized patients were defined as the 'experts' since they are the ones who experience the results of the delivery of health services by healthcare professionals and system operations. Thus, each country's team aimed to recruit 7–12 patients who had experienced a hospitalization within the last year to serve as an expert rater. The patient experts also had to be able to follow instructions for completing the evaluation of the survey questions and have enough years of education to complete the task. A patient

rater's ability to speak English was not required for this aspect of the cross-cultural evaluation process because of the difficulty in gauging the English fluency of the patient experts. The use of patient raters addresses Flaherty's evaluation criteria around content and contextual equivalence.

Once selected by the country's research team, each patient rater received oral and written instructions in their own language about evaluating the relevance of the survey questions to their hospitalization experience. The raters used content validity indexing (CVI) techniques and had the opportunity to make comments on each item and about the survey as a whole. Using the CVI approach, raters scored each survey question for their relevance to the patient care experience using the following scale: 1, = Not relevant, 2 = Somewhat relevant, 3 = Very relevant and 4 = Highly relevant. CVI techniques produce an item level score (an average of all raters evaluations of a question, known as an I-CVI) and then a scale-level score (S-CVI) which is the average of all item level scores for a question [33]. A common concern with the CVI approach is the possibility of chance agreement among raters occurring [33]. To address that concern, once the patient's scores completed, the research team used a formula that adjusts the CVI calculation to account for chance agreement between the raters [33]. The resulting *modified kappa* score can then be used to evaluate the cross-cultural relevance of a survey question. The score is a reflection of chance-agreement corrected proportions of patient agreement that raters scored and item as 'relevant' and 'highly relevant'.

Results

With seven languages involved in the rating process, a total of 70 patient raters were invited to evaluate the HCAHPS translations. Sixty-eight patient raters (97% participation rate) participated in the process, with only the Swiss-Italian group having 8 raters; all other language groups had 10. As Table 1 illustrates,

Table 1 Scale level results of the patient rater evaluations

Analysis results by language ($n = 68$)			With demographic question scores removed	
Language	Scale CVI with chance correction (k)[a]	Overall rating	Scale CVI with chance correction (k)	Overall rating
Dutch (Belgium)	0.89	Excellent	0.95	Excellent
Finnish	0.83	Excellent	0.86	Excellent
French (Belgium)	0.91	Excellent	0.91	Excellent
French (Switzerland)	0.75	Excellent	0.82	Excellent
German (Switzerland)	0.63	Good	0.68	Good
Greek	1.00	Excellent	0.99	Excellent
Italian (Switzerland)	0.65	Good	0.63	Good
Polish	0.74	Excellent	0.86	Excellent

[a]CVI with chance correction formula by Polit *et al.* [33] that produces a modified kappa score using the rating scale described by Cicchetti and Sparrow [34] and Fleiss [35]: 0.59 or less = poor; 0.60 to 0.73 = good; $\geq$ 0.74 = excellent.

the scale-level *modified kappa* scores ranged from 0.63 to 1.00. Swiss German and Italian translations received the lowest scores while Greek was the highest. Per the *modified kappa* scoring standards recommended by Cicchetti and Sparrow [34] and Fleiss [35], the team concluded all translations were acceptable for use and had high overall, scale-level relevance scores for their potential applicability to each country's patient care services experience. 'Excellent' ratings by the kappa standard were obtained for all translations except for the Swiss German and Swiss Italian which received 'good' overall ratings. Only demographic questions received 'poor' ratings at the item level.

Because of the scale-level scores for the Swiss German and Swiss Italian translations of the HCAHPS instrument, we explored the effect of removing the item-level scores about the demographic questions from the overall relevance ratings by the patient experts. This resulted in an increase in almost all of the scale-level scores. To determine if the scale-level scores were significantly different if the demographic questions were removed from the process, a *t*-test analysis was conducted. The result confirmed that the removal of the demographic questions produced a statistically significant increase in the relevance scores ($P = .018$, -0.069 to -0.009, 95% CI).

Comments by the patient raters further confirmed the effect of the personal questions on the cross-cultural relevance scores. Many patient raters commented that they did not understand the need to collect 'personal' information and that they did not see how things like education level, in particular, were relevant to patient satisfaction scores. The comments by Swiss German and Italian patient raters also shed some light on the lower overall scores as they slanted toward the negative about the entire survey.

Discussion

The results from this study suggest that patients view the HCAHPS as relevant to their patient care experiences in their home countries. The instrument, as a result, may adapt

well across cultures and developed country health systems for measuring patient satisfaction with in-patient acute care services. The experience of the cross-cultural, pre-data collection evaluation process does raise multiple methodological issues researchers may need to consider when designing multi-country comparative health services research studies.

To begin, the effect of the demographic questions on the overall scores of the instrument has several implications for comparative health services research. First, early evaluation by the research teams hinted at the potential problems that could arise with the demographic questions, in particular when trying to compare educational levels across countries. These issues, however, were mostly technical in nature, like trying to determine what constituted educational equivalence. Patients' scores and feedback, however, highlighted other concerns about demographic questions. On the positive side, the patient experts' comments provided the team with valuable feedback that allowed them to anticipate questions that might go unanswered during the survey process.

Negatively, however, the reaction of patients to sharing personal information on this kind of survey did affect the overall cross-cultural relevance scores of the survey. Even when researchers think it is important, patients or other research participants may not perceive standard demographic questions as relevant to a survey, thereby affecting the scores of an overall instrument. The improvement in the majority of relevance rating scores when the personal questions were removed from the overall scale score illustrates that phenomenon. Another implication regarding the effect of evaluating personal questions contained in established instruments is that researchers need to be more sensitive and judicious about asking for what patients may perceive to be unrelated personal information. Furthermore, the question of whether or not researchers even need to include personal questions in the expert evaluation process that uses CVI techniques remains unanswered and may be specific to the country. It may be worthwhile for researchers to have personal questions evaluated separately when integrated into a survey

instrument that will be applied in other contexts, cultures or countries.

Researchers also need to use their best judgment to determine the potential impact on results of expert rater identity. The patient raters' feedback, through scoring and comments, provided valuable insight to the team that was not identified by 'research' experts alone. This study suggests that combining expertise from researchers and subjects when adapting an instrument for use across cultures and countries may be a stronger approach methodologically to pre-data collection evaluation of a survey instrument. The approach, however, does require further study. Additional follow-up for low scoring questions once data is collected is also necessary to determine if rater evaluation of survey items can accurately predict missing data patterns or unexpected responses.

As with any approach, there are limitations to this study and the methods undertaken. Selection bias by the teams toward specific patient raters may have also occurred and we recommend that researchers develop clear guidelines for selection when choosing raters. Grant and Davis [36] provide some useful references for rater selection. Some raters also had difficulty with the concept of evaluating questions instead of answering them. Finally, since the predictive validity of this process for survey results has not yet finished, researchers should give due consideration of the strength of the approach.

To conclude, the value of this type of pre-data collection, cross-cultural instrument evaluation process is that it lays a solid foundation for comparative country studies of patient satisfaction. The potential threat to validity related to language translation is reduced significantly. Consequently, researchers and policymakers can increase their certainty that their explorations about the impact of health system structures on patient satisfaction across countries have accounted for the basic requirements of rigorous cross-cultural research, addressed role variations among healthcare workers and the financing mechanisms that may differentially affect access to services.

References

1. Institute of Medicine. *Crossing the Quality Chasm*. Washington, DC: Institute of Medicine, 1999.

2. Cohen G, Forbes J, Garraway M. Can different patient satisfaction survey methods yield consistent results? Comparison of three surveys. *BMJ* 1996;**313**:841–4.

3. Draper M, Cohen P, Buchan H. Seeking consumer views: what use are results of hospital patient satisfaction surveys? *Int J Qual Health Care* 2001;**13**:463–8.

4. Cheng S, Yang M, Chiang T. Patient satisfaction with and recommendation of a hospital: effects of interpersonal and technical aspects of care. *Int J Qual Health Care* 2003;**15**:345–55.

5. Perneger TV, Kossovsky MP, Cathieni F *et al.* A randomized trial of four patient satisfaction questionnaires. *Med Care* 2003;**41**:1343–52.

6. Boulding W, Glickman SW, Manary MP *et al.* Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. *Am J Manag Care* 2011;**17**:41–8.

7. Busse R, Valentine N, Lessof S *et al.* Being responsive to citizens' expectations: the role of health services in responsiveness and satisfaction. In: McKee M, Figueras J, Saltman R, editors. *Health Systems: Health, Wealth, Society and Well-being.* Maidenhead: OUP/McGraw-Hill; 2011.

8. Fujimura Y, Tanii H, Saijoh K. Inpatient satisfaction and job satisfaction/stress of medical workers in a hospital with the 7:1 nursing care system (in which 1 nurse cares for 7 patients at a time). *Environ Health Prev Med* 2011;**16**:113–22.

9. Murphy GT, Birch S, O'Brien-Pallas L *et al.* Nursing inputs and outcomes of hospital care: an empirical analysis of Ontario's acute-care hospitals. *Can J Nurs Res* 2011;**43**:126–46.

10. Sack C, Scherag A, Lütkes P *et al.* Is there an association between hospital accreditation and patient satisfaction with hospital care? A survey of 37,000 patients treated by 73 hospitals. *Int J Qual Health Care* 2011;**3**:278–83.

11. Tataw DB, Bazargan-Hejazi S, James FW. Health services utilization, satisfaction, and attachment to a regular source of care among participants in an urban health provider alliance. *J Health Hum Serv Adm* 2011;**34**:109–41.

12. Nápoles AM, Gregorich SE, Santoyo-Olsson J *et al.* Interpersonal processes of care and patient satisfaction: do associations differ by race, ethnicity, and language? *Health Serv Res* 2009;**44**:1326–44.

13. Coulter A, Cleary PD. Patients' experiences with hospital care in five countries. *Health Aff* 2001;**20**:43–53.

14. Jenkinson C, Coulter A, Bruster S. The Picker Patient Experience Questionnaire: development and validation using data from in-patient surveys in five countries. *Int J Qual Health Care* 2002;**14**:353–8.

15. Coulter A, Jenkinson C. European patients' views on the responsiveness of health systems and healthcare providers. *Eur J Public Health* 2005;**15**:355–60.

16. Clark PA, Leddy K, Drain M *et al.* State nursing shortages and patient satisfaction: more RNs-better patient experiences. *J Nurs Care Qual* 2007;**22**:119–27.

17. Jha AK, Orav EJ, Zheng J *et al.* Patients' perception of hospital care in the United States. *N Engl J Med* 2008;**359**:1921–31.

18. Kutney-Lee A, McHugh MD, Sloane DM *et al.* Nursing: a key to patient satisfaction. *Health Aff* 2009;**28**:669–77.

19. Sermeus W, Aiken LH, Heede KV *et al*. Nurse forecasting in Europe (RN4Cast): rationale; design and methodology. *BMC Nurs* 2011;**10**:1–9.

20. Darby C, Hays RD, Kletke P. Development and evaluation of the CAHPS hospital survey. *Health Serv Res* 2005;**40**: 1973–6.

21. Centers for Medicare & Medicaid Services [internet]. HCAHPS Fact Sheet *2010*. Washington, DC: Department of Health and Human Services, United States Federal Government. Accessed July 6, 2011, http://www.hcahpsonline.org/files/HCAHPS% 20Fact%20Sheet%202010.pdf.

22. O'Malley AJ, Zaslavsky AM, Hays RD *et al*. Exploratory factor analyses of the CAHPS hospital pilot survey responses across and within medical, surgical, and obstetric services. *Health Serv Res* 2005;**40**:2078–95.

23. Erkut S. Developing multiple language versions of instruments for intercultural research. *Child Dev Perspect* 2010;**4**:19–24.

24. Johnson TP. Methods and frameworks for crosscultural measurement. *Med Care* 2006;**44**:S17–20.

25. Tran TV. *Developing Cross-Cultural Measurement*. London: Oxford University Press, 2009.

26. Weeks A, Swerissen H, Belfrage J. Issues, challenges, and solutions in translating study instruments. *Eval Rev* 2007;**31**:153–65.

27. Maneersriwongul W, Dixon JK. Instrument translation process: a methods review. *J Adv Nurs* 2004;**48**:175–86.

28. Flaherty JA, Gaviria FM, Pathak D *et al*. Developing instruments for cross-cultural psychiatric research. *J Nerv Ment Dis* 1988;**176**:257–63.

29. Squires A. Methodological challenges in cross-language qualitative research: a research review. *Int J Nurs Stud* 2009;**46**:277–87.

30. Liu K, Squires A, You LM. A pilot study of a systematic method for translating patient satisfaction questionnaires. *J Adv Nurs* 2011;**67**:1012–21.

31. Squires A, Aiken LH, Van den Heede K *et al*. Language translation in health services research for multi-country comparative studies. *Inte J Nurs Stud* 2011 (In press), Accessed from http://www.sciencedirect.com/science/article/pii/S0020748912000600, http://dx.doi.org/10.1016/j.ijnurstu.2012.02.015.

32. Olson K. An examination of questionnaire evaluation by expert reviewers. *Field Methods* 2010;**22**:295–318.

33. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health* 2007;**30**:459–467.

34. Cicchetti DV, Sparrow S. Developing criteria for establishing interrater reliability of specific items: application to assessment of adaptive behavior. *Am J Ment Defic* 1981;**86**: 127–37.

35. Fleiss J. *Statistical Methods for Rates and Proportions*. 2nd edn. New York: John Wiley, 1981.

36. Grant JS, Davis LL. Selection and use of content experts for instrument development. *Res Nurs Health* 1997;**20**: 269–74.