Professional Practice Environment Scale-Chinese Version: Development and Psychometric Testing

Ching-I Teng¹ • Hao-Yuan Chang² • Jui-fen Rachel Lu³** • Chia-Yu Chou⁴

¹PhD, Professor, Graduate Institute of Business and Management, College of Management, Chang Gung University • ²PhD, RN, Assistant Professor, School of Nursing, Colleges of Medicine, National Taiwan University • ³ScD, Professor, Graduate Institute of Business and Management and Department of Health Care Management, College of Management, Chang Gung University • ⁴MS, Administrator, Center for Quality Management, National Taiwan University Hospital.

ABSTRACT
Background: The Professional Practice Environment (PPE) scale is widely used to assess the quality of the healthcare environment around the world. No validated Chinese-language scale that is designed to address this issue currently exists.

Purpose: The present study evaluates the construct validity of the Chinese-version PPE on a sample of 290 healthcare workers in Taiwan.

Methods: Forward and backward translations of the PPE scale of Halcomb et al. were used to ensure semantic equivalence. Further, multiple psychometric properties were examined.

Results: The developed scale showed sufficient equivalence. The results of a survey of 290 healthcare providers demonstrated that the developed Chinese-version PPE scale had high reliability (Cronbach’s alpha = .88) and validity.

Conclusions/Implications for Practice: The findings of the present study support the effectiveness and efficiency of the Chinese-version 30-item Halcomb’s PPE in assessing the professional practice environment in Taiwan.

KEY WORDS: professional practice environment, forward translation, backward translation, Chinese version, scale development.

Introduction

The health insurance system has intensified healthcare market competition under the global budget payment system (Lu & Hsieh, 2003). This competition may motivate hospitals to reduce their staffing costs and shift the emphasis of these institutions toward financial considerations to the detriment of professionalism and to enhance process efficiencies rather than workplace communications. Reduced staffing costs may undermine clinical practices, which warrants the development of a scale that allows the objective evaluation of the professional practice environment. This type of scale may help healthcare managers more thoroughly understand how their organizations may facilitate quality health services. There is currently no Chinese-language scale that is suitable for this purpose. This study was developed to address this need.

The Professional Practice Environment (PPE) scale of Halcomb, Davidson, Caldwell, Salamonson, and Rolley (2010) has shown adequate reliability and validity in assessing the professional practice environment. The PPE has fewer items (30) than other scales developed to address the same issue. The eight dimensions of the PPE scale allow healthcare managers to evaluate various aspects of their clinical practice. The above features show the quality of this scale. However, this scale lacks a Chinese version, which would facilitate the efforts of Chinese-speaking health professionals. Therefore, the research aim of this study is to present a Chinese version of the PPE scale of Halcomb et al. and to psychometrically test the Chinese version for future applications.

Literature Review

Professional practice environment

The professional practice environment refers to “an organizational context in which health care is delivered” (Erickson et al., 2004) and “characteristics of a work environment” (Halcomb et al., 2010). This study adopts the above definition used by Erickson et al. (2004).

Accepted for publication: August 5, 2014
*Address correspondence to: Jui-fen Rachel Lu, No. 259, Wenhua 1st Rd., Kweishan District, Taoyuan City 33330, Taiwan, ROC. Tel: +886 (03) 211-8800 ext. 5483; E-mail: rachel@mail.cgu.edu.tw
The authors declare no conflicts of interest.

Cite this article as:
The literature has identified five distinct dimensions of the practice environment: (a) nurse participation in hospital affairs; (b) nursing foundations for quality of care; (c) nurse manager ability, leadership, and support of nurses; (d) staffing and resource adequacy; and (e) collegial nurse-physician relations (Lake, 2002). Furthermore, the literature has identified the following eight dimensions of professional practice environments: (a) handling of disagreements and conflict, (b) internal work motivation, (c) control over practice, (d) leadership and autonomy in clinical practice, (e) staff relationship with physicians, (f) teamwork, (g) cultural sensitivity, and (h) communication about patients (Erickson et al., 2004; Halcomb et al., 2010). Table 1 lists each of these dimensions and their respective definitions.

Papastavrou et al. (2012) compared professional practice environments in seven countries and found that control over practice differs between countries located at relatively high northern latitudes (e.g., Northern Europe and North America) and those at lower northern latitudes (e.g., Mediterranean countries such as Greece). Papastavrou et al. (2014) further found that Greece scored higher than Cyprus in three factors of professional practice environment (i.e., handling disagreements and conflicts, leadership and autonomy in clinical practice, and teamwork). These findings show the importance of the professional practice environment to global nursing managers.

Yang, Liu, Huang, and Zhu (2013) found that the professional practice environment partially mediates the impact of empowerment on organizational commitment. Suhonen et al. (2014) found that the professional practice environment relates significantly to individualized care. The literature further indicates the relevance of professional practice environment to global health service providers.

A good practice environment elevates healthcare outcomes (Grumbach & Bodenheimer, 2004), further increasing patient satisfaction. Furthermore, a good practice environment elevates the satisfaction of healthcare providers (Manojlovich, 2005) and reduces their turnover intention (El-Jardali et al., 2011). Research is thus warranted to develop valid scales for evaluating the professional practice environment. The next section reviews the literature on this issue.

**Scales for the professional practice environment**

The PPE scale of Erickson (2000) consists of 35 items with eight dimensions. Erickson et al. (2004) modified the PPE scale by adding five and omitting two items. The modified PPE scale thus has 38 items with sufficient reliability (Cronbach’s $\alpha = .93$) and interitem correlation ($r = .81$; Erickson et al., 2004).

Halcomb et al. (2010) further modified the wording of the 38-item PPE scale of Erickson et al. (2004) and evaluated the face validity. Halcomb et al. later surveyed 342 Australian general practice nurses. On the basis of the factor analysis results, Halcomb et al. omitted eight items that cover two questions on internal work motivation, three on control over practice, two on leadership and autonomy in clinical practice, and one on teamwork. The final modified PPE scale of Halcomb et al. consists of 30 items and involves a response option ranging from 1 (lowest) to 4 (highest).

The modified PPE scale of Halcomb et al. (2010) retained adequate reliability and validity, showing its value in being translated into other languages. Halcomb et al. compared the 38-item and 30-item PPE scales in terms of reliability, indicating that differences in the reliability coefficient ranged from −.01 to .03. This finding implies that the 30-item PPE scale of Halcomb et al. is adequately reliable.

**Methods**

**Forward and Backward Translations**

During the forward translation stage, three bilingual experts were invited to translate the English version of the PPE scale of Halcomb et al. (2010) with permission into

<table>
<thead>
<tr>
<th>TABLE 1. Definitions of Each Dimension in the Professional Practice Environment Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dimension</strong></td>
</tr>
<tr>
<td>Handling of disagreements and conflict</td>
</tr>
<tr>
<td>Internal work motivation</td>
</tr>
<tr>
<td>Control over practice</td>
</tr>
<tr>
<td>Leadership and autonomy in clinical practice</td>
</tr>
<tr>
<td>Staff relationship with physicians</td>
</tr>
<tr>
<td>Teamwork</td>
</tr>
<tr>
<td>Cultural sensitivity</td>
</tr>
<tr>
<td>Communication about patients (patient-related communication)</td>
</tr>
</tbody>
</table>
Chinese. Each item was translated into one or two Chinese items from September to October 2011. All Chinese items associated with a single English item were then synthesized into one Chinese item, totaling 30 items in Chinese in the translated scale.

After the forward translation, the backward translation process involved another two bilingual experts who were equally qualified as the forward-translation experts. The two experts were asked to translate each of the 30 Chinese items into up to three English items. Where the translated English items had the same meaning as the corresponding English items in the original scale, these items were considered to have been backward translated successfully. If the translated English items did not have the same meaning, the experts were again consulted to resolve the problem.

Pilot Study to Modify the Wording of the Translated Scale

Nineteen senior nurses were recruited for a pilot test on the translated Chinese scale and asked to provide their comments on its wording.

Research Design, Sample, and Data Collection

This study adopted a cross-sectional research design that used questionnaires to collect data. Nurses at a medical center in northern Taiwan were surveyed. In total, this study recruited participants from 54 ward units. Data were gathered from March 6 to April 3, 2012.

The first round surveyed all of the nurses who worked day shifts during the data collection period. Nurses were required to satisfy the following inclusion criteria to participate: (a) employed in the wards as full-time, frontline, and registered nurses; (b) employed in the hospital during at least the immediately preceding 6-month period; and (c) assigned to dayshift work. The latter criterion reflected the fact that dayshift nurses are required to communicate more frequently with other healthcare professionals to facilitate teamwork, maintain staff relationships with physicians, and handle disagreements and conflict. In total, 299 nurses were solicited, and 290 returned valid completed surveys, yielding a response rate of 97%.

The second survey round was conducted roughly 1 month later, from May 10 to June 5. The questionnaires contained the same items. Nurse participants who participated in the second survey round were sent questionnaires via the post office to reduce interference with their clinical routes, but full postal addresses of 82 participants were not available. Therefore, 208 questionnaires were sent, and 99 valid responses were collected, yielding a response rate of 47.6%.

Ethical Considerations

This study received approval from the institutional review board and the nursing department of the medical center. Before data collection, research assistants were trained on implementing the standardized research procedure and on related ethical issues. The research assistants approached eligible nurses and briefed them on the research objectives of this study. All respondents signed an informed consent form before participation.

All nurses included in this study were assigned a unique code to link their responses to patients’ responses. Research assistants then made appointments with the nurses to gather the questionnaires over a 3-day period. Because the research assistants were not employed by the medical center, it is believed that respondents should be more inclined to respond honestly. The second survey round involved sending questionnaires to those participants who provided their postal addresses.

Instrument Psychometric Properties and Test–Retest Reliability

This study developed a Chinese-language version of the PPE scale of Halcomb et al. (2010) and tested the reliability of this scale using Cronbach’s α values, composite reliability, and average variance extracted (AVE). Moreover, indicator loadings were used to test convergent validity. Finally, the AVE and the squared correlation were compared with test discriminant validity.

Instrument Parallel-Form Reliability

The Professional Work Controls Scale (Jaworski & MacInnis, 1989) was developed using a random sample from the marketers’ list of the American Marketing Association. Participants were asked to rate the items using a response scale ranging from 1 (very disagreeable) to 5 (very agreeable), in which higher scores denote stronger control over their work environments. In developing the Professional Work Controls Scale, 379 participants returned valid responses, and analysis of the responses yielded a Cronbach’s α of .89.

Ramaswami (1996) modified the scale of Jaworski and MacInnis (1989) by replacing “department” with “division” and replacing “marketing professionals” with “members.” When surveying another 318 marketers, Ramaswami found that the modified scale had a Cronbach’s α value of .85. Furthermore, Ramaswami used the confirmatory factor analysis (CFA) method to confirm the factor structure of the modified scale.

The scale of Ramaswami (1996) was adopted as the parallel form of the translated Chinese version, along with wording modifications that replaced “division” with “ward unit,” “members” with “nursing professionals,” and “productivity” with “healthcare services.” These modifications ensured that the wording of the scale more accurately reflected the current research context.

Instrument Criterion-Related Validity

This study adopted job satisfaction as the first criterion for evaluating the professional practice environment. Manojlovich
(2005) indicated that a good practice environment increases the job satisfaction of nurses. Teng, Shyu, and Chang (2007) recently developed an abbreviated (three-item) version of the job satisfaction scale for nurses. This study evaluated job satisfaction among nurses based on this scale.

Turnover intention was adopted as the second criterion for the professional practice environment. A good practice environment has been associated with lower turnover intention (El-Jardali et al., 2011). Because Teng et al. (2007) had evaluated the turnover intention of nurses, this study adopted their scale to evaluate turnover intention.

### Results

Each subsection in this section corresponds to its peer subsection in the Methods section.

#### Forward and Backward Translations

Forward and backward translations were used to develop a semantically equivalent scale for assessment. After forward and backward translations, the Chinese version of the PPE scale was made semantically equivalent to the original in English.

#### Pilot Study to Modify the Wording of the Translated Scale

One pilot study was used to modify the wording of the scale to ensure the clarity of the scale to nurses. On the basis of the comments of the senior nurses, “staff” was changed to “nurses in my unit,” and “leader” was changed to “nursing supervisor.” Moreover, respondents suggested adding examples for “other healthcare providers in the community” to increase the clarity of this item. Hence, “public health center,” “long-term care center,” and “nursing care center” were added as examples of other healthcare providers in the community.

#### Demographic Distribution of the Sample

All participants were women; 95.9% were younger than 40 years old, 97.6% had received an associate or bachelor’s degree or higher, 71.8% had less than 10 years of nursing experience, and 51.8% had a level of competency in the clinical ladder of N2 (acute nursing) or N3 (holistic nursing). The level of competency in the clinical ladder in Taiwan evaluates nursing knowledge skills, as detailed in Teng, Shyu, et al. (2012). Table 2 lists the demographics of the nurse participants.

Gender was excluded from the following analysis because all participants reported their gender as female. On the basis of analysis of variance, this study examined whether and how age levels (three) and educational levels (three) impacted the survey responses. No significant impact was identified apart from the finding that age impacts two items in the dimension of internal work motivation (“Inadequate working relationships with other healthcare providers limit effectiveness” and “Other healthcare providers seem to have a low opinion of this unit”) and that educational level impacts one item in the dimension of teamwork (“Other healthcare providers seem to have a low opinion of this unit”). Analytical results indicate that the sociodemographic variables influence the responses to the 30 items in this scale only slightly.

#### Assessment of Psychometric Properties

Next, the psychometric properties of the developed scale such as reliability and validity were examined. The Kaiser–Meyer–Olkin Measure of Sampling Adequacy equaled .83, exceeding the criterion of > .80. Moreover, the data passed Bartlett’s test of sphericity ($p < .05$). These analytical results indicated that the data were valid for further factor analysis. The psychometric properties of the translated scale were assessed directly using CFA. Table 3 summarizes the CFA results.

The Cronbach’s $\alpha$ value for all of the items used to evaluate each construct exceeded .67, except for the items used to evaluate teamwork. Performance of the Cronbach’s $\alpha$ values

### Table 2.

**Participant Profile ($N = 290$)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$n$</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>288</td>
<td>99.3</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>0.7</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20–29</td>
<td>169</td>
<td>58.3</td>
</tr>
<tr>
<td>30–39</td>
<td>109</td>
<td>37.6</td>
</tr>
<tr>
<td>40–49</td>
<td>9</td>
<td>3.1</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>1.0</td>
</tr>
<tr>
<td>Highest educational level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school</td>
<td>3</td>
<td>1.0</td>
</tr>
<tr>
<td>University/college</td>
<td>279</td>
<td>96.2</td>
</tr>
<tr>
<td>Graduate institute</td>
<td>4</td>
<td>1.4</td>
</tr>
<tr>
<td>Missing</td>
<td>4</td>
<td>1.4</td>
</tr>
<tr>
<td>Nursing work experience (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$&lt;1$</td>
<td>37</td>
<td>12.8</td>
</tr>
<tr>
<td>1–4</td>
<td>97</td>
<td>33.5</td>
</tr>
<tr>
<td>5–9</td>
<td>74</td>
<td>25.5</td>
</tr>
<tr>
<td>10–14</td>
<td>52</td>
<td>17.9</td>
</tr>
<tr>
<td>15–19</td>
<td>23</td>
<td>7.9</td>
</tr>
<tr>
<td>$\geq20$</td>
<td>4</td>
<td>1.4</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
<td>1.0</td>
</tr>
<tr>
<td>Level of competency in the clinical ladder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N/N0</td>
<td>52</td>
<td>17.9</td>
</tr>
<tr>
<td>N1</td>
<td>32</td>
<td>11.0</td>
</tr>
<tr>
<td>N2</td>
<td>79</td>
<td>27.3</td>
</tr>
<tr>
<td>N3</td>
<td>71</td>
<td>24.5</td>
</tr>
<tr>
<td>N4</td>
<td>54</td>
<td>18.6</td>
</tr>
<tr>
<td>Missing</td>
<td>2</td>
<td>0.7</td>
</tr>
</tbody>
</table>
approximately fulfills the reliability criterion (> .70). Furthermore, the Cronbach’s α value for the entire scale equaled .88, which further fulfilled the reliability criterion. Except for the items used to evaluate teamwork, all of the composite reliability values exceeded .79, and all of the AVE values exceeded .53. According to the statistical testing results, the measurements in this study have a tolerable reliability.

As for convergent validity, most items had indicator loadings exceeding .50, with the exception of three items used to evaluate the handling of disagreements and conflict and the (three) items used to evaluate teamwork. Except for these six items, all other items performed acceptably. As for the discriminant validity, squared correlation for any pair of constructs was smaller than the AVE values of each construct in the pair, thus fulfilling the discriminant validity criterion.

The measurement model fits the data tolerably ($\chi^2 = 1851.96, df = 377$, root mean square error of approximation [RMSEA] = .11, normed fit index = .90, comparative fit index = .92, incremental fit index = .92, goodness-of-fit index = .71, and standardized root mean square residual = .085). The normed $\chi^2$ (i.e., $\chi^2/df$) = 4.91, which satisfied the criterion ($\chi^2/df < 5$). The value of normed fit index equaled .90, which fulfilled the criterion, and the value of incremental fit index equaled .92, which satisfied the criterion. However,
the value of RMSEA equaled .11, which did not satisfy the requirements (RMSEA < .10). The fit indices should be evaluated as a whole rather than individually, and not all fit indices must fulfill criteria in the literature (Hair, Anderson, Tatham, & Black, 1998).

The substandard performance of RMSEA may be because of those items used to evaluate teamwork. These items have different targets for evaluation, namely, other healthcare professionals, other units, and other healthcare providers in the community. The three targets may draw on various sources of support and thus degrade the internal consistency of the items used to evaluate this dimension. The degraded internal consistency further deteriorates the validity and model fit performances.

Table 4 lists the correlations among the constructs of this study. The constructs correlated well with each other (r ≥.21, p < .01), with the exception of correlations with teamwork (r = .11 to .08), the common method variance (CMV) that used one construct CMV to explain all of the study items. The CMV model fits poorly with the data (χ² = 506.69, df = 405). The χ² difference was 3654.73, and the degree-of-freedom difference was 28. Therefore, the χ² threshold value was 41.34 (i.e., χ²(α = .05, df = 28) = 41.34). Comparing the χ² difference with its threshold value revealed that the χ² difference (3654.73) exceeded its threshold value (41.34), indicating that the model without the CMV significantly outperforms the model with the CMV. Thus, the lack of the CMV in this study indicates the quality of the instrument.

**Assessment of Test–Retest Reliability**

The test–retest reliability involves two waves of data collection. Because it is common that not all participants in the first wave respond to the second wave, the possibility of nonresponse bias should be examined. The no-response bias may result from the difference between respondents and nonrespondents. Therefore, this study also examined whether respondents and nonrespondents differed in specific socio-demographic variables and work-related variables, namely, gender, age, educational level, nursing work experience, and nursing certification level. Mann–Whitney tests were used. As noted, all participants were women, and thus, there was no gender difference between the respondents and nonrespondents. Furthermore, the two groups were similar in age (p = .62), educational level (p = .23), length of nursing work experience (p = .93), and nursing certification level (p = .97).

Finally, test results indicate that respondents and nonrespondents were similar in terms of the examined sociodemographic and work-related variables. The above confirms the randomness of the nonresponse bias in this study.

The average scores for items used to evaluate each dimension were calculated for each participant. The average scores in the first wave were then correlated with the average scores in the second wave. The resultant correlation coefficients were used as the indicators for evaluating the test–retest reliability.

The test–retest reliability for each dimension ranged from .20 to .56. In particular, the test–retest reliability was .30 (p < .01) for handling of disagreements and conflict, .52 (p < .01) for internal work motivation, .51 (p < .01) for control over practice, .56 (p < .01) for leadership and autonomy in clinical practice, .46 (p < .01) for staff relationship with physicians, .42 (p < .01) for teamwork, .40 (p < .01) for cultural sensitivity, and .20 (p < .05) for communication about patients.

Overall, the test–retest reliability for the translated Chinese version was adequate.

**Assessment of Parallel-Form Reliability**

Parallel-form reliability may be used to examine whether a scale generates scores that are consistent with a different scale used to measure a similar construct. In this assessment, responses to the five items of the modified version of the scale of Ramaswami (1996) were averaged for each individual nurse. The average scores were correlated with the average scores for all 30 items in the Chinese version of the PPE scale. Restated, scores on the Professional Work Controls

**TABLE 4.**

**Correlations Among Study Constructs**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Handling disagreement and conflict</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Internal work motivation</td>
<td>.40**</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Control over practice</td>
<td>.22**</td>
<td>.30**</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Leadership and autonomy in clinical practice</td>
<td>.23**</td>
<td>.28**</td>
<td>.42**</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Staff relationship with physicians</td>
<td>.35**</td>
<td>.37**</td>
<td>.29**</td>
<td>.46**</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Teamwork</td>
<td>.07</td>
<td>–.11</td>
<td>–.09</td>
<td>–.05</td>
<td>–.10</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>7. Cultural sensitivity</td>
<td>.44**</td>
<td>.49**</td>
<td>.25**</td>
<td>.31**</td>
<td>.51**</td>
<td>.08</td>
<td>–</td>
</tr>
<tr>
<td>8. Communication about patients</td>
<td>.37**</td>
<td>.41**</td>
<td>.24**</td>
<td>.21**</td>
<td>.43**</td>
<td>–.09</td>
<td>.65**</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.
Scale were correlated with scores on the Chinese version of the PPE scale. Responses from 99 nurses were used to generate the correlation that was then used as the indicator of parallel-form reliability. The parallel-form reliability was .72 ($p < .01$), indicating adequate scale reliability.

**Assessment of Criterion-Related Validity**

The criterion-related validity was assessed by correlating the scores of our scale with scores on two criteria: job satisfaction and turnover intention. The correlation between the professional practice environment and job satisfaction was .53 ($p < .01$), indicating sufficient criterion-related validity for the evaluations. In addition, this study correlated the eight dimensions with the score of job satisfaction. Analytical results indicated that the correlations ranged from .30 to .51 ($p < .01$), except for the correlation regarding teamwork ($r = .08$, $p = .44$). The above results further show that most of the dimensions of the practice environment positively contribute to the job satisfaction of healthcare professionals.

The correlation between the average responses to all 30 items used to evaluate a practice environment and to evaluate turnover intention was $- .51$ ($p < .01$). This finding suggests that a practice environment is a valid predictor of turnover intention. In addition, the eight dimensions were lowly-to-moderately related to turnover intention ($r$ ranged from $- .23$ to $- .52$, $p < .05$), with the exception of the correlation related to teamwork ($r = - .03$, $p = .78$). Analytical results showed that most of the dimensions of the practice environment contributed to a low level of turnover intention.

To provide further evidence on validity, five experts were asked to evaluate the validity of each item using a 4-point scale that ranged from 1 (very inappropriate for measuring this construct) to 4 (very appropriate for measuring this construct). Twenty-five of the 30 items received a rating of 3 or 4 by all experts. The overall content validity index was 94.7% (see Lynn, 1986). These results further support the content validity of the items.

**Discussion**

**Main Findings and Contributions**

This study found that the Chinese-language version of the PPE scale developed in this study (our version, hereafter) has adequate semantic equivalence to the scale of Halcomb et al. (2010). Our version has adequate reliability, as evaluated using three criteria. Our version also has acceptable validity in terms of convergent validity, discriminant validity, and content validity (as rated by experts). Our version further performs tolerably in multiple fit indices. Moreover, the eight dimensions in our version are distinct from each other, as supported by minimal CMV. In addition, the risk of nonresponse biases affecting the findings of this study is low. These findings as a whole indicate that our version is a reliable and valid tool for measuring the professional practice environment.

This study exhibited that our version has adequate performance in test–retest reliability, parallel-form reliability, and criterion-related validity. These findings provide further evidence in terms of reliability and validity of our version and its usefulness to nursing managers.

This study translated the 30-item PPE scale of Halcomb et al. (2010) into a Chinese-language version and thus contributes to current literature by providing a reliable and valid instrument for evaluating professional practice environments in bilingual/multilingual settings. Therefore, our version increases the scope of application of the original PPE scale and enables researchers to evaluate practice environments in Chinese-speaking populations and subsequently investigate the impact of the practice environment on healthcare providers and patient outcomes.

This study mirrors Erickson et al. (2004) in using eight dimensions to represent the practice environment construct and extends their work to show that all of these dimensions are applicable in a non-English-language setting.

Related studies (Grumbach & Bodenheimer, 2004; Manojlovich, 2005) have shown the merits of a good practice environment in patient and personnel outcomes. These merits motivated the authors of this study to provide a reliable and valid instrument to evaluate practice environment quality. This instrument contributes to the above literature by enabling the assessment of practice environments and facilitating the evaluation of efforts to improve these environments.

**Implications for Healthcare Managers and Nursing Supervisors**

The significant impact of cultural differences on cooperative behavior is a key issue for healthcare managers (Teng, Lee, Chu, Chang, & Liu, 2012). One dimension of the PPE scale measures cultural sensitivity, which may be used by managers to assess and improve their own cultural sensitivity to improve their managerial effectiveness (Thomas & Ravlin, 1995). As language and culture often go hand in hand (Kramsch, 1998), it is vital that managers recognize the importance of the intersection of language with culture in the professional setting. For healthcare professionals to fully and effectively understand the questions being posed about their practice environment and to most accurately convey their opinions to healthcare managers, they preferably need to be able to read and answer those questions in their mother tongue. This argument is evidenced by the general preference for and ability to use a mother tongue instead of a second language (Gulbrandsen, Schroeder, Milerad, & Nylenna, 2002). Thus, the translated version of the PPE scale provides a vital tool for extracting information that might otherwise have been lost when respondents answer in a language other than their mother tongue and perhaps do not fully grasp the nuances of the language.

The Chinese version of the PPE scale is applicable in various Chinese-speaking settings. Because Chinese-speaking people...
comprise a significant proportion of the world’s population (Chen, Chou, & Hwang, 2006), the translated version has potentially widespread applicability. This Chinese version may thus be used by healthcare managers worldwide to evaluate and improve their services in Chinese-language settings.

This study recommends that managers use the PPE-Chinese version to evaluate the practice environment, elicit the baseline values for the eight dimensions at their institute or unit, and then implement improvement activities according to the scores of eight dimensions. Nursing supervisors and healthcare managers may compare preintervention scores with postintervention scores to assess the effectiveness of improvement activities. This comparison provides an effective means of evaluating the effectiveness of the improvement activities. In addition, this scale may be used as a tool to help managers identify those aspects in need of further improvement. The entire process indicates the direction for continuous improvement within the practice environment, which boosts healthcare outcomes, increases health service provider satisfaction, and increases health service provider retention in the organization (El-Jardali et al., 2011; Grumbach & Bodenheimer, 2004; Manojlovich, 2005).

Research Limitations and Future Research Directions
Because of time and resource constraints, this study did not use the Delphi method. However, to overcome this limitation, we asked related experts to rate the validity of each item. Future studies may consider using the Delphi method to increase the robustness of scale measures.

A further limitation of this study is that the reliability and validity of the three items used to evaluate teamwork were unsatisfactory. This is an area that should be improved in future studies on this scale. Future studies with different objectives may devise new items or revise the existing three items to evaluate the teamwork dimensions to further improve the psychometric properties of the translated scale.

Conclusions
This study developed a Chinese-language version of the PPE scale using a forward–backward translation process to verify semantic equivalence with the original (English-version) scale. Analysis results indicate that the Chinese version offers sufficient reliability and validity in terms of internal consistency, convergent validity, and discriminant validity. In addition, the Chinese version showed good parallel-form reliability and good criterion-related validity using job satisfaction and turnover intention as two criteria. In summary, the Chinese version is a reliable and valid instrument for evaluating professional practice environments in Chinese-speaking populations. Furthermore, this instrument is applicable to elicit the quality of a professional practice environment and to measure the effectiveness of improvement activities. This study recommends that future studies and clinical healthcare managers apply this instrument to various clinical environments to further improve health service environments.

Acknowledgments
The authors thank Taiwan National Science Council (now Ministry of Science and Technology) for the funding support of this study (NSC-100-2632-H-182-001-MY2).

References


專業執業環境量表中文版的發展與心理計量之測試

陳景宜1  張皓媛2  盧瑞芬3*  周家玉4

1長庚大學企業管理研究所教授  2國立臺灣大學護理學系助理教授  3長庚大學企業管理
研究所暨醫務管理學系教授  4臺灣大學醫學院附設醫院品質管理中心管理師

背景
專業執業環境量表（PPE）已被廣泛用於測量醫療執業環境，具有良好的信效度。然而，
目前尚未有一個中文版的專業執業環境量表。

目的
本研究的目的為發展 Halcomb 版本之 PPE 中文版，其內容之信效度測試。

方法
本研究以 Halcomb 之 30 題 PPE 量表為本，採用前譯與回譯的方式，將原著 PPE 進行轉
譯，並檢測量表同等性及各項心理計量特性。

結果
此量表與原量表間具有足夠的量表同等性，對 290 位護理人員的調查結果顯示此量表具
有高度的信度（Cronbach’s Alpha= .88）與效度。

結論／
實務應用
此份 PPE 中文版可以有效地用來評估台灣的專業執業環境。

關鍵詞：專業執業環境、前譯、回譯、中文版本、量表發展。

接受刊載：103年8月5日
*通訊作者地址：盧瑞芬  33330桃園市龜山區文化一路259號
電話：（03）2118800－5483  E-mail: rachel@mail.cgu.edu.tw