Measurement Invariance and Latent Mean Differences in the Nurses' Emotional Labour Scale

Sun-Hee KIM1• Eunkyung LEE2• Jin-Hwa PARK3•

ABSTRACT

Background: The measurement invariance and latent mean differences in emotional labor across different hospital and monthly salary levels among registered nurses have never been confirmed for the Emotional Labour Scale. These issues may influence the application and efficacy of this scale in practice.

Purpose: This study was developed to evaluate the factor structure of the nurses’ Emotional Labour Scale and to examine the measurement invariance and latent mean differences for this scale across different hospital and monthly salary levels.

Methods: Data were collected from 461 registered nurses working in four general hospitals and 12 long-term care hospitals. Confirmatory factor analysis and a multigroup confirmatory factor analysis were performed to determine the internal structure and measurement invariance of the Emotional Labour Scale.

Results: The results of the confirmatory factor analysis indicate that the factor structure model proposed by the original scale fits well with our data as well as configurual invariance, factor loading invariance, intercept invariance, and uniqueness invariance. Moreover, factor variance/covariance invariance across two hospital levels as well as configural invariance, factor loading invariance, and intercept invariance across two monthly salary levels were supported. The mean score for emotional control effort in the profession of general hospital nurses was lower than that for long-term care hospital nurses. No statistically significant latent mean differences were found across monthly salary levels.

Conclusions/Implications for Practice: The findings show the Emotional Labour Scale to be a valid and reliable tool for assessing registered nurses and also comparing the mean score for emotional labor across hospital and monthly salary levels to be feasible. The scale may contribute to the development of human resource strategies.

KEY WORDS:
confirmatory factor analysis, emotion, hospital, measurement invariance, nurse.

Introduction

Emotional labor has been studied primarily among service providers such as flight attendants (Hochschild, 2012) and began to be studied in sociology and business approximately 20 years ago (Grandey & Sayre, 2019). There has been increasing interest in emotional labor among Korean nurses from the perspective of nursing service (S. H. Kim & Ham, 2015; J. E. Kim et al., 2019). Because emotional labor is a concept rooted in the premise of interactions with others and refers to the efforts to control and express emotions to adhere to the norms required by clients or organizations (M. J. Kim, 2017), the practice of emotional labor may vary according to the characteristics of clients and organizations. In contrast to other service jobs that serve a physically and mentally healthy target population, nursing services are provided to patients with physical and mental health problems. In addition, because nurses provide primary care at the patient bedside, their emotional labor differs from that of other professionals (Gray, 2009; Williams, 2013).

Emotional labor exists for nurses in all nursing care settings, as it relates to their interpersonal interactions with patients, guardians, and other health professionals. In particular, nurses show therapeutic emotional labor in their relationships with patients and guardians (Delgado et al., 2017). Nurses perceive increased autonomy as professionals when providing care and services for patients with a professional attitude (Hong, 2016). Moreover, the facial expressions, attitudes, and kindness that nurses show during the process of care are directly linked to patients’ satisfaction (Andrew et al., 2011) and facilitate an empathetic experience with patients (Msiska et al., 2014). The emotional management efforts of nurses to use positive facial expressions and control their feelings increase patient comfort and sense of security. Thus, this concept of emotional labor has been considered an important topic in the study of nursing work (Mann & Cowburn, 2005).

As described above, understanding the unique features of nurses’ work is critical to understanding nurses' emotional...
labor. However, most Korean studies on emotional labor in nurses have utilized a translated and revised version (S. H. Kim & Ham, 2015) of a foreign instrument originally developed for use on hotel workers (Morris & Feldman, 1996). Moreover, this instrument comprises factors such as the frequencies and varieties of emotional display, attentiveness for emotional display, and emotional dissonance (Morris & Feldman, 1996) and is limited in its ability to assess the unique attributes of nurses' emotional labor (Hong & Kim, 2019). In addition, validity and reliability-testing problems were encountered in developing the Korean version of this instrument. To overcome these issues, Hong (2016) developed the Emotional Labour Scale for nurses to measure the properties of emotional labor performed by Korean registered nurses with patients.

Hong (2016) and Hong and Kim (2019) identified the attributes of emotional labor of hospital nurses using a sample of 304 participants. The construct validity of the instrument was tested in these studies using exploratory factor analysis, the group comparison method, and item–total correlation, whereas criterion validity was tested based on correlation with Brotheridge and Lee's (2003) emotional scale. The Emotional Labour Scale was developed by Hong and Hong and Kim using sample populations of nurses in a metropolitan area, and applicability is limited to general hospitals and upper-level general hospitals. The scale has not yet been tested in other geographic settings or types of medical institutions. Therefore, to increase the generalizability of this scale, measurement invariance should be tested by ensuring that those who recorded the same score on a specific scale remain at the same level of the measurement area regardless of individual characteristics. Therefore, the adequacy evaluation of the Emotional Labour Scale for long-term care hospitals and wages should be investigated, and the consistency of its factor structure should be tested.

Many studies employ confirmatory factor analysis during the process of instrument development or assessment to examine and validate instrument suitability. In addition, confirming that the factor structure of an instrument is maintained equally between groups is crucial to ensuring that scores may be compared between groups without error (S.-H. Kim, 2019). Emotional labor and other factors related to emotional labor (e.g., burnout) are known to vary according to hospital and nurse salary levels (K. R. Lee & Kim, 2016; Yom et al., 2017). In a previous study (K. R. Lee & Kim, 2016), nurses’ degree of emotional labor was statistically significantly higher in large-sized hospitals than in small- and medium-sized hospitals but was not affected by level of income. However, in Yom et al., the surface acting of emotional labor was found to differ between yearly income groups. Moreover, the mean of emotional labor was higher in low-salary nurses, although the difference did not achieve statistical significance (J. Lee, 2019). In another study, higher levels of emotional labor demand were found to be associated with lower salary levels. However, instruments are typically developed without establishing equivalence across groups for major factor-structure constructs. No study on the emotional labor of nurses in the literature has examined measurement invariants across different hospital and salary levels, although, for example, nurses in long-term care settings who earn a certain salary and care for patients with widely different levels of illness severity may interpret the meaning of an item very differently from nurses who work in a general hospital located in a metropolitan area. Therefore, this study was designed to examine measurement invariance in the Emotional Labour Scale, which was not tested by either Hong (2016) or Hong and Kim (2019) during scale development. This will contribute to ensuring the correct use of this scale and the correct interpretation of results and is expected to increase the usefulness of this instrument in measuring emotional labor in nurses. The aims of this study were to examine the factor structure of the Emotional Labour Scale and to assess measurement invariance and latent mean differences across different hospital and nurse salary levels.

**Methods**

**Design and Participants**

In this study, a descriptive cross-sectional survey was used to assess the factorial structure of the Emotional Labour Scale and test for measurement invariance and latent mean differences across hospital and monthly salary levels. Nurses working in a general or long-term care hospital who signed an informed consent form were enrolled as participants.

The minimum sample size for conducting a valid confirmatory factor analysis of the Emotional Labour Scale was calculated based on the model fit (Preacher & Coffman, 2006). The results designated a minimum required sample size of 197 with a degree of freedom of 101, a significance of .05, a power of .90, a null root mean square error of approximation (RMSEA) of .00, and an alternative RMSEA of .05. Thus, the authors designated a target total sample size of 450, with 225 participants assigned to each of the two groups and a withdrawal buffer of 10%. The questionnaire was distributed to 500 participants. After excluding the 26 questionnaires that were not returned and the 13 with incomplete responses, data from 461 participants were available for analysis in this study.

**Instrument**

The Emotional Labour Scale (Hong, 2016) is a 16-item scale consisting of three factors: emotional control effort in profession (seven items), patient-focused emotional suppression (five items), and emotional pretense by norms (four items). Each item is rated using a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5), with higher scores indicating greater emotional labor. At the time of development, the Cronbach’s alpha of the Emotional Labour Scale was .81 for the entire scale, .88 for the emotional control effort in profession factor, .77 for the patient-focused emotional suppression factor, and .69 for the emotional pretense by norms factor. In this study, the Cronbach’s alpha was .84 (95% CI [.82, .86])
for the entire scale, .82 (95% CI [.80, .85]) for the emotional control effort in profession factor, .87 (95% CI [.85, .89]) for the patient-focused emotional suppression factor, and .69 (95% CI [.64, .73]) for the emotional pretense by norms factor. The authors secured permission from Hong before using the instrument.

Data Collection
Data were collected from June 13 to August 18, 2017. Four general hospitals and 12 long-term care hospitals were convenience sampled from Daegu City and Gyeongsangbuk-do Province. Next, the heads of the nursing department at each hospital were contacted via email and phone and informed of the study purpose, participant inclusion requirements, methods, and sample size. Upon obtaining permission to collect data, a research assistant visited the hospital to obtain signed informed consent from the participants and then distributed and retrieved the questionnaires. The participants were each given a small gift as a token of appreciation for their cooperation.

Ethical Considerations
This study was approved by the institutional review board of the authors’ university (approval number: CUIRB-2017-0010) before beginning the study. To protect the participants, they were informed of the purpose, method, and procedures of the study as well as of their freedom to withdraw from the study at any point. In addition, the authors informed participants of the measures that would be taken to ensure participant anonymity and data confidentiality. The participants then signed a consent form and completed the questionnaire.

Statistical Analysis
The data were analyzed using IBM SPSS Statistics 25.0 and Amos 20 software (IBM Inc., Armonk, NY, USA). Demographic characteristics were presented as frequency and percentage. The differences in demographic characteristics between the groups were analyzed using a chi-square test. Univariate normal distribution was tested using skewness and kurtosis. First, the factor structure of emotional labor was analyzed using confirmatory factor analysis. Before this analysis, multivariate normal distribution was tested, and estimates that did not satisfy multivariate normality were analyzed using bootstrapping. There were no missing data on emotional labor in this study. Model fit was examined comprehensively using the χ² statistic (df, p value), normed χ² (normed chi-square, NC), comparative fit index (CFI), goodness-of-fit index (GFI), adjusted GFI (AGFI), RMSEA, and standardized root mean residual (SRMR). The criteria for the model fit indices were p ≥ .05 for χ² statistic, 2.00–5.00 for NC, CFI ≥ .90, GFI ≥ .90, AGFI ≥ .90, RMSEA ≤ .08, and SRMR ≤ .08 (Bae, 2016). To confirm the validity of the factors, convergent validity was established based on a factor loading ≥ .5 (critical ratio, CR ≥ 1.97), construct reliability ≥ .7, and average variance extracted (AVE) ≥ .5, whereas discriminant validity was established based on the AVE being greater than the square of the correlation coefficients between factors and the correlation coefficient plus or minus 2 times the standard error (correlation coefficient ± 2 × standard error) not including the value of 1 (Bae, 2016).

Next, multigroup confirmatory factor analysis was conducted to test for measurement invariance across groups. The levels of invariance were tested using the following invariance steps: configural, factor loading, intercept, uniqueness, and factor variance/covariance (Bae, 2016). Next, the fit of each model was examined. After comparing latent means, if the configural invariance, factor loading invariance, and intercept invariance were established, the latent mean differences across two groups were then examined in a model in which the factor loadings and intercepts were constrained to be equal. In this study, Hedges’s g and Glass’s delta were used as the effect size, with a value of .20 interpreted as small, a value of .50 interpreted as medium, and a value of .80 interpreted as large. Hedges’s g was used because each group differed significantly in terms of hospital-level percentage breakdown. Factor variance invariance was not confirmed in this study. Therefore, Glass’s delta was used, as each monthly salary group had different standard deviations (Ellis, 2020).

Although χ² difference is frequently utilized to test for measurement invariance, this parameter is known to be sensitive to sample size, which hinders accurate analysis of differences. Therefore, differences in model fit indices were utilized as an alternative approach. In this study, as suggested by Chen (2007), the differences in the CFI were used as the main criterion, whereas the differences in the RMSEA and SRMR were used as the subcriterion to identify the measurement invariance model. The comparison criteria differed according to the analytical model used. For studies, including this study, with less than 300 cases per group, the criteria for equivalence of fit between two models are ΔCFI ≤ −.005, ΔRMSEA ≤ −.010, and ΔSRMR ≤ −.025 for the test of factor loading invariance and ΔCFI ≤ −.005, ΔRMSEA ≤ −.010, and ΔSRMR ≤ −.005 for the test of intercept and uniqueness invariances.

Results
Characteristics of Subjects
Most participants were in their 20s (39.8%), followed by 30s (22.6%). Most were women (96.5%), held a bachelor’s degree (52.9%), and were in a clinical career stage with more than 6 years of experience (58.1%). The highest percentage of participants worked in geriatric units (36.9%), followed by medical units (23.6%) and surgical units (17.6%). The most common current job position was staff nurse (81.1%). Most nurses were assigned fewer than 30 patients per shift on average (54.6%), and most nurses had a monthly salary of less than approximately 2,200 U.S. dollars (51.9%). The characteristics of the general and long-term care hospital nurses are shown in Table 1.

With the exception of gender (χ² = 1.82, p = .177) and current position (χ² = 0.91, p = .340), all demographic
characteristics, including age ($\chi^2 = 164.15, p < .001$), educational level ($\chi^2 = 34.90, p < .001$), stage of clinical career ($\chi^2 = 45.07, p < .001$), number of patients assigned per shift ($\chi^2 = 213.05, p < .001$), and monthly salary ($\chi^2 = 93.52, p < .001$), were found to significantly relate to hospital level.

### Confirmatory Factor Analysis for the Emotional Labour Scale

Model fit indices are shown in Table 2. Confirmatory factor analysis was performed to test the factor structure of the Emotional Labour Scale and, with the exception of NC and SRMR, the model fit indices did not meet the criteria: $\chi^2 = 483.216 (df = 101, p < .001)$, NC = 4.784, CFI = .864, GFI = .880, AGFI = .839, RMSEA = .091, and SRMR = .078. After checking the highest modification index, we set a covariance between the errors of Items 1 and 2 and examined the model fit once more. Only NC, RMSEA, SRMR, and GFI met the criteria, whereas the remaining fit indices did not: $\chi^2 = 384.839 (df = 100, p < .001)$, NC = 3.848, CFI = .899, GFI = .904, AGFI = .869, RMSEA = .079, and SRMR = .071. Next, the highest modification index was rechecked, and a covariance between the errors of Items 11 and 12 was set for another test of model fit. With the exception of $\chi^2$, which is influenced by sample size, and AGFI, the model fit indices all satisfied the criteria: $\chi^2 = 325.602 (df = 99, p < .001)$, NC = 3.289, CFI = .920, GFI = .918, AGFI = .887, RMSEA = .071, and SRMR = .070.

### Table 1

**Characteristics of Participants (N = 461)**

| Variable                                | Total          | General Hospital (n = 249) | Long-Term Care Hospital (n = 212) | $\chi^2$ | $p$   |
|-----------------------------------------|----------------|---------------------------|----------------------------------|---------|------|
| Age (years) a                           |                |                           |                                  |         |      |
| 20s                                     | 182 (39.8%)    | 153 (61.7%)               | 29 (13.9%)                       | 164.15  | < .001|
| 30s                                     | 103 (22.6%)    | 63 (25.4%)                | 40 (19.1%)                       |         |      |
| 40s                                     | 82 (17.9%)     | 25 (10.1%)                | 57 (27.3%)                       |         |      |
| 50s or more                             | 90 (19.7%)     | 7 (2.8%)                  | 83 (39.7%)                       |         |      |
| Gender                                  |                |                           |                                  | 1.82    | .177 |
| Male                                    | 16 (3.5%)      | 6 (2.4%)                  | 10 (4.7%)                        |         |      |
| Female                                  | 445 (96.5%)    | 243 (97.6%)               | 202 (95.3%)                      |         |      |
| Educational level a                     |                |                           |                                  | 34.90   | < .001|
| Associate                               | 187 (41.6%)    | 72 (29.4%)                | 115 (56.1%)                      |         |      |
| Bachelor                                | 238 (52.9%)    | 160 (65.3%)               | 78 (38.0%)                       |         |      |
| Master or higher                        | 25 (5.5%)      | 13 (5.3%)                 | 12 (5.9%)                        |         |      |
| Stage of career a                       |                |                           |                                  | 45.07   | < .001|
| Novice (< 3)                            | 112 (24.7%)    | 85 (34.1%)                | 27 (13.2%)                       |         |      |
| Competent (3 to < 6)                    | 78 (17.2%)     | 54 (21.7%)                | 24 (11.7%)                       |         |      |
| Proficient (≥ 6)                        | 264 (58.1%)    | 110 (44.2%)               | 154 (75.1%)                      |         |      |
| Nursing unit                            |                |                           |                                  | 348.62  | < .001|
| Medical                                 | 109 (23.6%)    | 100 (40.2%)               | 9 (4.3%)                         |         |      |
| Surgical                                | 81 (17.6%)     | 81 (32.5%)                | 0 (0.0%)                         |         |      |
| Emergency room                          | 10 (2.2%)      | 8 (3.2%)                  | 2 (0.9%)                         |         |      |
| Intensive care unit                     | 52 (11.3%)     | 27 (10.8%)                | 25 (11.8%)                       |         |      |
| Geriatric unit                          | 170 (36.9%)    | 0 (0.0%)                  | 170 (80.2%)                      |         |      |
| Others                                  | 39 (8.4%)      | 33 (13.3%)                | 6 (2.8%)                         |         |      |
| Current position                        |                |                           |                                  | 0.91    | .340 |
| Staff nurse                             | 374 (81.1%)    | 206 (82.7%)               | 168 (79.2%)                      |         |      |
| Head nurse/unit manager/charge nurse    | 87 (18.9%)     | 43 (17.3%)                | 44 (20.8%)                       |         |      |
| Number of patients assigned per shift a |                |                           |                                  | 213.05  | < .001|
| 29 or below                             | 224 (54.6%)    | 205 (87.2%)               | 39 (18.4%)                       |         |      |
| 30 or more                              | 203 (45.4%)    | 30 (12.8%)                | 173 (81.6%)                      |         |      |
| Monthly salary (USD) a                  |                |                           |                                  | 93.52   | < .001|
| < 2,200                                 | 238 (51.9%)    | 77 (31.0%)                | 161 (76.3%)                      |         |      |
| ≥ 2,200                                 | 221 (48.1%)    | 171 (69.0%)               | 50 (23.7%)                       |         |      |

Note. LTC = long-term care, 
* Missing values were excluded.
With the exception of Item 13 (factor loading = .46), the factor loadings for all of the items in the Emotional Labour Scale satisfied the criteria. Moreover, the criteria for construct reliability ($\geq .78$) and AVE (.58) were also met. Thus, convergent validity was established (Table 3). AVE was greater than the square of the coefficients of correlation between Factors 1 and 2 (.39), between Factors 1 and 3 (.37), and between Factors 2 and 3 (.47). Furthermore, the values for correlation coefficients (between Factors 1 and 2, Factors 1 and 3, and Factors 2 and 3) $\pm 2 \times SE$ did not include 1, at .30–.37, .35–.39, and .43–.51, respectively. Thus, discriminant validity was established.

### Measurement Invariance of the Emotional Labour Scale

**Measurement invariance across hospital level**

To test for measurement invariance across hospital level, we examined the fit of each model (Table 4). The configural invariance

### Table 2

**Model Fit Indices of Confirmatory Factor Models**

| Model                                      | $\chi^2$ | df | NC  | CFI  | GFI  | AGFI | RMSEA | 90% RMSEA | SRMR |
|--------------------------------------------|----------|----|-----|------|------|------|-------|-----------|------|
| Initial model ($n = 461$)                  | 483.216  | 101| 4.784| .864 | .880 | .839 | .091  | .083–.099 | .078 |
| Model 1 ($n = 461$)                        | 384.839  | 100| 3.848| .899 | .904 | .869 | .079  | .070–.087 | .071 |
| Model 2 ($n = 461$)                        | 325.602* | 99 | 3.289| .920 | .918 | .887 | .071  | .062–.079 | .070 |
| Model of general hospital sample ($n = 249$)| 185.408* | 99 | 1.873| .930 | .914 | .882 | .059  | .046–.072 | .074 |
| Model of long-term care sample ($n = 212$)  | 243.837* | 99 | 2.463| .908 | .875 | .829 | .083  | .070–.097 | .082 |
| Model of monthly salaries under 2,200 USD sample ($n = 238$)| 260.484* | 99 | 2.631| .895 | .875 | .829 | .083  | .071–.095 | .085 |
| Model of monthly salaries more than 2,200 USD sample ($n = 221$)| 205.271* | 99 | 2.073| .919 | .899 | .861 | .070  | .056–.083 | .073 |

Note. Model 1 has a covariance between the errors of Items 1 and 2; Model 2 has two covariances between the errors of Items 1 and 2 and between Items 11 and 12. df = degrees of freedom; NC = normed chi-square; CFI = comparative fit index; GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; RMSEA = root mean square error of approximation; 90% RMSEA = the lower limit and upper limit of a 90% confidence interval for the population; SRMR = standardized root mean residual.

* $p < .001$.

### Table 3

**Factor Loading of the Observational Variables in the Emotional Labour Scale ($N = 461$)**

| Factor/Observational Variable | Beta | CR     | $p$  | Construct Reliability | AVE  |
|-------------------------------|------|--------|------|-----------------------|------|
| Emotional control efforts in profession |       |        |      |                       |      |
| Item 1                        | .66  | 16.64  | .004 | .99                   | 0.98 |
| Item 2                        | .64  | 10.33  | .006 |                       |      |
| Item 3                        | .58  | 10.30  | .015 |                       |      |
| Item 4                        | .59  | 11.41  | .009 |                       |      |
| Item 5                        | .68  | 10.45  | .018 |                       |      |
| Item 6                        | .60  | 10.89  | .009 |                       |      |
| Item 7                        | .63  |        | .020 |                       |      |
| Patient-focused emotional suppression |       |        |      |                       |      |
| Item 8                        | .80  | 21.12  | .009 |                       | .89  |
| Item 9                        | .89  | 17.83  | .006 |                       | 0.83 |
| Item 10                       | .78  | 15.12  | .014 |                       |      |
| Item 11                       | .68  | 12.28  | .007 |                       |      |
| Item 12                       | .57  |        | .008 |                       |      |
| Emotional pretense by norms   |       |        |      |                       |      |
| Item 13                       | .46  | 7.03   | .004 |                       | .78  |
| Item 14                       | .50  | 7.70   | .009 |                       | 0.58 |
| Item 15                       | .70  | 7.23   | .007 |                       |      |
| Item 16                       | .72  |        | .013 |                       |      |

Note. CR = critical ratio; AVE = average variance extracted.
was supported (NC = 2.168, CFI = .918, RMSEA = .050, SRMR = .074). In addition, factor loading invariance (ΔCFI = −.003, ΔRMSEA < .001, ΔSRMR = .006), intercept invariance (ΔCFI = −.003, ΔRMSEA = −.001, ΔSRMR < .001), uniqueness invariance (ΔCFI = −.003, ARMSEA = −.001, ΔSRMR < .001), and factor variance/covariance invariance (ΔCFI = −.003, ΔRMSEA < .001, ΔSRMR = .010) were supported.

**Measurement invariance across monthly salary level**

To test for measurement invariance across two monthly salary levels, we examined the fit of each model (Table 4). The configural invariance was supported (NC = 2.352, CFI = .906, RMSEA = .054, SRMR = .085). Factor loading invariance (ΔCFI < .001, ΔRMSEA < .001, ΔSRMR < .003) and intercept invariance (ΔCFI = −.001, ΔRMSEA = −.003, ΔSRMR < .001) were also supported. However, uniqueness invariance (ΔCFI = −.011, ΔRMSEA = .001, ΔSRMR = .001) was not supported.

**Latent Mean Differences in the Emotional Labour Scale**

**Latent mean differences across hospital level**

The only latent mean for emotional control efforts in the profession was statistically significantly lower by .103 in general hospital nurses than in long-term care hospital nurses (CR = −2.430, p = .015). Hedges’s g of Factor 1 (emotional control efforts in profession) was 0.262, indicating a small effect size, whereas those of Factors 2 and 3 were 0.056 and 0.029, respectively (Table 5).

**Latent mean differences across monthly salary level**

No statistically significant latent mean differences were identified for emotional control efforts in profession, patient-focused emotional suppression, or emotional pretense by norms across monthly salary level (CR = −0.367, p = .691; CR = 0.772, p = .381; CR = 1.441, p = .126). Glass’s deltas for Factors 1, 2, and 3 were 0.037, 0.078, and 0.163, respectively, indicating a small effect size (Table 5).

**Discussion**

This study was designed to examine the factor structure of emotional labor developed by Hong (2016) and Hong and Kim (2019) in measuring and evaluating nurses’ emotional labor and to verify whether measurement invariance was maintained across hospital and salary levels. Thus, a survey was conducted on registered nurses working in general and long-term care hospitals, and factor and measurement invariance analyses that discriminated among hospital and salary levels were conducted on the Emotional Labour Scale.
Confirmatory factor analysis was conducted on the 16 items and three factors to identify the factor structure of the Emotional Labour Scale, whereas multigroup confirmatory factor analysis was conducted to examine whether the scale contains measurement invariance across two hospital levels and nurses’ monthly salary levels. Furthermore, latent mean analysis was performed to examine potential variation in the means of the latent factors of emotional labor across two hospital levels and monthly salary levels.

After conducting a confirmatory factor analysis for each factor structure model to examine the fit of the factor structure model of the Emotional Labour Scale, Model 2, which has two covariances between the errors of Items 1 and 2 and between Items 11 and 12 (which had a high modification index), was identified as the most suitable. Item 1 is “I try to be kind to patients genuinely from my heart,” whereas Item 2 is “I try to change my emotions to be positive to meet my patients’ expectations.” Item 11 is “I control my mind, thinking that patience is a virtue,” whereas Item 12 is “I tolerate unfair treatment to maintain a good work atmosphere in the ward.” Because each question has a similar meaning in Korean, it was thought that this may have raised the modification index.

Then, as a procedure before the measurement invariance verification, Model 2 (the model identified in the confirmatory factor analysis as most appropriate) was analyzed separately by hospital and salary levels, with the results assessed to identify match quality with the data in each group. As a result, the three-factor factor structure suggested by Hong (2016) and Hong and Kim (2019) was found to be the most suitable. We successfully replicated the result of the factor structure of the Emotional Labour Scale across hospital and monthly salary levels. This means that nurses in this study interpreted the items on the Emotional Labour Scale similarly regardless of work setting and salary level. In the future, further research should be conducted to explore the possibility of applying the factor structure suggested by the original author transversely (across different countries) using international comparative research.

In the tests for measurement invariance in the Emotional Labour Scale across hospital and monthly salary levels conducted in this study, the scale was shown to have factor loading invariance, intercept invariance, uniqueness invariance, and factor variance/covariance invariance, which is considered the highest level of invariance, across hospital levels. This supports that the scale is valid for application at both of the hospital levels examined, which is a major contribution of this study. In other words, empirical evidence is provided in this study of the usefulness of the Emotional Labour Scale in analyzing the between-group differences in both a general hospital and a long-term care hospital. Furthermore, the Emotional Labour Scale exhibited invariance up to intercept invariance across the two monthly salary levels (i.e., less than and more than 2,200 USD). The equivalence of residual variances with factor variances and covariances was not supported. Therefore, the mean of the latent variable between the two groups may be compared in the future, but caution is needed in the interpretation of the mean difference because of the differences in factor and residual variances.

General hospital nurses earned a lower mean score for F1 (emotional control efforts in profession) than their long-term care hospital peers. This result suggests that the latter display greater emotional control efforts. This implies that long-term care hospital nurses engage in greater emotional labor such as controlling their emotions or breaking down barriers when engaging in interactions with their patients, who typically stay in the hospital longer than general-hospital patients, and thus form stronger carer–patient bonds (Gray, 2009). In this study, the effect size on emotional control efforts in profession, a latent variable, was found to be small. The lack of significant between-group differences in latent mean values suggests the presence of similar variations of means in observed variables that take measurement error into consideration between the groups, with no significant measurement error.

The original scale developed by Hong (2016) and Hong and Kim (2019) targeted nurses working in general hospitals and upper-level general hospitals. However, in this study, nurses from tertiary hospitals (upper-level general hospitals) were excluded, and nurses working in long-term care hospitals and general hospitals were used for comparison. Thus, the Emotional Labour Scale has been used to examine nurses at each hospital level. Moreover, whereas Hong and Hong and Kim studied nurses from healthcare institutions in the Seoul metropolitan area, in this study, nurses were enrolled from healthcare institutions in Daegu City and Gyeongsangbuk-do to test for measurement invariance across regions. In addition, although nurses were distinguished into two groups based on monthly salary level, this allocation does not necessarily correlate with job position. Hence, subsequent studies should be developed to test for measurement invariance across job positions and analyze variance attributable to job position to confirm that the scale may be used regardless of job position. In general, measurement invariance between groups must be verified in scales (even standardized scales) to allow their measured concept to be applied across two or more groups. In the absence of verification, ascertaining whether between-group differences are caused by a true score or a systematic error in estimation of a value is difficult (Han et al., 2019). In addition, further studies to verify longitudinal measurement invariance as well as cross-sectional measurement invariance are suggested. Although no single statistical criterion is applicable to all situations, several important methodological issues based on a previous study (Han et al., 2019), including the sample size and equality across groups, were considered in this study. Because the sample sizes were not equal and the sample size in each group was less than 300, we used a stricter criterion (ΔCFI). The effect size measures were also considered.

**Implications for Nursing Management**

On the basis of the emotional labor scores of nurses, different hospital levels were found to maintain factor variance/covariance...
invariance, and wages (monthly salaries) were found to maintain up-to-intercept invariance. This supports the validity of the Emotional Labour Scale and its reliability for use in making comparisons and conducting between-group analyses distinguished by hospital level and salary level. The verification of measurement invariance supports the use of the Emotional Labour Scale to accurately measure the emotional labor of nurses across diverse settings to improve human resources management.

Emotional labor in nurses is widely considered to be higher than ever now because of factors including the COVID-19 pandemic. The findings of this study have significant implications for nursing management, as the scale may be utilized to provide valid baseline data for managing emotional labor in nurses. Furthermore, by better understanding emotional labor in nurses, nursing organizations can develop more-effective strategies to reduce job stress, increase job satisfaction, and reduce turnover intention. The Chinese version of the Emotional Labour Scale originally developed by Hong and Kim (2019) has also recently been developed and tested for validation (Ying et al., 2020). The data obtained using this scale may be useful in developing human resources strategies and management policies for nurses.

Author Contributions

Study conception and design: SHK
Data collection: EL, JHP
Data analysis and interpretation: SHK
Drafting of the article: All authors
Critical revision of the article: SHK, JHP

Received: November 26, 2020; Accepted: September 24, 2021
*Address correspondence to: Jin-Hwa PARK, PhD, APRN-BC, RN, College of Nursing, Daegu Catholic University, 33, Duryugongwon-ro 17-gil, Nam-gu, Daegu 42472, Republic of Korea. Tel: +82-53-650-4754; E-mail: parkjh07@cu.ac.kr
The authors declare no conflicts of interest.

Cite this article as:
Kim, S.-H., Lee, E., & Park, J.-H. (2022). Measurement invariance and latent mean differences in the nurses’ emotional labour scale. The Journal of Nursing Research, 30(6), Article e239. https://doi.org/10.1097/jnr.00000000000000519

References

Andrew, S., Salamonson, Y., Everett, B., Halcomb, E. J., & Davidson, P. M. (2011). Beyond the ceiling effect: Using a mixed methods approach to measure patient satisfaction. International Journal of Multiple Research Approaches, 5, 52–63. https://doi.org/10.5172/mra.2011.5.1.52

Bae, B. R. (2016). Mplus 7.0 structural equation modeling. Chungram Publishing. (Original work published in Korean)

Bhave, D. P., & Glomb, T. M. (2009). Emotional labour demands, wages and gender: A within-person, between-jobs study. Journal of Occupational & Organizational Psychology, 82(3), 683–707. https://doi.org/10.1348/096317907x360684

Brotheridge, C. M., & Lee, R. T. (2003). Development and validation of the emotional labour scale. Journal of Occupational and Organizational Psychology, 76, 365–379. https://doi.org/10.1348/09631790376847229

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling, 14(3), 464–504. https://doi.org/10.1080/10705511071301834

Delgado, C., Upton, D., Ranse, K., Furness, T., & Foster, K. (2017). Nurses’ resilience and the emotional labour of nursing work: An integrative review of empirical literature. International Journal of Nursing Studies, 70, 71–88. https://doi.org/10.1016/j.ijnurstu.2017.02.008

Ellis, P. D. (2020). How to calculate an effect size. In P. D. Ellis (Ed.), Meta-Analysis Made Easy (pp. 37–42). Kings Press.

Gray, B. (2009). The emotional labour of nursing—Defining and managing emotions in nursing work. Nurse Education Today, 29(2), 168–175. https://doi.org/10.1016/j.nedt.2008.08.003

Han, K., Colarelli, S. M., & Weed, N. C. (2019). Methodological and statistical advances in the consideration of cultural diversity in assessment: A critical review of group classification and measurement invariance testing. Psychological Assessment, 31(12), 1481–1496. https://doi.org/10.1037/pas0000731

Hochschild, A. R. (2012). The managed heart: Commercialization of human feeling (3rd ed.). University of California Press.

Hong, J. (2016). Development and validation of the emotional labor scale for nurses [Doctoral dissertation, Ewha Womans University]. RISS. Http://www.riss.kr/link?id=T13975482 (Original work published in Korean)

Hong, J., & Kim, O. (2019). Methodological and statistical advances in the consideration of cultural diversity in assessment: A critical review of group classification and measurement invariance testing. Psychological Assessment, 31(12), 1481–1496. https://doi.org/10.1037/pas00000731

Kim, J. E., Park, J. H., & Park, S. H. (2019). Anger suppression and rumination sequentially mediates the effect of emotional labor in Korean nurses. International Journal of Environmental Research and Public Health, 16(5), Article 799. https://doi.org/10.3390/ijerph16050799

Kim, M. J. (2017). A study on the re-conceptualization of emotional labor. Korean Management Consulting Review, 17(2), 73–87. https://www.earticle.net/Article/A308833 (Original work published in Korean)

Kim, S.-H. (2019). The breast feeding adaptation scale-short form: Development and testing of its psychometric properties and measurement invariance. Child Health Nursing Research, 25(3), 333–343. https://doi.org/10.4094/chnr.2019.25.3.333

Kim, S. H., & Ham, Y. (2015). A meta-analysis of the variables related to the emotional labor of nurses. Journal of Academy Nursing Administration, 21(3), 263–276. https://doi.org/10.1111/jkana.2015.21.3.263 (Original work published in Korean)

Lee, J. (2019). Emotional labor, burnout and job satisfaction among Korean clinical nurses. Medico-Legal Update, 19(1), 446–451. https://doi.org/10.5958/0974-1283.2019.00084.7

Lee, K. R., & Kim, J. M. (2016). Effects of emotional labor on burnout in nurses: Focusing on the moderating effects of social
intelligence and emotional intelligence. *Journal of Academy Nursing Administration, 22*(1), 22–32. https://doi.org/10.1111/jkana.2016.22.1.22 (Original work published in Korean)

Mann, S., & Cowburn, J. (2005). Emotional labour and stress within mental health nursing. *Journal of Psychiatric and Mental Health Nursing, 12*(2), 154–162. https://doi.org/10.1111/j.1365-2850.2004.00807.x

Morris, J. A., & Feldman, D. C. (1996). The dimensions, antecedents, and consequences of emotional labor. *The Academy of Management Review, 21*(4), 986–1010. https://doi.org/10.2307/259161

Msiska, M., Smith, P., Fawcett, T., & Nyasulu, B. M. (2014). Emotional labour and compassionate care: What’s the relationship? *Nurse Education Today, 34*(9), 1246–1252. https://doi.org/10.1016/j.nedt.2014.03.002

Preacher, K. J., & Coffman, D. L. (2006, May). *Computing power and minimum sample size for RMSEA* [Computer software]. http://www.quantpsy.org/rmsea/rmsea.htm

Williams, A. (2013). Hochschild (2003)—The managed heart: The recognition of emotional labour in public service work. *Nurse Education Today, 33*(1), 5–7. https://doi.org/10.1016/j.nedt.2012.07.006

Ying, Y., Yao, Y., Wei, W., Hu, Y., Zhang, Y., & Chen, M. (2020). Translation and psychometric validation of the Chinese version of the Emotional Labour Scale for nurses. *Journal of Nursing Management, 28*(4), 948–958. https://doi.org/10.1111/jonm.13029

Yom, Y.-H., Son, H., Lee, H. Z., & Kim, M. A. (2017). The relationship between physical discomfort, burnout, depression, social supports and emotional labor of clinical nurses in Korea. *Journal of Korean Clinical Nursing Research, 23*(2), 222–235. https://doi.org/10.22650/JKCNR.2017.23.2.222 (Original work published in Korean)