phrp
OSONG PUBLIC HEALTH AND RESEARCH PERSPECTIVES

pISSN 2210-9099 eISSN 2233-6052
https://doi.org/10.24171/j.phrp.2025.0440
Osong Public Health Res Perspect 2026;17(3):272-281

Original Article

Factor structure and measurement invariance of the EuroQol 5-dimensional questionnaire: a secondary analysis of the Korea Health Panel survey

Byungsun Park[1], **Hanna Choi**[2]

[1]Department of Social Welfare, Kangwon National University, Wonju, Republic of Korea
[2]Department of Nursing Science, Kyungmin University, Uijeongbu, Republic of Korea

ABSTRACT

Objectives: This study evaluated the EuroQol 5-dimensional questionnaire, 3-level version (EQ-5D-3L), using Korea Health Panel (KHP) data by examining its factor structure, measurement invariance across gender and age groups, and longitudinal measurement invariance.
Methods: Panel 1 data from the second survey year (2009), when the EQ-5D-3L was first introduced in the KHP, through the 12th year (2017) were analyzed, along with panel 2 data from 2019 to 2021. Confirmatory factor analysis and measurement invariance tests by gender and age groups were conducted within each period. Longitudinal measurement invariance was also evaluated for each period.
Results: A 1-factor model demonstrated good fit for the EQ-5D-3L. In panel 1, full measurement invariance across gender and age groups was supported. In panel 2, partial invariance was achieved after relaxing constraints on item 5. Longitudinal measurement invariance was supported over 5- and 10-year intervals in panel 1 and over a 3-year interval in panel 2, indicating temporal stability of the measurement model.
Conclusion: The EQ-5D-3L used in the KHP panel 1 and panel 2 datasets demonstrates a stable 1-factor structure and acceptable measurement invariance across key subgroups and over time. These findings support the use of the EQ-5D-3L as an appropriate instrument for assessing health-related quality of life among Korean adults and for longitudinal analyses within large-scale panel surveys.

Keywords: Confirmatory factor analysis; EuroQol five-dimensional questionnaire; Health-related quality of life; Korea Health Panel; Measurement invariance

Received: October 13, 2025
Revised: February 4, 2026
Accepted: March 9, 2026

Corresponding author:
Hanna Choi
Department of Nursing Science, Kyungmin University, 545 Seobu-ro, Uijeongbu 11618, Republic of Korea
E-mail: hanna.choi.kr@gmail.com

Introduction

Background/Rationale

Health-related quality of life (HRQOL) is widely recognized as a multidimensional construct

The opinions expressed by authors contributing to this journal do not necessarily reflect the opinions of the Korea Disease Control and Prevention Agency and Korea National Institute of Health.

© 2026 Korea Disease Control and Prevention Agency.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

reflecting an individual's perceived health status. The EuroQol 5-dimensional questionnaire (EQ-5D) is one of the most widely used instruments for measuring HRQOL because of its simplicity and international comparability [1]. The scale has been translated into more than 130 languages and is applied across a wide range of research and policy contexts owing to its cross-cultural validity, ease of administration, and strong psychometric performance [2,3]. The EQ-5D assesses 5 dimensions of health: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [4]. Respondents indicate their health status using 3 response levels (1 = no problems, 2 = some problems, 3 = extreme problems), producing a 5-item health profile [2].

The 3-level version of the EQ-5D (EQ-5D-3L) is easy to understand and requires minimal time to complete, making it widely applicable in both clinical and population-based studies. However, because of its well-documented ceiling effect, the instrument may have limited sensitivity for detecting subtle differences in health status or changes over time, particularly in general population samples [5–7]. To address this limitation, the EuroQol Group developed the 5-level version of the EQ-5D (EQ-5D-5L), which retains the same 5 dimensions but expands the response scale to 5 severity levels [8]. Consequently, a growing body of research has examined the psychometric properties and practical applications of both the EQ-5D-3L and EQ-5D-5L instruments [2,4,9,10].

In the Republic of Korea, the EQ-5D-3L is widely used to measure HRQOL. In particular, it is included in the Korea Health Panel (KHP), a nationally representative panel survey that collects detailed information on health status, health care utilization, and medical expenditures [11]. The KHP has provided data from panel 1 covering the period from 2008 to 2018, and panel 2 has been conducted since 2019. Although the EQ-5D-3L is routinely used in the KHP to measure quality of life, the factor structure and measurement invariance of the instrument have not been systematically evaluated using these panel data.

Measurement invariance refers to the degree to which a measurement instrument assesses the same construct in the same way across groups or time points, thereby enabling valid comparisons [12]. When comparing scores across groups or across time using the same scale, it is necessary to assume that the measurement properties (e.g., structure and meaning) of the instrument remain equivalent across those contexts [13]. For example, when comparing EQ-5D responses across languages—as in studies examining equivalence between English and Chinese versions—the stability and comparability of the measurement instrument must first be established [14]. Although several studies have investigated the measurement invariance of the EQ-5D-3L [2,4,9,14,15], evidence regarding longitudinal measurement invariance (LMI) remains limited.

HIGHLIGHTS

- Using Korea Health Panel data, we evaluated the EuroQol 5-dimensional questionnaire, 3-level version and confirmed a robust 1-factor structure with good model fit.
- Measurement invariance held across gender and age in panel 1; in panel 2, partial invariance was achieved when relaxing constraints on item 5.
- Longitudinal invariance was supported over 5–10 years (panel 1) and 3 years (panel 2), supporting valid group comparisons and longitudinal use in Korean adults.

Objectives

In this context, the objectives of this study were threefold. First, using panel 1 and panel 2 data from the KHP, we examined the factor structure of the EQ-5D-3L and tested measurement invariance across gender and age groups. Second, using panel 1 data, for which the panel survey has been completed, we evaluated short-term LMI over 5 years and long-term LMI over 10 years. Third, using panel 2 data that are currently being collected, we examined LMI across available survey waves. Through these analyses, we aimed to provide empirical evidence supporting the validity and applicability of the EQ-5D-3L for comparative and longitudinal research on HRQOL using KHP data.

Materials and Methods

Study Design and Setting

This study was a cross-sectional observational analysis based on secondary data from the KHP.

Setting and Participants

In this study, data from panel 1 (2008–2018) and panel 2 (2019–2021) of the KHP were used for analysis. The KHP is a nationally representative panel survey that provides data for detailed analyses of health care utilization, medical expenditures, and related factors [11]. It serves as an important source of evidence for establishing health and medical policy in the Republic of Korea. The survey was conducted using probability-proportional-to-size 2-stage stratified cluster sampling and included approximately 8,000 households nationwide.

The data used in this study included panel 1 data from the second survey year (2009), when the EQ-5D-3L was first introduced, through the 12th survey year (2017), and panel 2 data from the first survey year (2019) through the third survey year (2021), which were the most recent data released at the time of analysis. The analytic sample comprised 12,606 participants based on the second year of panel 1 and 11,593 participants based on the first year of panel 2.

Variables

The EQ-5D-3L, a measure of HRQOL, comprises 5 items: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each item is rated on a 3-point scale: 1=no problems, 2=some or moderate problems, and 3=extreme problems or unable to perform the activity. Higher scores indicate poorer quality of life [2,16].

Cronbach's α values for the EQ-5D-3L across survey years were as follows: panel 1, 2009 (0.704), 2010 (0.701), 2011 (0.728), 2012 (0.757), 2013 (0.760), 2015 (0.787), and 2017 (0.790); Panel 2, 2019 (0.762), 2020 (0.769), and 2021 (0.760).

Data Analysis

Participants and data collection

Confirmatory factor analysis (CFA) and measurement invariance analyses were conducted to evaluate the validity and measurement invariance of the EQ-5D-3L used in the KHP. The analyses used data from both panel 1 (2008–2018) and panel 2 (2019–2021), in which the EQ-5D-3L was administered. For panel 1, data from 12,606 individuals were analyzed using responses anchored to the second survey year (2009), when the EQ-5D-3L was first introduced. For panel 2, data from 11,593 individuals were analyzed using responses from the first survey year (2019).

Confirmatory factor analysis

CFA was conducted to examine the factor structure of the EQ-5D-3L. Specifically, we compared alternative factor structure models discussed in previous studies and selected the model judged to provide the most appropriate representation of the data.

Measurement invariance analysis

To evaluate measurement invariance across gender and age groups, we first examined the fit of the factor structure model within each group and then performed multi-group confirmatory factor analysis. Invariance testing proceeded sequentially through configural, weak, strong, and strict invariance [12,17–19]. For group comparisons, participants were categorized by gender as male or female and by age as younger than 65 years or 65 years and older. This approach allowed us to examine whether the EQ-5D-3L functioned consistently across gender and age groups.

Longitudinal measurement invariance

LMI was tested to assess the stability of measurement over time. For the factor structure model, all latent factors and observed indicators across the time points included in each analysis were specified simultaneously, and invariance testing proceeded through configural, weak, strong, and strict models [13,20–22].

Statistical Methods

IBM SPSS ver. 24.0 (IBM Corp.) was used to analyze the demographic characteristics of the study participants and the characteristics of the main variables. Mplus ver. 8.4 (Muthén & Muthén) was used to evaluate the factor structure model and measurement invariance. Because the EQ-5D-3L items are categorical, the weighted least squares mean and variance adjusted (WLSMV) estimator was used. Missing data were handled using the default WLSMV procedure in Mplus, which applies pairwise deletion for the estimation of polychoric correlations and thresholds. This approach uses all available pairs of observations; however, because pairwise deletion may still be sensitive to the missing-data mechanism, we additionally report the proportion of missing responses for each EQ-5D-3L item at each wave and interpret the longitudinal findings cautiously. Model fit was evaluated using the comparative fit index (CFI) and Tucker-Lewis index (TLI), with values of 0.90 or higher indicating acceptable fit, and the standardized root mean square residual (SRMR) and root mean square error of approximation (RMSEA), with values of 0.08 or lower indicating acceptable fit [23]. For model comparisons, invariance (or negligible differences) was considered supported when the decrease in CFI was no greater than 0.010 ($\Delta CFI \geq -0.010$), the increase in RMSEA was no greater than 0.015 ($\Delta RMSEA \leq 0.015$), and the increase in SRMR was no greater than 0.010 ($\Delta SRMR \leq 0.010$), which is appropriate for large-sample settings [24].

Ethics Statement

This study used publicly available secondary data without any personally identifiable information. As such, it does not constitute human subjects research and was exempt from Institutional Review Board (IRB) approval. All procedures were conducted in accordance with the ethical principles of the Declaration of Helsinki.

https://doi.org/10.24171/j.phrp.2025.0440

Results

Participants

Table 1 presents the demographic characteristics of participants from the 2 baseline periods, 2009 and 2019, including gender, age, marital status, and educational level. The gender distribution was similar across periods, with the proportion of male participants increasing slightly from 44.0% in 2009 to 44.6% in 2019. The mean age increased from 48.82 years (standard deviation [SD], 16.1) to 56.48 years (SD, 16.4). Marital status showed a decrease in the proportion of married participants, from 72.5% to 71.2%, and increases in the proportions of divorced participants, from 8.9% to 11.1%, and separated participants, from 2.4% to 5.4%. Educational level remained relatively consistent, with approximately one-third of participants having a university degree or higher in both periods.

Confirmatory Factor Analysis

Previous studies have proposed several factor structure models for the EQ-5D-3L, including a 1-factor model [2,4] and a 2-factor model that distinguishes activities (mobility, self-care, and usual activities) from symptoms (pain/discomfort and anxiety/depression). Accordingly, we compared these alternative structures to identify the most appropriate factor structure for the present data (Figure 1). The main results showed that all tested models demonstrated acceptable fit in both panel 1 and panel 2. Detailed factor loadings for all evaluated models are presented in Table S1. Although the χ^2 statistic was lower

Table 1. Participants' demographic characteristics

Characteristic	Panel 1, 2nd data (2009) (*n* = 12,606)	Panel 2, 1st data (2019) (*n* = 11,593)
Sex		
Male	5,544 (44.0)	5,170 (44.6)
Female	7,062 (56.0)	6,423 (55.4)
Age (y)	48.82 ± 16.1 (18, 98)	56.48 ± 16.4 (19, 99)
Marital status		
Married	9,142 (72.5)	8,250 (71.2)
Bereaved	69 (0.5)	74 (0.6)
Divorced	1,117 (8.9)	1,289 (11.1)
Separated	302 (2.4)	627 (5.4)
Single	1,976 (15.7)	1,352 (11.7)
Educational background		
Below elementary school graduation	2,792 (22.1)	2,578 (22.2)
Middle school graduation	1,470 (11.7)	1,593 (13.7)
High school graduation	4,102 (32.5)	3,498 (30.2)
University graduation or higher	4,242 (33.7)	3,924 (33.8)

Data are presented as *n* (%) or mean ± standard deviation (minimum, maximum).

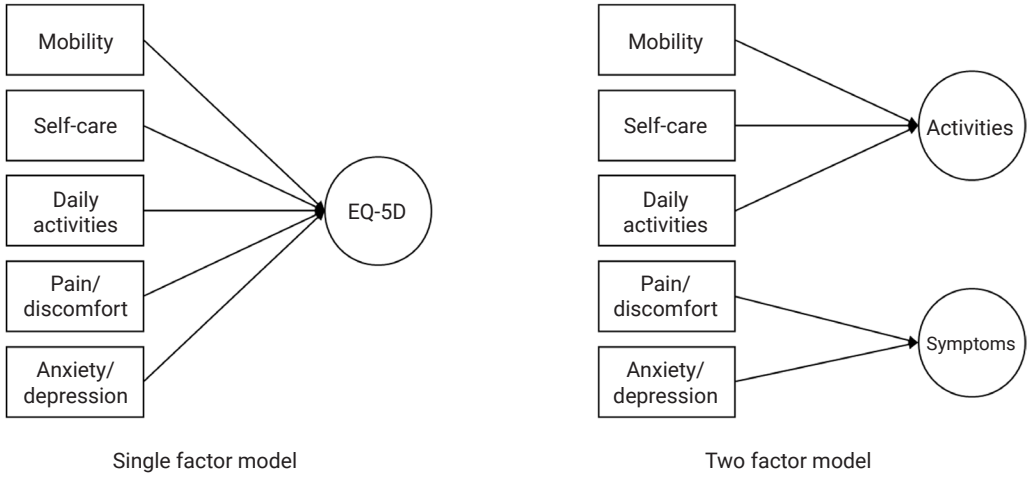


Figure 1. Factor structure model of EuroQol 5-dimensional questionnaire, 3-level version (EQ-5D-3L).

for the 2-factor model than for the 1-factor model, this difference is likely attributable to the sensitivity of the χ^2 test to trivial deviations in very large samples (panel 1, n=12,606; panel 2, n=11,593). In such settings, it is generally more informative to evaluate changes in incremental fit indices, such as the CFI and TLI, and absolute fit indices, such as the RMSEA, than to rely solely on χ^2 differences [24]. When comparing the 2 models, the ΔCFI, ΔSRMR, and ΔRMSEA values remained within the recommended cutoffs shown in Table 2, indicating that the difference in fit between the models was negligible. Under the principle of parsimony, simpler models are preferred when they explain the data comparably well. Because the 1-factor model already demonstrated excellent fit (CFI > 0.99, TLI > 0.99, RMSEA < 0.06), adopting a more complex 2-factor model was not necessary. In addition, the strong fit of the 1-factor model supports the interpretation that the 5 items reflect a common latent construct. Given the minimal gain in model fit and the consistent replication of the 1-factor structure across both periods, the 1-factor model was selected as the final model.

Validation of the Group-Specific Baseline Model

Before testing measurement invariance across gender and age groups, we examined the fit of the baseline model separately within each group. The main results are presented in Table 3. For the second survey year of panel 1 and the first survey year of panel 2, the CFA results indicated that all group-specific models demonstrated adequate fit.

Measurement Invariance

Measurement invariance across gender and age groups was evaluated using 2 datasets. First, Table 4 presents the results based on the second survey year of panel 1. In this analysis,

Table 2. Factor structure analysis of the EQ-5D-3L

	χ^2	df	CFI	TLI	SRMR	RMSEA (90% CI)
Panel 1, 2nd data in 2009 (n=12,606)						
One factor	187.412	5	0.996	0.991	0.035	0.054 (0.047–0.061)
Two factors	122.743	4	0.997	0.993	0.025	0.049 (0.041–0.056)
Panel 2, 1st data in 2019 (n=11,593)						
One factor	179.536	5	0.998	0.995	0.027	0.055 (0.048–0.062)
Two factors	123.406	4	0.998	0.996	0.019	0.051 (0.043–0.059)

EQ-5D-3L, EuroQol5-dimensional questionnaire,3-level version; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker-Lewis index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation; CI, confidence interval.

Table 3. Results of analysis of the baseline model for data by gender and age groups of EQ-5D-3L

	Model fit index					
	χ^2	df	CFI	TLI	SRMR	RMSEA (90% CI)
Panel 1, 2nd data in 2009 (n=12,606)						
Sex						
Male (n=5,544)	53.428	5	0.997	0.994	0.027	0.042 (0.032–0.052)
Female (n=7,062)	109.636	5	0.996	0.992	0.038	0.054 (0.046–0.064)
Age						
Under 65 years (n=9,979)	88.177	5	0.994	0.988	0.035	0.041 (0.034–0.049)
Older 65 years (n=2,627)	109.212	5	0.992	0.985	0.045	0.089 (0.075–0.104)
Panel 2, 1st data in 2019 (n=11,593)						
Sex						
Male (n=5,170)	55.660	5	0.998	0.997	0.025	0.044 (0.034–0.055)
Female (n=6,423)	132.613	5	0.997	0.994	0.029	0.063 (0.054–0.073)
Age						
Under 65 years (n=7,388)	82.525	5	0.994	0.989	0.039	0.046 (0.037–0.055)
Older 65 years (n=4,205)	109.946	5	0.997	0.994	0.027	0.071 (0.060–0.082)

EQ-5D-3L, EuroQol5-dimensional questionnaire,3-level version; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker-Lewis index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation; CI, confidence interval.

https://doi.org/10.24171/j.phrp.2025.0440

strict invariance was supported across both gender groups and age groups, indicating that measurement invariance was well maintained in panel 1.

Next, Table 5 presents the results for the first wave of panel 2 (2019). The configural model showed acceptable fit in both the gender-group and age-group analyses, indicating that the overall 1-factor structure was comparable across groups. However, when factor loadings were constrained to equality, corresponding to weak (metric) invariance, the increase in RMSEA relative to the configural model exceeded the recommended threshold (ΔRMSEA > 0.015), suggesting that full metric invariance was not supported in panel 2 [24].

Accordingly, we proceeded to test partial measurement

Table 4. Results of measurement invariance verification for gender and age groups of EQ-5D-3L (panel 1 data)

Invariance model	Model fit index					Model comparison		
	χ^2	df	CFI	SRMR	RMSEA (90% CI)	ΔCFI	ΔSRMR	ΔRMSEA
1st period 2nd data (2009)								
Sex group								
Configural	167.031	10	0.996	0.033	0.050 (0.043–0.057)			
Weak	138.746	14	0.997	0.034	0.038 (0.032–0.043)	0.001	0.001	−0.012
Strong	170.025	18	0.996	0.034	0.037 (0.032–0.042)	−0.001	0.000	−0.001
Strict	191.703	23	0.996	0.035	0.034 (0.030–0.039)	0.000	0.001	−0.003
Age group								
Configural	196.866	10	0.993	0.037	0.054 (0.048–0.061)			
Weak	152.500	14	0.995	0.038	0.040 (0.034–0.045)	0.002	0.001	−0.014
Strong	251.910	18	0.991	0.039	0.045 (0.041–0.050)	−0.004	0.001	0.005
Strict	371.193	23	0.987	0.040	0.049 (0.045–0.053)	−0.004	0.001	0.004

ΔCFI, ΔRMSEA, and ΔSRMR were computed as (fit index of the more constrained model) minus (fit index of the less constrained model). Thus, negative ΔCFI indicates a decrease in CFI under additional constraints.

EQ-5D-3L, EuroQol5-dimensional questionnaire, 3-level version; df, degrees of freedom; CFI, comparative fit index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation; CI, confidence interval; ΔCFI, change in comparative fit index; ΔSRMR, change in standardized root mean square residual; ΔRMSEA, change in root mean square error of approximation.

Table 5. Results of measurement invariance verification for gender and age groups of EQ-5D-3L (panel 2 data)

Invariance model	Model fit index					Model comparison		
	χ^2	df	CFI	SRMR	RMSEA (90% CI)	ΔCFI	ΔSRMR	ΔRMSEA
2nd period 1st data (2019)								
Sex group								
Configural	194.145	10	0.997	0.028	0.056 (0.050–0.063)			
Weak	134.378	14	0.998	0.028	0.039 (0.033–0.045)	0.001	0.000	−0.017
Partial weak (with factor loading of item 5 free)	173.535	13	0.998	0.028	0.046 (0.040–0.052)	0.001	0.000	−0.010
Partial strong	190.764	17	0.997	0.028	0.042 (0.037–0.047)	−0.001	0.000	−0.004
Partial strict	192.380	22	0.998	0.029	0.037 (0.032–0.041)	0.001	0.001	−0.005
Age group								
Configural	189.829	10	0.996	0.036	0.056 (0.049–0.063)			
Weak	134.620	14	0.997	0.036	0.039 (0.033–0.045)	0.001	0.000	−0.017
Partial weak (with factor loading of item 5 free)	166.224	13	0.997	0.036	0.045 (0.039–0.051)	0.001	0.000	−0.011
Partial strong	250.564	17	0.995	0.037	0.049 (0.043–0.054)	−0.002	0.001	0.004
Partial strict	372.138	22	0.993	0.049	0.052 (0.048–0.057)	−0.002	0.012	0.003

ΔCFI, ΔRMSEA, and ΔSRMR were computed by comparing each model with the immediately preceding, less constrained model (e.g., metric vs. configural; strong vs. metric; strict vs. strong). For partial invariance models, changes were computed relative to the corresponding preceding model in the partial sequence (e.g., partial metric vs. configural; partial strong vs. partial metric; partial strict vs. partial strong). Negative ΔCFI indicates a decrease in CFI under additional constraints.

EQ-5D-3L, EuroQol5-dimensional questionnaire,3-level version; df, degrees of freedom; CFI, comparative fit index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation; CI, confidence interval; ΔCFI, change in comparative fit index; ΔSRMR, change in standardized root mean square residual; ΔRMSEA, change in root mean square error of approximation.

invariance by relaxing equality constraints on items showing evidence of non-invariance [20,25–27]. Inspection of modification indices (MI) and expected parameter changes (EPC) identified item 5 (anxiety/depression) as the primary source of non-invariance in the weak invariance model, with substantial misfit observed in both the gender-group analysis (MI, 32.096; EPC, 0.262) and the age-group analysis (MI, 27.892; EPC, 0.213). On this basis, we relaxed the loading constraint for item 5 and evaluated partial invariance primarily using changes in approximate fit indices and substantive interpretability rather than relying solely on raw chi-square statistics under WLSMV estimation [23]. This adjustment also acknowledges the possibility that the anxiety/depression dimension was particularly sensitive to socioenvironmental changes during panel 2, including the coronavirus disease 2019 (COVID-19) pandemic.

Subsequent analyses showed that, after freeing item 5, the partial weak and partial strong invariance models were supported for both gender and age groups (Table 5). For gender groups, partial strict invariance was also supported. For age groups, however, the transition from the partial strong to the partial strict invariance model resulted in a change in SRMR that exceeded the recommended threshold (ΔSRMR > 0.010). Therefore, in panel 2, comparisons between age groups should be interpreted at the level of partial strong invariance, and strict invariance across age groups was not fully supported [24].

Validation of Baseline Models for Each Year for LMI

To evaluate LMI, analyses were conducted using data from both panel 1 and panel 2. For panel 1, short-term longitudinal invariance was examined using 5 waves from the second through the sixth survey years (LMI model 1), and long-term longitudinal invariance was examined using 5 waves from the second through the 10th survey years at 2-year intervals (LMI model 2). For panel 2, the analysis used the first 3 released waves, corresponding to the first through third survey years, to generate empirical evidence on the LMI of the EQ-5D-3L in the KHP.

Before testing LMI, CFA was performed to confirm that the annual datasets included in each analysis fit the selected EQ-5D-3L factor structure model. The results are presented in Table 6. All annual datasets demonstrated adequate fit to the study model.

Longitudinal Measurement Invariance

The results of the LMI analyses are presented in Table 7. In panel 1, both LMI model 1 and LMI model 2 were supported through the strict invariance level. In panel 2, LMI model 3 was also supported through the strict invariance level. These

Table 6. Baseline model analysis for each year's data to verify longitudinal measurement invariance

Division	Model fit index					
	χ^2	df	CFI	TLI	SRMR	RMSEA (90% CI)
LMI model 1						
Panel 1 2nd–6th data (5 wave) (*n* = 12,606)						
2nd (2009) (*n* = 12,606)	187.424	5	0.996	0.991	0.035	0.054 (0.047–0.061)
3rd (2010) (*n* = 10,719)	237.787	5	0.994	0.987	0.040	0.066 (0.059–0.073)
4th (2011) (*n* = 9,986)	235.915	5	0.994	0.988	0.038	0.068 (0.061–0.076)
5th (2012) (*n* = 9,186)	180.837	5	0.998	0.995	0.032	0.062 (0.054–0.070)
6th (2013) (*n* = 8,598)	159.142	5	0.997	0.994	0.029	0.060 (0.052–0.068)
LMI model 2						
Panel 1 2nd–10th data (5 wave) (*n* = 12,606)						
2nd (2009) (*n* = 12,606)	187.424	5	0.996	0.991	0.035	0.054 (0.047–0.061)
4th (2011) (*n* = 9,986)	235.914	5	0.994	0.988	0.038	0.068 (0.061–0.076)
6th (2013) (*n* = 8,598)	159.142	5	0.997	0.994	0.029	0.060 (0.052–0.068)
8th (2015) (*n* = 7,624)	106.522	5	0.998	0.997	0.022	0.052 (0.043–0.060)
10th (2017) (*n* = 7,086)	69.947	5	0.999	0.998	0.019	0.043 (0.034–0.052)
LMI model 3						
Panel 2 1st–3rd data (3 wave) (*n* = 11,593)						
1st (2019) (*n* = 11,593)	179.534	5	0.998	0.995	0.027	0.055 (0.048–0.062)
2nd (2020) (*n* = 10,034)	200.206	5	0.997	0.994	0.030	0.062 (0.055–0.070)
3rd (2021) (*n* = 9,091)	161.861	5	0.997	0.994	0.030	0.059 (0.051–0.067)

LMI, longitudinal measurement invariance; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker-Lewis index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation; CI, confidence interval.

https://doi.org/10.24171/j.phrp.2025.0440

Table 7. LMI verification and model comparison of the EQ-5D-3L scale

Invariance model	Model fit index					Model comparison		
	χ^2	df	CFI	SRMR	RMSEA (90% CI)	ΔCFI	ΔSRMR	ΔRMSEA
Panel 1 2nd–6th data								
Configural	1507.117	215	0.993	0.032	0.022 (0.021–0.023)			
Weak	1332.760	231	0.994	0.034	0.019 (0.018–0.020)	0.001	0.002	−0.003
Strong	1568.244	247	0.993	0.033	0.021 (0.020–0.022)	−0.001	−0.001	0.002
Strict	1514.836	267	0.993	0.034	0.019 (0.018–0.020)	0.000	0.001	−0.002
Panel 1 2nd–10th data								
Configural	1086.739	215	0.996	0.028	0.018 (0.017–0.019)	-	-	-
Weak	1099.223	231	0.996	0.032	0.017 (0.016–0.018)	0.000	0.004	−0.001
Strong	1199.627	247	0.995	0.030	0.017 (0.017–0.018)	−0.001	−0.002	0.000
Strict	1287.631	267	0.995	0.033	0.017 (0.016–0.018)	0.000	0.003	0.000
Panel 2 1st–3rd data								
Configural	667.491	72	0.996	0.027	0.027 (0.025–0.029)	-	-	-
Weak	539.324	80	0.997	0.028	0.022 (0.020–0.024)	0.001	0.001	−0.005
Strong	654.577	88	0.997	0.028	0.024 (0.022–0.025)	0.000	0.000	0.002
Strict	606.196	98	0.997	0.028	0.021 (0.020–0.023)	0.000	0.000	−0.003

LMI, longitudinal measurement invariance; EQ-5D-3L, EuroQol5-dimensional questionnaire,3-level version; df, degrees of freedom; CFI, comparative fit index; SRMR, standardized root mean square residual; RMSEA, root mean square error of approximation; CI, confidence interval; ΔCFI, change in comparative fit index; ΔSRMR, change in standardized root mean square residual; ΔRMSEA, change in root mean square error of approximation.

findings indicate that the EQ-5D-3L can be used appropriately in both short-term and long-term longitudinal studies of HRQOL.

Discussion

Key Results

This study examined the factor structure and measurement invariance of the EQ-5D-3L across time, gender, and age using nationally representative data from the KHP. The confirmatory factor analysis results consistently supported the original 5-item, 1-factor structure across all survey waves. This finding reinforces previous research suggesting that the EQ-5D descriptive system can be represented parsimoniously by a single latent construct [2,4]. Overall, the results indicate that the EQ-5D-3L maintains a stable and interpretable measurement structure over time while also revealing meaningful period- and group-specific differences in the degree of measurement invariance achieved. These findings highlight the importance of clearly specifying the conditions under which comparisons of HRQOL are interpreted across demographic groups and survey periods.

In panel 1 (2009–2017), strict measurement invariance was supported across both gender and age groups. This result indicates that the EQ-5D-3L functioned equivalently across these groups during this period, permitting valid comparisons of latent means, variances, and other structural parameters [23,24]. The robustness of measurement invariance observed in panel 1 is consistent with previous studies reporting stable psychometric properties of the EQ-5D-3L in general population samples [6,9,12]. Taken together, these findings support the use of the EQ-5D-3L for subgroup comparisons in Korean population health data collected prior to 2018 without substantial concern regarding measurement bias.

In contrast, the results for panel 2 (2019–2021) were more nuanced. Although configural invariance was supported across both gender and age groups, full metric invariance was not supported according to commonly recommended criteria based on changes in approximate fit indices [24]. Subsequent partial measurement invariance analyses identified item 5 (anxiety/depression) as the primary source of non-invariance. After relaxing the equality constraint for this item, the partial invariance models demonstrated acceptable fit. Specifically, partial strict invariance was supported for gender groups, whereas only partial strong invariance was supported for age groups. Accordingly, comparisons between age groups in panel 2 should be interpreted at the level of partial strong invariance rather than strict invariance, consistent with established methodological guidance for partial measurement invariance [20,25–27].

The differential performance of the anxiety/depression item in panel 2 suggests that this dimension may be particularly sensitive to temporal and contextual influences. The period after 2019 coincided with substantial socioenvironmental disruption, including the COVID-19 pandemic, which has been widely associated with changes in population mental health.

Compared with physical health domains such as mobility or self-care, mental health-related dimensions may be more responsive to external stressors and broader shifts in social context. Consequently, changes in social and psychological conditions during this period may have influenced how respondents across age groups interpreted and reported anxiety or depressive symptoms. These contextual effects could reduce measurement equivalence even when the overall factor structure remains stable. Importantly, this pattern should not be interpreted as a weakness of the EQ-5D-3L itself but rather as reflecting context-dependent variation in how mental health experiences are reported within population surveys.

With respect to longitudinal stability, the LMI analyses indicated that the measurement structure of the EQ-5D-3L remained stable across both short-term and longer-term intervals within each panel. These findings suggest that, despite the context-specific variation observed for the anxiety/depression dimension in panel 2, the underlying construct measured by the EQ-5D-3L remained largely consistent over time. As a result, the instrument appears suitable for longitudinal analyses of HRQOL in panel data when appropriate measurement invariance assumptions are verified.

From a methodological perspective, the present findings highlight the practical value of partial measurement invariance approaches when strict invariance assumptions are not fully satisfied. By identifying and freeing the parameter associated with item 5, the analysis preserved comparability for the remaining dimensions of the EQ-5D-3L while avoiding overly restrictive model constraints. This strategy aligns with contemporary recommendations in the structural equation modeling literature, which emphasize balancing statistical fit criteria with substantive interpretability when evaluating measurement invariance in large-scale survey data [20,23,25].

Several limitations should be considered when interpreting these findings. First, the EQ-5D-3L is known to exhibit ceiling effects, particularly in general population samples, which may limit its sensitivity to subtle changes in health status [6,9]. Such ceiling effects may partially contribute to the apparent stability observed in some invariance tests, particularly in panel 2. Second, although the present study focused on establishing measurement equivalence, it did not examine substantive differences in latent means or variances across groups. Future research should extend this work by investigating group differences in HRQOL once appropriate levels of measurement invariance have been established. Third, the age classification used in this study dichotomized respondents at 65 years to reflect policy-relevant definitions in the Republic of Korea. However, this approach may obscure heterogeneity within broader age categories. Future studies could adopt more granular age-group classifications or continuous measurement invariance approaches to better capture age-related variation in HRQOL. Finally, differences in sampling design and survey context between panel 1 and panel 2 of the KHP may have influenced response patterns. These structural differences should be considered when interpreting comparisons across survey periods.

In summary, the findings of this study demonstrate that the EQ-5D-3L generally maintains a stable measurement structure over time in the Korean population. At the same time, the results underscore the importance of careful, context-sensitive interpretation of group comparisons during periods characterized by substantial social and environmental change. Explicit consideration of partial measurement invariance is therefore essential for the valid use of the EQ-5D-3L in contemporary population health research and policy evaluation [23,24].

Conclusion

This study supports the validity and applicability of the EQ-5D-3L by examining its measurement invariance using KHP data. The findings provide empirical evidence that the instrument can be used appropriately in comparative and longitudinal studies of HRQOL across gender and age groups in the Republic of Korea.

Supplementary Material

Table S1. Participants' demographic characteristics. Supplementary data are available at https://doi.org/10.24171/j.phrp.2025.0440.

Notes

Ethics Approval

This study used de-identified secondary data from the KHP. Because the dataset is publicly available in anonymized form, IRB approval for this secondary analysis was waived in accordance with applicable regulations and institutional policy.

Conflicts of Interest

The authors have no conflicts of interest to declare.

Funding

This work was supported by the Ministry of Education of the Republic of Korea, the National Research Foundation of Korea (NRF-2023S1A5A2A01077879), and the 2024 Academic Research Support Program at Gangneung-Wonju National University.

Availability of Data

The data that support the findings of this study are available from the KHP. Access is subject to the KHP data access policy and may require an application; the authors did not receive special access privileges.

References

1. Bhadhuri A, Kind P, Salari P, et al. Measurement properties of EQ-5D-3L and EQ-5D-5L in recording self-reported health status in older patients with substantial multimorbidity and polypharmacy. Health Qual Life Outcomes 2020;18:317.
2. Buchholz I, Janssen MF, Kohlmann T, et al. A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. Pharmacoeconomics 2018;36:645–61.
3. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res 2013;22:1717–27.
4. Feng YS, Kohlmann T, Janssen MF, et al. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res 2021;30:647–73.
5. Macran S, Weatherly H, Kind P. Measuring population health: a comparison of three generic health status measures. Med Care 2003;41:218–31.
6. Johnson JA, Coons SJ. Comparison of the EQ-5D and SF-12 in an adult US sample. Qual Life Res 1998;7:155–66.
7. Sullivan PW, Lawrence WF, Ghushchyan V. A national catalog of preference-based scores for chronic conditions in the United States. Med Care 2005;43:736–49.
8. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20:1727–36.
9. Thompson AJ, Turner AJ. A comparison of the EQ-5D-3L and EQ-5D-5L. Pharmacoeconomics 2020;38:575–91.
10. Zhang RY, Wang W, Zhou HJ, et al. A head-to-head comparison of the EQ-5D-3L index scores derived from the two EQ-5D-3L value sets for China. Health Qual Life Outcomes 2022;20:80.
11. Korea Institute for Health and Social Affairs (KIHASA); National Health Insurance Service (NHIS) (KR). Korea Health Panel survey [Internet]. KIHASA; 2024 [cited 2024 Feb 4]. Available from: https://www.khp.re.kr:444. Korean.
12. Leitgob H, Seddig D, Asparouhov T, et al. Measurement invariance in the social sciences: historical development, methodological challenges, state of the art, and future perspectives. Soc Sci Res 2023;110:102805.
13. Liu Y, Millsap RE, West SG, et al. Testing measurement invariance in longitudinal data with ordered-categorical measures. Psychol Methods 2017;22:486–506.
14. Luo N, Chew LH, Fong KY, et al. Do English and Chinese EQ-5D versions demonstrate measurement equivalence?: an exploratory study. Health Qual Life Outcomes 2003;1:7.
15. Jiang R, Rand K, Kuharic M, et al. EQ-5D-5L measurement properties are superior to EQ-5D-3L across the continuum of health using US value sets. Health Qual Life Outcomes 2022;20:134.
16. Balestroni G, Bertolotti G. EuroQol-5D (EQ-5D): an instrument for measuring quality of life. Monaldi Arch Chest Dis 2012;78:155–9.
17. Lasker J. Measurement invariance testing works. PsyArXiv [Preprint] 2022 Apr 20. https://doi.org/10.31234/osf.io/r7e6f.
18. Svetina D, Rutkowski L, Rutkowski D. Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using Mplus and the lavaan/semtools packages. Struct Equ Modeling Multidiscip J 2020;27:111–30.
19. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ Res Methods 2000;3:4–70.
20. Mackinnon S, Curtis R, O'Connor R. A tutorial in longitudinal measurement invariance and cross-lagged panel models using lavaan. Meta-Psychology 2022;6:1–20.
21. Millsap RE. Testing measurement invariance using item response theory in longitudinal data: an introduction. Child Dev Perspect 2010;4:5–9.
22. Zhang L, Li Z. Factor structure and longitudinal measurement invariance of the K6 among a national representative elder sample of China. BMC Public Health 2022;22:1789.
23. Kline RB. Principles and practice of structural equation modeling. 5th ed. Guilford Press; 2023.
24. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Modeling Multidiscip J 2007;14:464–504.
25. Dimitrov DM. Testing for factorial invariance in the context of construct validation. Meas Eval Couns Dev 2010;43:121–49.
26. Sass DA. Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. J Psychoeduc Assess 2011;29:347–63.
27. Bowen NK, Masa RD. Conducting measurement invariance tests with ordinal data: a guide for social work researchers. J Soc Social Work Res 2015;6:229–49.