Depression is a major social issue that not only affects the interpersonal relationships and overall quality of life of individuals, but also potentially leads to suicide in many cases. The total number of people with depression is estimated to exceed 300 million worldwide, and depression has been ranked the single largest contributor to global disability (World Health Organization [WHO], 2017). Depression is the leading contributor to deaths by suicide, with close to 800,000 deaths per year worldwide (WHO, 2019). According to a meta-analysis that evaluated the aggregate prevalence of depression in multiple countries between 1994 and 2014, it was estimated that 10.8% of the people in the world are affected by depression at some point in their lives (Lim et al., 2018). A recent study by Liu et al. (2020) reported that the number of incident cases of depression worldwide has increased from 172 million in 1990 to 258 million in 2017, representing an alarming increase of 49.86%.

Similar to many nations, the negative effects of depression are also evident in Korea. Epidemiological surveys on mental health conducted every 5 years by the Ministry of Health & Welfare of Korea (2021) have reported that the prevalence of depression has increased gradually in Korea (2001: 4.0%; 2011: 6.7%; and 2021: 7.7%). In addition, Korea has been ranked top among the Organisation for Economic Cooperation and Development (OECD) countries in terms of suicide rates for over a decade (OECD, 2021).

Owing to the seriousness of depression, the early detection and treatment of risk groups are of significant importance. For the early detection and treatment of depression, an accurate evaluation of the existence and severity of depressive symptoms must be performed. Therefore, researchers and clinicians have dedicated much attention to the development of robust testing tools to measure depression accurately and quickly. Consequently, various measures of depression have been developed. The Center for Epidemiological Studies-Depression Scale (CES-D), a self-reported measure comprising 20 items, is one of the most widely used scales worldwide for measuring the degree of depression in the general population (Hann et al., 1999; Perreira et al., 2005; Vilagut et al., 2016).

While the full version of the CES-D scale has been frequently used in both research and clinical settings, its length (20 items) poses problems in large-scale survey research, where several measurements are usually incorporated (Boey, 1999). Thus, researchers have attempted to develop abbreviated versions of the CES-D scale to reduce the participants’ response burden (Carpenter et al., 1998). As a result, several abbreviated versions have been developed, including 5-item (Shrout & Yager, 1989), 8-item (Karim 1117799

"Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage)."
et al., 2015), 9-item (Santor & Coyne, 1997), 10-item (Andresen et al., 1994; Cole et al., 2004; Kohout et al., 1993; Meadows et al., 2006), 11-item (Kohout et al., 1993), and 12-item (Poulin et al., 2005) versions.

Among these abbreviated versions, the CES-D-11 is the most commonly used tool to measure depression in Korea. Hence, it is not surprising that many studies in Korea have attempted to examine the measurement properties of the CES-D-11. Overall, studies conducted in Korea on this scale can be divided into two broad categories. The first category of studies is related to identifying the factor structure of the scale. For example, Gweon (2009) and Kim and Kim (2008) reported that the four-factor model of depressed affect, positive affect, somatic complaints, and interpersonal problems presented by the original author (Radloff, 1977) were the most suitable. Conversely, Lee and Kang (2009) found that the most suitable scale was the five-factor model of depressed affect, interpersonal relationships, positive affect, slow activity, and physical condition. More recently, Hoe et al. (2015) investigated both the 4- and 5-factor models and recommended the use of the former, as it is consistent with the original authors’ suggestion. The second category of studies on the CES-D-11 is related to measurement invariance. For instance, Hoe et al. (2015) investigated whether the measure is invariant across gender and age groups, and deduced that factor mean invariance was supported for gender and scalar invariance across age groups.

Based on the above review of studies conducted in Korea, there is some evidence that the Korean version of the four-factor CES-D-11 is a valid measure suitable for use across gender and some age groups. However, the longitudinal invariance of the Korean version of this scale has received limited attention despite the fact that it is important to examine the presence (or absence) of such invariance when an instrument is administered in a longitudinal study that tracks changes over time. Moreover, previous studies have emphasized that it is not possible to determine whether temporal changes in a construct are due to actual changes or changes in the structure, or measurement of the construct over time without verifying longitudinal measurement invariance (Esnaola et al., 2019; Liu & West, 2018). Nevertheless, evidence of measurement invariance over time in the Korean version of the CES-D-11 is scarce in the existing literature.

Therefore, to bridge this gap in the literature, this research aims to examine the longitudinal measurement invariance of the Korean version of the CES-D-11 across time points to determine whether the scale has satisfactory properties for longitudinal comparisons and whether it can be effectively used to examine symptom changes across multiple time points. For this purpose, both short- and long-term longitudinal invariance were examined using two datasets. In Study 1, short-term longitudinal invariance was examined using baseline and second, third, and fourth follow-up data. In Study 2, long-term longitudinal invariance was examined using baseline and 4th, 7th, and 10th follow-up data.

**Methods**

**Participants**

The current study was conducted using data from the Korean Welfare Panel Study (KoWePS), which included data from a nationally representative sample of South Korean households. Households were selected using a stratified multistage probability sampling design, and data on household members aged 18 years or above were collected annually through face-to-face interviews, beginning in 2006.

The KoWePS extraction frame includes 230,000 enumeration districts excluding islands and special facilities from 90% of the Korean census population as of 2005. In the first stage, a total of 517 enumeration districts were sampled using 90% of the population census data. In the second stage, a total of 3,500 households with less than 60% of the median income and 3,500 households with more than 60% of the median income were extracted. Finally, panel households were selected using the stratified double extraction method totaling 7,000 households. For the purpose of this study, data for all household members from the baseline (2006) to the 10th survey year (2015) were included in the survey.

**Measure**

The CES-D-11 was included in the KoWePS to measure symptoms associated with depression experienced over the previous week, with four response options: (0) Rarely or none of the time (<1 day); (1) Some or a little of the time (2–3 days); (2) Occasionally or a moderate amount of time (4–5 days); and (3) Most or all of the time (6–7 days).

The psychometric properties of the CES-D-11 have been reported in the literature. Cronbach’s alphas for the 11 items ranged from .71 to .87 (Carpenter et al., 1998; Gellis, 2010; Kohout et al., 1993). Considering that a criterion of .70 to .90 is proposed as a measure of good internal consistency (Nunnally & Bernstein, 1994), the scale’s reliability was found to be satisfactory. The CES-D-11 scale has a high correlation of .95 on the 20-item scale (Kohout et al., 1993) and retains almost (87%) of the variance of the CES-D-20 (Covinsky et al., 2010). In addition, factor analytic studies have indicated that the two scales capture the same dimensions of depression with similar precision, including depressed affect, positive affect, somatic complaints, and interpersonal problems (Gellis, 2010; Kohout et al., 1993).

The measurement translation procedures and measurement properties, including the reliability and validity of the Korean version of the CES-D-11 scale have been reported in detail by Cho and Kim (1998). The authors translated and back-translated the scale twice to derive the final version. The final version’s reliability assessed by Cronbach’s $\alpha$ was .893 and the authors reported that the scale had a strong concurrent and good discriminant validity.
Analysis Plan

In this study, the data analysis was conducted using the following steps. First, based on previous studies that reported the factor structure of the CES-D-11 scale (Hoe et al., 2015; Kohout et al., 1993), confirmatory factor analysis was performed for each time point. Second, data were consecutively analyzed from the first to fourth waves in Study 1 to examine short-term longitudinal invariance. Third, long-term longitudinal invariance of the scale was examined in Study 2. Beginning with the first wave, data were extracted every 3 years up to the 10th year (waves 1, 4, 7, and 10).

A configural or form invariance model was initially estimated, with the loadings and thresholds being freely estimated. Next, a metric or weak invariance model was estimated, in which the factor loadings were constrained to be equal across time points. Then, a scalar or strong invariance model was estimated, in which the loadings and thresholds were constrained to be equal across time points. Finally, in addition to factor structure, loadings, and thresholds, a uniqueness or strict invariance model was estimated, in which residual variances were constrained to be equal across time points (Liu et al., 2017; Marsh et al., 2018; Meredith, 1993; Richardson, et al., 2020; Widaman et al., 2010; Winter & Depaoli, 2020).

A list-wise deletion was employed to address the missing data in the data analyses. The final sample size was 10,098 in Study 1 and 7,077 in Study 2, which were large enough for the estimation.

Results

Demographic Characteristics

Participants’ demographic characteristics are presented in Table 1. In Study 1, 56.6% of the participants were female, and the $M_{age}$ was 51 years ($SD=16.8$; range = 18–99). In addition, the majority of the participants were married (69.6%),
and more than 54% had a high school education or higher. In Study 2, the mean age was 50.9 years (SD = 15.7; range = 18–99), while 58.3% of the sample were women. Moreover, approximately two-thirds (72.2%) of the participants were married, and more than half of the sample (53.2%) had a high school education or higher.

Psychometric Properties of the CES-D-11

The CES-D-11 scale items with their factor structure and psychometric properties are presented in Table 2. The scale consists of four factors including depressed affect (three items), positive affect (two items), somatic complaints (four items), and interpersonal problems (two items).

The reliability of the scale was assessed using Cronbach’s alpha and McDonald’s omega coefficient. The McDonald’s omega coefficient ranged from .864 to .894, and Cronbach’s alpha ranged from .852 to .887, indicating that the scale has satisfactory internal consistency.

The floor and ceiling effects of the scale were calculated as the percentage of participants who reported the lowest score of 0 or highest score of 3 for each of the 11 items. Floor and ceiling effects were considered present if more than 15% of the participants had either the lowest possible score (floor effect) or the highest possible score (ceiling effect; Terwee et al., 2007). As shown in Table 2, there were no ceiling effects for the CES-D-11 throughout the study periods. However, all items showed floor effects ranging between 43% and 96%. Considering that the sample of this study was not a clinical population but the general public, one possible explanation for the presence of floor effects is that only a limited number of participants had depressive symptoms.

Confirmatory Factor Analysis of the Baseline Model

Before examining the longitudinal measurement invariance, it is important to establish a baseline model that fits well with the data across time points (Byrne & Watkins, 2003; Sass, 2011). Although one study reported that the Korean version of the CES-D-11 consists of a 5-factor model (Lee & Kang, 2009), other studies (Gweon, 2009; Hoe et al., 2015; Kim & Kim, 2008) concluded that the scale is suitable as a 4-factor model, consistent with the original author’s suggestion (Kohout et al., 1993). Thus, the 4-factor model was adopted in the current study as a baseline and tested to ascertain that it fits well with the data across time points.

As presented in Table 3, the CFI and TLI values were greater than the cut-off point of 0.95, while the SRMR and RMSEA values were less than the cut-off point of 0.08, indicating that the baseline model matched well with the data at all time points. This also allowed for further investigation of the longitudinal measurement invariance.

Descriptive Statistics

Table 4 presents the descriptive statistics of the factor scores at the baseline and follow-up periods in this study. The mean of the depressed affect ranged from 0.92 to 1.55, positive affect ranged from 0.8 to 1.65, somatic complaints ranged from 1.72 to 2.63, and interpersonal problems ranged from 0.09 to 0.24. The medians and inter-quartile range (Q1–Q3) were also reported to describe the distribution of each factor. The median scores for the depressed affect ranged from 0 to 1 (Q1–Q3, 0–3), positive affect ranged from 0 to 1 (Q1–Q3, 0–3), somatic complaints ranged from 1 to 2 (Q1–Q3, 0–4), and interpersonal problems ranged from 0 to 1 (Q1–Q3, 0–1).

In addition, Cronbach’s alphas, which were calculated for each factor throughout the time points showed that they were all within satisfactory levels; that is, they ranged from .705 to .848, except for positive affect at time 7 (.636).

Table 5 presents the correlation coefficients of the factor scores over time. The correlation coefficients of the factors ranged from r = .06 to .78 and they were all statistically significant. Because of the large sample size, the statistical significance of the correlations may not have a practical implication. Thus, Fisher’s Z transformed effect sizes were calculated to examine the magnitude of the relationship between the variables. The effect sizes ranged from .01 to 1.01, and the average effect size was 2.91, according to Cohen (1988), which corresponds to a medium effect size.

Longitudinal Measurement Invariance

The baseline model used in this study is shown in Figure 1. The responses to the 11 CES-D items within each measurement occasion were regressed on four common factors. The common factors were allowed to correlate across time intervals, and the residuals of the same response variables were allowed to correlate across time intervals simultaneously. In subsequent analyses, the models were specified by progressively constraining additional parameters (factor loadings, item thresholds, and residual variances) to remain equal across time.

The results of Study 1, which examined the short-term longitudinal invariance of the CES-D-11, are present in Table 6. The baseline model of configural invariance was acceptable (CFI = 0.976; TLI = 0.967; RMSEA = 0.017). Next, the metric invariance model fit was adequate (CFI = 0.974; TLI = 0.967; RMSEA = 0.026), whereas the differences in CFI and RMSEA between the configural and metric invariance models were negligible (ΔCFI = −0.001; ΔRMSEA = −0.001). The scalar invariance model provided a satisfactory fit (CFI = 0.973; TLI = 0.967; RMSEA = 0.026), whereas the changes in CFI and RMSEA were negligible (ΔCFI = −0.001; ΔRMSEA = −0.001). Finally, the residual invariance model was shown to adequately fit the data (CFI = 0.969; TLI = 0.963; RMSEA = 0.027), with negligible differences in CFI and RMSEA between the strong and strict invariance
Table 2. Scale Items, Factors, and Reliabilities of the Scale.

| Item                  | Wave 1 (N = 10,098)                          | Wave 2 (N = 10,098)                          | Wave 3 (N = 10,098)                          | Wave 4 (N = 10,098)                          |
|-----------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| **Factor**            | **Content**                         | Score (0–3) M (SD) Floor/ceiling effects (%) Reliability | Score (0–3) M (SD) Floor/ceiling effects (%) Reliability | Score (0–3) M (SD) Floor/ceiling effects (%) Reliability |
| **Depressed affect (DA)** | I felt quite depressed             | 0.52 (0.793) 63.4/3.7 Cronbach’ α = .887 | 0.54 (0.808) 62.0/3.9 Cronbach’ α = .869 | 0.49 (0.729) 63.2/2.3 Cronbach’ α = .861 |
|                       | I felt lonely                      | 0.49 (0.801) 66.5/4.0 McDonald’s ω = .894 | 0.51 (0.815) 56.3/4.2 McDonald’s ω = .869 | 0.44 (0.730) 67.4/2.4 McDonald’s ω = .863 |
| Positive affect (PA)  | My heart felt sad                 | 0.46 (0.763) 67.1/3.3 McDonald’s ω = .894 | 0.49 (0.776) 64.9/3.5 McDonald’s ω = .878 | 0.41 (0.691) 68.6/2.0 McDonald’s ω = .870 |
|                       | I felt that I was doing generally well | 0.69 (0.905) 55.1/6.1 | 0.65 (0.918) 59.9/6.0 | 0.56 (0.832) 62.9/3.9 | 0.59 (0.842) 60.1/4.0 |
|                       | I went on without much complaints  | 0.95 (1.02) 43.7/10.8 | 0.82 (1.01) 53.2/9.0 | 0.74 (0.943) 54.7/6.4 | 0.75 (0.911) 51.8/5.5 |
| Somatic complaints (SC) | I did not feel like eating; my appetite was poor | 0.48 (0.787) 66.6/3.6 | 0.53 (0.827) 64.3/4.6 | 0.50 (0.783) 65.0/3.4 | 0.497 (0.787) 66.3/3.4 |
|                       | I felt difficulty in everything I did | 0.82 (0.909) 45.4/6.8 | 0.90 (0.934) 61.0/8.2 | 0.80 (0.867) 44.6/5.1 | 0.76 (0.870) 47.4/5.2 |
|                       | I could not sleep well             | 0.74 (0.941) 53.6/7.2 | 0.72 (0.919) 53.6/6.5 | 0.66 (0.868) 72.1/3.7 | 0.67 (0.884) 55.5/5.4 |
|                       | I did not have the courage to carry out something | 0.47 (0.815) 69.0/4.4 | 0.49 (0.830) 68.5/4.7 | 0.42 (0.773) 72.1/3.7 | 0.38 (0.740) 74.3/3.2 |
| Interpersonal problems (IP) | I felt that people were treating me coldly | 0.14 (0.447) 89.5/0.8 | 0.11 (0.402) 91.1/0.6 | 0.10 (0.370) 91.9/0.4 | 0.09 (0.353) 92.6/0.4 |
|                       | I felt that people disliked me      | 0.10 (0.381) 92.0/0.5 | 0.09 (0.348) 93.2/0.3 | 0.07 (0.297) 94.0/0.2 | 0.07 (0.315) 94.0/0.3 |

| Item                  | Wave 1 (N = 7,077)                          | Wave 4 (N = 7,077)                          | Wave 7 (N = 7,077)                          | Wave 10 (N = 7,077)                         |
|-----------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| **Factor**            | **Content**                         | Score (0–3) M (SD) Floor/ceiling effects (%) Reliability | Score (0–3) M (SD) Floor/ceiling effects (%) Reliability | Score (0–3) M (SD) Floor/ceiling effects (%) Reliability |
| **Depressed affect (DA)** | I felt quite depressed             | 0.50 (0.779) 64.5/3.4 Cronbach’ α = .883 | 0.42 (0.700) 68.1/2.1 Cronbach’ α = .857 | 0.38 (0.663) 70.2/1.7 Cronbach’ α = .852 |
|                       | I felt lonely                      | 0.47 (0.784) 67.9/3.6 McDonald’s ω = .891 | 0.38 (0.680) 70.9/1.9 McDonald’s ω = .864 | 0.30 (0.598) 75.8/1.1 McDonald’s ω = .864 |
| Positive affect (PA)  | My heart felt sad                 | 0.45 (0.747) 67.6/3.0 McDonald’s ω = .891 | 0.37 (0.661) 71.3/1.7 McDonald’s ω = .864 | 0.26 (0.568) 79.0/1.0 McDonald’s ω = .864 |
|                       | I felt that I was doing generally well | 0.68 (0.903) 56.1/6.0 | 0.56 (0.821) 61.5/3.5 | 0.34 (0.667) 76.1/1.5 | 0.37 (0.692) 73.7/1.5 |
|                       | I went on without much complaints  | 0.94 (1.02) 44.6/10.6 | 0.72 (0.897) 53.0/5.0 | 0.46 (0.819) 70.7/4.4 | 0.43 (0.757) 70.5/2.4 |
| Somatic complaints (SC) | I did not feel like eating; my appetite was poor | 0.46 (0.772) 67.7/3.5 | 0.46 (0.759) 67.6/2.8 | 0.34 (0.668) 75.2/2.1 | 0.37 (0.712) 74.6/2.3 |
|                       | I felt difficulty in everything I did | 0.80 (0.901) 46.1/6.5 | 0.74 (0.854) 48.5/4.8 | 0.55 (0.778) 59.4/3.5 | 0.53 (0.772) 61.0/3.0 |
|                       | I could not sleep well             | 0.73 (0.936) 53.7/7.1 | 0.65 (0.872) 56.2/5.1 | 0.53 (0.815) 63.4/4.1 | 0.55 (0.839) 63.3/4.1 |
|                       | I did not have the courage to carry out something | 0.44 (0.786) 70.3/3.9 | 0.35 (0.698) 75.7/2.5 | 0.30 (0.647) 77.6/2.3 | 0.35 (0.683) 74.3/2.1 |
| Interpersonal problems (IP) | I felt that people were treating me coldly | 0.13 (0.430) 90.3/0.8 | 0.08 (0.302) 93.4/0.3 | 0.05 (0.252) 95.7/0.1 | 0.08 (0.323) 93.2/0.3 |
|                       | I felt that people disliked me      | 0.09 (0.359) 92.6/0.4 | 0.06 (0.287) 95.3/0.3 | 0.04 (0.218) 96.8/0.1 | 0.06 (0.287) 94.6/0.2 |
Table 3. Confirmatory Factor Analysis of the Baseline Model at Each Time Point.

| Model fit indices | Study 1 | Study 2 |
|-------------------|---------|---------|
|                   | Time 1  | Time 1  | Time 2  | Time 4  | Time 7  | Time 10 |
| $\chi^2$          | 1,476.258 | 1,035.233 | 1,274.340 | 1,476.258 | 947.810 | 1,123.762 |
| df                | 38      | 38      | 38      | 38      | 38      | 38      |
| CFI               | 0.973   | 0.961   | 0.965   | 0.961   | 0.959   | 0.960   |
| TLI               | 0.961   | 0.979   | 0.976   | 0.973   | 0.971   | 0.972   |
| SRMR              | 0.025   | 0.022   | 0.023   | 0.025   | 0.028   | 0.025   |
| RMSEA             | 0.059 (0.056–0.061) | 0.048 (0.046–0.051) | 0.052 (0.049–0.054) | 0.055 (0.052–0.057) | 0.056 (0.053–0.059) | 0.061 (0.058–0.064) |

Table 4. Descriptive Statistics of the Factor Scores at Each Time Point.

| Time Point | Factor               | M (SD)   | Median (Q1–Q3) | Variance | Cronbach's alpha |
|------------|----------------------|----------|----------------|----------|------------------|
| Study 1    | Time 1 (2006)        | Depressed affect 1.47 (2.12) | 0 (0–3) | 4.265 | .848 |
|            | Positive affect 1.65 (1.73) | 1 (0–3) | 2.925 | .728 |
|            | Somatic complaints 2.51 (2.68) | 2 (0–4) | 7.162 | .776 |
|            | Interpersonal problems 0.24 (0.75) | 0 (0–0) | .565 | .780 |
|            | Time 2 (2007)        | Depressed affect 1.55 (2.08) | 1 (0–3) | 4.342 | .837 |
|            | Positive affect 1.47 (1.71) | 1 (0–3) | 2.911 | .714 |
|            | Somatic complaints 2.63 (2.62) | 2 (1–4) | 6.843 | .731 |
|            | Interpersonal problems 0.20 (0.68) | 0 (0–0) | .459 | .769 |
|            | Time 3 (2008)        | Depressed affect 1.34 (1.86) | 0 (0–2) | 3.451 | .831 |
|            | Positive affect 1.29 (1.56) | 1 (0–2) | 2.445 | .711 |
|            | Somatic complaints 2.38 (2.46) | 2 (0–4) | 6.034 | .706 |
|            | Interpersonal problems 0.17 (0.59) | 0 (0–0) | .353 | .726 |
|            | Time 4 (2009)        | Depressed affect 1.25 (1.84) | 0 (0–2) | 3.400 | .844 |
|            | Positive affect 1.34 (1.55) | 1 (0–2) | 2.387 | .711 |
|            | Somatic complaints 2.30 (2.44) | 2 (0–4) | 5.952 | .727 |
|            | Interpersonal problems 0.16 (0.61) | 0 (0–0) | .370 | .788 |
| Study 2    | Time 1 (2006)        | Depressed affect 1.42 (2.02) | 0 (0–2) | 4.092 | .848 |
|            | Positive affect 1.61 (1.70) | 1 (0–3) | 2.905 | .725 |
|            | Somatic complaints 2.44 (2.62) | 2 (0–4) | 6.872 | .770 |
|            | Interpersonal problems 0.22 (0.71) | 0 (0–0) | .504 | .755 |
|            | Time 4 (2009)        | Depressed affect 1.18 (1.77) | 0 (0–2) | 3.144 | .837 |
|            | Positive affect 1.29 (1.51) | 1 (0–2) | 2.289 | .708 |
|            | Somatic complaints 2.20 (2.36) | 2 (0–3) | 5.559 | .721 |
|            | Interpersonal problems 0.14 (0.55) | 0 (0–0) | .307 | .771 |
|            | Time 7 (2012)        | Depressed affect 0.95 (1.58) | 0 (0–2) | 2.481 | .823 |
|            | Positive affect 0.80 (1.28) | 0 (0–2) | 1.635 | .636 |
|            | Somatic complaints 1.72 (2.22) | 1 (0–3) | 4.911 | .754 |
|            | Interpersonal problems 0.09 (0.41) | 0 (0–0) | .172 | .705 |
|            | Time 10 (2015)       | Depressed affect 0.92 (1.58) | 0 (0–1) | 2.490 | .846 |
|            | Positive affect 0.81 (1.29) | 0 (0–1) | 1.673 | .743 |
|            | Somatic complaints 1.81 (2.37) | 1 (0–3) | 5.600 | .792 |
|            | Interpersonal problems 0.14 (0.56) | 0 (0–0) | .310 | .798 |

In Study 2, the long-term longitudinal measurement invariance was examined over a 10-year period. The findings are presented in Table 6.

The baseline model of configural invariance was acceptable (CFI=0.970; TLI=0.960; RMSEA=0.028). Next, the models ($\Delta$CFI=$-0.004$; $\Delta$RMSEA=$0.001$). Based on these findings, the residual invariance of the CES-D-11 scores across time was supported. Overall, the results of Study 1 suggest that the four-factor model of the CES-D-11 had strict invariance over the 4-year period.
Table 5. Correlation Coefficients Among the Factor Scores at Each Time Point.

| Study 1 | Study 2 | Time 1 | Time 4 | Time 7 | Time 10 |
|---------|---------|--------|--------|--------|---------|
|         |         | DA     | PA     | SC     | IP      | DA     | PA     | SC     | IP      | DA     | PA     | SC     | IP      | DA     | PA     | SC     | IP      |
| Time 1  | DA      | 1      | .553** | .760** | .431** | .336** | .210** | .316** | .105** | .335** | .192** | .304** | .150** | .282** | .200** | .258** | .135** |
|         | PA      | .561** | 1      | .588** | .261** | .249** | .171** | .262** | .089** | .225** | .150** | .232** | .114** | .228** | .170** | .225** | .127** |
|         | SC      | .766** | .592** | 1      | .363** | .305** | .206** | .349** | .092** | .314** | .194** | .326** | .137** | .287** | .212** | .296** | .146** |
|         | IP      | .441** | .281** | .378** | 1      | .146** | .091** | .131** | .126** | .174** | .115** | .139** | .139** | .116** | .077** | .094** | .108** |
| Time 2  | DA      | .439** | .323** | .395** | .190** | 1      | .486** | .694** | .373** | .350** | .204** | .304** | .154** | .284** | .193** | .258** | .151** |
|         | PA      | .301** | .240** | .306** | .144** | .547** | 1      | .526** | .239** | .218** | .186** | .210** | .109** | .202** | .165** | .224** | .107** |
|         | SC      | .398** | .320** | .429** | .159** | .706** | .583** | 1      | .290** | .338** | .221** | .354** | .140** | .303** | .235** | .337** | .149** |
|         | IP      | .158** | .126** | .144** | .178** | .362** | .257** | .306** | 1      | .107** | .080** | .068** | .133** | .103** | .064** | .073** | .111** |
| Time 3  | DA      | .392** | .290** | .360** | .175** | .449** | .281** | .395** | .159** | 1      | .471** | .712** | .319** | .349** | .238** | .315** | .190** |
|         | PA      | .247** | .217** | .259** | .137** | .276** | .242** | .291** | .120** | .508** | 1      | .469** | .252** | .216** | .186** | .220** | .118** |
|         | SC      | .351** | .285** | .390** | .154** | .406** | .308** | .453** | .158** | .697** | .544** | 1      | .258** | .325** | .256** | .355** | .177** |
|         | IP      | .151** | .103** | .127** | .142** | .157** | .126** | .133** | .186** | .353** | .264** | .306** | 1      | .152** | .115** | .114** | .158** |
| Time 4  | DA      | .355** | .264** | .323** | .164** | .406** | .262** | .355** | .157** | .448** | .293** | .406** | .178** | 1      | .564** | .725** | .474** |
|         | PA      | .221** | .193** | .221** | .103** | .270** | .235** | .284** | .110** | .281** | .269** | .298** | .137** | .498** | 1      | .591** | .346** |
|         | SC      | .332** | .271** | .357** | .136** | .368** | .289** | .403** | .134** | .412** | .320** | .469** | .176** | .698** | .539** | 1      | .398** |
|         | IP      | .138** | .103** | .121** | .137** | .163** | .113** | .145** | .156** | .172** | .118** | .149** | .207** | .394** | .248** | .324** | 1      |

Note: DA = depressed affect; PA = positive affect; SC = somatic complaints; IP = interpersonal problems. The lower left diagonal of the table corresponds to the correlation table of subfactors at each time in Study 1. The upper right is the correlation table of sub-factors at each time in Study 2.

*p < .05. **p < .01.
Figure 1. Initial model used for the test of longitudinal invariance.
Note. For simplicity, the description of the model is only provided for time periods 1 and 2. In the actual analyses, data from all four-time points were simultaneously modeled.
metric invariance model fit was adequate (CFI = 0.968; TLI = 0.959; RMSEA = 0.021), whereas the differences in CFI and RMSEA between the configural and metric invariance models were negligible (ΔCFI = −0.002; ΔRMSEA = 0.001). The scalar invariance model provided a satisfactory fit (CFI = 0.964; TLI = 0.955; RMSEA = 0.022), whereas the changes in CFI and RMSEA were negligible (ΔCFI = −0.004; ΔRMSEA = 0.001). Finally, the residual invariance model was shown to adequately fit the data (CFI = 0.932; TLI = 0.919; RMSEA = 0.040). However, the model comparison fit indices indicated that the difference in CFI (i.e., ΔCFI = −0.032) exceeded the cut-off value of <0.01, failing to support the residual invariance model. Overall, the results of Study 2 indicated that the four-factor model of the CES-D-11 had scalar-level invariance over a 10-year period.

**Discussion and Conclusion**

This study presents a number of pertinent findings based on the results. First, the baseline model of the Korean version of the CES-D-11 scale, whose factor structure has been verified in previous studies, was tested to determine whether it adequately fits the data. The examination of the fit indices indicated that the baseline model matched well with the data at all time points in this study. Second, the results of Study 1, which examined short-term longitudinal measurement invariance, indicate that the strict invariance model holds true. Third, the results of Study 2, which examined long-term longitudinal measurement invariance, indicate that the longitudinal invariance model holds up to the scalar level.

Although an increasing number of longitudinal studies have evaluated the changes in the CES-D-11 scale (Chung & Kim, 2021; Jo & Choi, 2019; Lee, 2021; S. Lee & Park, 2021), only a few studies have systematically tested the assumption of temporal invariance. This is problematic since even if any changes over time are observed in the underlying construct, it is difficult to determine whether the changes are real or due to changes in the scale’s psychometric properties without evidence of measurement invariance. To the best of our knowledge, this is the first study to examine the longitudinal measurement invariance of the Korean version of the CES-D-11 scale using a large representative sample.

A strict level of invariance is ideal because it provides confidence that the group mean differences in the scale scores are driven by real group differences and not by other factors. However, achieving residual invariance can be difficult (Chen, 2007) and many researchers suggest that meeting scalar level invariance is considered sufficient to meaningfully compare factors or observed means (Bowen & Masa, 2015; Marsh et al., 2018; Richardson et al., 2020; Sedig & Leitgöb, 2018). The overall results from both the short- and long-term investigations in this study indicated that the Korean version of the CES-D-11 scale had scalar-level invariance over time.

Therefore, it can be concluded that the Korean version of the CES-D-11 is a valid measure for assessing both short- and long-term depressive symptoms over time. The results of this study also indicate that any observed changes in scale scores over time can be interpreted as actual changes.

The present study provides relevant implications for future research since the data used in this study were collected using the probability sampling method, and the sample size was sufficiently large to make a valid generalization. Thus, the findings of this study can be generalized to the Korean population. In addition, this study simultaneously investigated both the short- and long-term invariance of the Korean version of the CES-D-11. Previous studies on longitudinal invariance were generally short-term, thus limiting the generalizability of the findings over a longer time interval. However, this study confirmed that the scale could be used to track symptom changes in depression for up to 10 years.

Although the present study reveals important findings, it has a few limitations. First, the study participants were predominantly recruited from the general community. Hence, future research should evaluate scale scores in clinical samples. Second, one of the primary methodological issues in longitudinal studies is attrition. This study was not exempt from this issue. For example, in the baseline model, the number of people who responded to depression was 13,774; however, this number decreased to 7,077 in the 10th year (dropout rate of approximately 48%). At this time, it is unknown whether the participants who remained in this study differed significantly from those who dropped out. As missing cases

---

**Table 6. Fit Indices and Model Comparison.**

| Invariance model                        | Study 1 Baseline (Time 1) to Year 4 (Time 4) follow-up | Study 2 Baseline (Time 1) to Year 10 (Time 10) Follow-up |
|-----------------------------------------|--------------------------------------------------------|---------------------------------------------------------|
| Model fit index                         | χ² | df | TLI | CFI   | RMSEA | SRMR | CFI   | RMSEA | SRMR | CFI   | RMSEA | SRMR | CFI   | RMSEA | SRMR |
| Configural                              | 5,932.261 | 716 | 0.967 | 0.975 | 0.017 | 0.026 | —     | —     | —     | —     | —     | —     | —     | —     | —     |
| Weak                                    | 6,037.004 | 737 | 0.967 | 0.974 | 0.018 | 0.026 | —     | —     | —     | —     | —     | —     | —     | —     | —     |
| Strong                                   | 6,248.973 | 758 | 0.967 | 0.973 | 0.018 | 0.026 | —     | —     | —     | —     | —     | —     | —     | —     | —     |
| Strict                                   | 7,144.699 | 791 | 0.963 | 0.969 | 0.020 | 0.027 | —     | —     | —     | —     | —     | —     | —     | —     | —     |
| Configural                              | 5,107.187 | 716 | 0.960 | 0.970 | 0.020 | 0.028 | —     | —     | —     | —     | —     | —     | —     | —     | —     |
| Weak                                    | 5,352.219 | 737 | 0.959 | 0.968 | 0.021 | 0.029 | —     | —     | —     | —     | —     | —     | —     | —     | —     |
| Strong                                   | 5,976.580 | 758 | 0.955 | 0.964 | 0.022 | 0.030 | —     | —     | —     | —     | —     | —     | —     | —     | —     |
| Strict                                   | 10,621.627 | 791 | 0.919 | 0.932 | 0.030 | 0.040 | —     | —     | —     | —     | —     | —     | —     | —     | —     | —     |
cause problems in longitudinal studies, additional analyses were conducted to examine the impact of the missing data. The missing data were imputed and complete datasets were created using the expectation-maximization (EM) algorithm. The newly created datasets were analyzed, and the results were compared with those of the present study (data not shown). In Study 1, the missing imputed data showed a strong level of invariance whereas a strict level of invariance was observed in the present study. In Study 2, the same level of strong invariance was observed for both datasets. Taken together, it appears that missing data did have some impact; however, it was not strong enough to influence the overall conclusion of this study, considering that a scalar or strong invariance is considered sufficient.

In terms of future research directions, it has been well reported in the existing literature that the ways of expressing depressive symptoms vary by culture and country. For example, it is possible that the floor effects observed throughout the items in this study might be related to the Eastern collectivistic cultures, where the expression of depressive affect is more likely to be devalued (Zhang et al., 2011). In contrast, studies on self-esteem suggest a tendency for people from collectivist cultures to exhibit a neutral response bias and avoid the extreme ends of rating scales (Schmitt & Allik, 2005). Future research needs to identify the specific variables or mechanisms associated with the complex interplay between culture and the items of rating scales.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs
Keungeun Lee https://orcid.org/0000-0002-8330-6106
Sung-Woo Bae https://orcid.org/0000-0003-3200-5175

References
Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults: Evaluation of a short-form of the CES-D. American Journal of Preventive Medicine, 10(2), 77–84.
Boey, K. W. (1999). Cross-validation of a short form of the CES-D in Chinese elderly. International Journal of Geriatric Psychiatry, 14(8), 608–617.
Bowen, N. K., & Masa, R. D. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. Journal of the Society for Social Work and Research, 6(2), 229–249.
Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. Journal of Cross-Cultural Psychology, 34(2), 155–175.
Carpenter, J. S., Andrykowski, M. A., Wilson, J., Hall, L. A., Rayens, M. K., Sachs, B., & Cunningham, L. (1998). Psychometrics for two short forms of the center for epidemiologic studies-depression scale. Issues in Mental Health Nursing, 19(5), 481–494.
Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. Structural Equation Modeling: A Multidisciplinary Journal, 14(3), 464–504.
Cho, M. J., & Kim, K. H. (1998). Use of the center for epidemiologic studies depression (CES-D) scale in Korea. The Journal of Nervous & Mental Disease, 186(5), 304–310.
Chung, S., & Kim, J. (2021). Social determinants of depression among Korean adults: Results from a longitudinal study. Mental Health & Social Work, 49(1), 229–258.
Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Routledge.
Cole, J. C., Rabin, A. S., Smith, T. L., & Kaufman, A. S. (2004). Development and validation of a Rasch-derived CES-D short form. Psychological Assessment, 16(4), 360–372.
Covinsky, K. E., Yaffe, K., Lindquist, K., Cherkasova, E., Yelin, E., & Blazer, D. G. (2010). Depressive symptoms in middle age and the development of later-life functional limitations: The long-term effect of depressive symptoms. Journal of the American Geriatrics Society, 58(3), 551–556.
Esnoula, I., Benito, M., Antonio-Agirre, I., Axpe, I., & Lorenzo, M. (2019). Longitudinal measurement invariance of the satisfaction with life scale in adolescence. Quality of Life Research, 28(10), 2831–2837.
Gellis, Z. D. (2010). Assessment of a brief CES-D measure for depression in homebound medically ill older adults. Journal of Gerontological Social Work, 53(4), 289–303.
Gweon, H. S. (2009). Effects of problem drinking of elderly on life satisfaction mediated by depression and self-esteem: A latent means analysis application between poor and non-poor elderly. Journal of the Korean Gerontological Society, 29(4), 1521–1538.
Hann, D., Winters, K., & Jacobsen, P. (1999). Measurement of depressive symptoms in cancer patients: Evaluation of the center for epidemiological studies depression scale (CES-D). Journal of Psychosomatic Research, 46(5), 437–443.
Hoe, M. S., Park, B. S., & Bae, S. W. (2015). Testing measurement invariance of the 11-item Korean version CES-D scale: Used in the Korea welfare panel study. Mental Health & Social Work, 43(2), 313–339.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternative. Structural Equation Modeling, 6(1), 1–55.
The Jamovi Project. (2019). Jamovi (Version 0.9) [Computer Software]. https://www.jamovi.org
Jo, M., & Choi, H. (2019). The reciprocal relationship between self-esteem and depression in Korean adults. The Journal of Humanities and Social science, 10(4), 1049–1062.
Karim, J., Weisz, R., Bibi, Z., & Rehman, S. (2015). Validation of the eight-item center for epidemiologic studies depression scale (CES-D) among older adults. Current Psychology, 34(4), 681–692.
Kim, K. H., & Kim, J. H. (2008). The effects of self-esteem on the relationship between the elderly depression and life satisfaction. Family and Culture, 20(2), 95–116.
Kline, R. B. (2005). Principles and practice of structural equation modeling (2nd ed.). The Guilford Press.
Kohout, F. J., Berkman, L. F., Evans, D. A., & Cornoni-Huntley, J. (1993). Two shorter forms of the CES-D depression symptoms index. *Journal of Aging and Health, 5*(2), 179–193.

Lee, H. H. (2021). Trajectory of development of depression and problem drinking in adults: Focused on the convergence factors of basic livelihood receipt and disabled people. *Journal of the Korea Convergence Society, 12*(5), 303–311.

Lee, H. J., & Kang, S. K. (2009). The relationships between stressors, psychosocial resources, and depression among individuals with disabilities. *Mental Health & Social Work, 33*, 193–217.

Lee, S., & Park, W. (2021). Influence of multidimensional poverty experience on a longitudinal-change patterns of depression in elderly. *The Journal of Humanities and Social Sciences, 21*(4), 405–416.

Lim, G. Y., Tam, W. W., Lu, Y., Ho, C. S., Zhang, M. W., & Ho, R. C. (2018). Prevalence of depression in the community from 30 countries between 1994 and 2014. *Scientific Reports, 8*(1), 1–10.

Liu, Q., He, H., Yang, J., Feng, X., Zhao, F., & Lyu, J. (2020). Changes in the global burden of depression from 1990 to 2017: Findings from the global burden of disease study. *Journal of Psychiatric Research, 126*, 134–140.

Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods, 22*(3), 486–506.

Liu, Y., & West, S. G. (2018). Longitudinal measurement non-invariance with ordered-categorical indicators: How are the parameters in second-order latent linear growth models affected? *Structural Equation Modeling: A Multidisciplinary Journal, 25*(5), 762–777.

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multigroup factor analysis comparison of latent means across many groups. *Psychological Methods, 23*(3), 524–545.

Meadows, S. O., Brown, J. S., & Elder, G. H. (2006). Depressive symptoms, stress, and support: Gendered trajectories from adolescence to young adulthood. *Journal of Youth and Adolescence, 35*(1), 93–103.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Ministry of Health & Welfare of Korea. (2021). *The survey of mental disorders in Korea 2021.* [https://mhs.ncmh.go.kr/front/en/infographic.do](https://mhs.ncmh.go.kr/front/en/infographic.do)

Muthén, L. K., & Muthén, B. O. (2019). *Mplus. Version 8.4* (Computer program). Muthén & Muthén.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Organization for Economic Co-operation and Development (OECD). (2021). *Health status: Suicide rates.* Author.

Perreira, K. M., Deeb-Sossa, N., Harris, K. M., & Bollen, K. (2005). What are we measuring? An evaluation of the CES-D across race/ethnicity and immigrant generation. *Social Forces, 83*(4), 1567–1601.

Poulin, C., Hand, D., & Boudreau, B. (2005). Validity of a 12-item version of the CES-D center for epidemiological studies depression scale used in the national longitudinal study of children and youth. *Chronic Diseases in Canada, 26*(2–3), 65–72.

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401.

Richardson, G. B., Smith, R., Lowe, L., & Acquavita, S. P. (2020). Structure and longitudinal invariance of the short alcohol and alcohol problems perception questionnaire. *Journal of Substance Abuse Treatment, 115*, 108041.

Santor, D. A., & Coyne, J. C. (1997). Shortening the CES-D to improve its ability to detect cases of depression. *Psychological Assessment, 9*(3), 233–243.

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*(4), 347–363.

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg self-esteem scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology, 89*(4), 623–642.

Seddig, D., & Leitgöb, H. (2018). Approximate measurement invariance and longitudinal confirmatory factor analysis: Concept and application with panel data. *Survey Research Methods, 12*(1), 29–41.

Shrout, P. E., & Yager, T. J. (1989). Reliability and validity of screening scales: Effect of reducing scale length. *Journal of Clinical Epidemiology, 42*(1), 69–78.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Buter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42.

Vilagut, G., Forero, C. G., Barbaglia, G., & Alonso, J. (2016). Screening for depression in the general population with the center for epidemiologic studies depression (CES-D): A systematic review with meta-analysis. *PloS One, 11*(5), e0155431.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives, 4*(1), 10–18.

Winter, S. D., & Depaoli, S. (2020). An illustration of Bayesian approximate measurement invariance with longitudinal data and a small sample size. *International Journal of Behavioral Development, 44*(4), 371–382.

World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates.* Author. [https://apps.who.int/iris/rest/bitstreams/1080542/retrieve](https://apps.who.int/iris/rest/bitstreams/1080542/retrieve)

World Health Organization. (2019). *Suicide in the world: Global health estimates.* Author. [https://apps.who.int/iris/rest/bitstreams/1244794/retrieve](https://apps.who.int/iris/rest/bitstreams/1244794/retrieve)

Zhang, B., Fokkema, M., Cuipers, P., Li, J., Smits, N., & Beekman, A. (2011). Measurement invariance of the Center for Epidemiological Studies Depression Scale (CES-D) among Chinese and Dutch elderly. *BMC Medical Research Methodology, 11*(1), 1–10.