

Article

# Factor Structure and Validation of the 12-Item Korean Version of the General Health Questionnaire in a Sample of Early Childhood Teachers

Boram Lee <sup>1</sup> and Yang-Eun Kim <sup>2,\*</sup>

<sup>1</sup> Department of Early Childhood Education, College of Health and Welfare, Woosong University, Daejeon 34606, Korea; blee@wsu.ac.kr

<sup>2</sup> Department of Global Child Education, College of Health and Welfare, Woosong University, Daejeon 34606, Korea

\* Correspondence: yekim@wsu.ac.kr

**Abstract:** The 12-item General Health Questionnaire (GHQ-12) is designed to detect a diagnosable psychiatric disorder and has demonstrated positive psychometric properties in adult populations. Despite these findings, the psychometric properties of the GHQ-12 have hardly been examined with regard to early childhood teachers. This study purposed to examine the factor structure of the GHQ-12 and to assess its psychometric properties vis-à-vis a sample of Korean early childhood teachers. An aggregate of 252 participants completed the Korean version of the GHQ-12 in tandem with other psychiatric measures, including the Patient Health Questionnaire-9 (PHQ-9) and the Beck Depression Inventory (BDI). The resulting data were subjected to confirmatory factor analyses to compare the goodness-of-fit of the previously proposed models of the GHQ-12. The three-factor model comprising anhedonia/sleep disturbance, social performance and loss of confidence was found by the goodness-of-fit indices to excellently fit our study sample. The average variance extracted and all factor loadings exceeded the recommended threshold of 0.50; hence, convergent validity was established. The criterion posited by Fornell and Larcker verified the discriminant validity. The instrument evidenced superior reliability evinced by its adequate internal consistency and composite reliability. This evidence allows the assertion that the GHQ-12 may be deployed as a screening tool for the evaluation of general symptoms of psychiatric disorders in Korean early childhood teachers.



**Citation:** Lee, B.; Kim, Y.-E. Factor Structure and Validation of the 12-Item Korean Version of the General Health Questionnaire in a Sample of Early Childhood Teachers. *Educ. Sci.* **2021**, *11*, 243. <https://doi.org/10.3390/educsci11050243>

Academic Editor: Eila Jeronen

Received: 5 May 2021

Accepted: 17 May 2021

Published: 19 May 2021

**Keywords:** factor structure; GHQ-12; Korean early childhood teachers; mental health; psychometric properties

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Recently conducted studies have consistently demonstrated the importance of the mental health of early childhood teachers [1,2]. It is critical to attend to the mental health of early childhood teachers for several reasons. Mental health problems such as depression and anxiety have become prominent national concerns in South Korea (hereafter Korea) [3,4]. This paper focuses on teachers in early childhood education settings, a professional group that experiences one of the highest levels of job-related stress. Teachers thus represent a vulnerable cohort that is at high risk of developing mental disorders [5,6]. Several researchers have repeatedly identified the common stressors of early childhood practitioners: work overload, time pressure, difficulties with administration or management and the need to manage behavioral problems in children [6–8]. Studies have also cited the challenges of dealing with parents who treat preschools as child-minding services and the performance of other non-teaching tasks as additional stressors for this group [9,10]. Further, the Korean public does not fully recognize the professional stature of early childhood teachers: the prevailing perception of this group as low in status is combined with

meager remunerations [11]. These elements also contribute to the poor mental health of early childhood teachers. It is widely acknowledged that the mental wellbeing of early childhood teachers significantly influences instructional effectiveness. It also affects the personal growth, emotional development and academic performance of the children in their charge [12,13]. Thus, concerns about the mental health of early childhood teachers assume immense individual and social importance. The increasing scholarly interest in the mental health of early childhood teachers has created a greater demand for valid and reliable research instruments that can appropriately measure the psychological distress of this group. Such evaluations are crucial before apt interventions to promote the mental health of this professional category can be planned, implemented and appraised. Our study tested the validity of the General Health Questionnaire–12 (GHQ-12) [14], a widely used instrument, as a measure of the mental health aspects pertaining specifically to a sample of early childhood educators.

## 2. Literature Review

### 2.1. Application of the GHQ-12

Several extant instruments measure symptoms indicative of psychological distress or psychiatric disorders. The General Health Questionnaire (GHQ) devised by Goldberg is one of the most widely applied assessments of the severity of symptoms associated with psychological distress [14]. The GHQ-12 index was also originally intended to screen for general (non-psychotic) psychiatric morbidity. The original GHQ comprised 60 items, but abridged versions have been developed and modified (e.g., GHQ-30, GHQ-28, GHQ-20 and GHQ-12). The GHQ-12 is one of the most used of such shorter adaptations. The popularity of the GHQ-12 in comparison to longer versions is attributable largely to its ease of use, brevity, self-reporting format and reliability in generating robust results [15,16]. In fact, the GHQ-12 was adopted by a World Health Organization study screening for psychological disorders in primary care because it was considered the most valid among similar vetting tools [17,18].

Empirical evidence suggests that the GHQ-12 evinces adequate internal consistency and superior sensitivity and specificity [19,20]. The GHQ-12 has been widely applied in multiple settings since its development. It has been utilized in both clinical and non-clinical samples, in different cultures and for different age groups [14,18]. To date, the GHQ-12 has been translated into 38 languages, making it accessible to practitioners and researchers across many parts of the world. Some studies have used the GHQ-12 and attended to its applicability in Korea and the reliability and validity of the instrument have been demonstrated [21,22]. However, the research subjects of such studies have mostly been members of the general population or university students. The Korean version of the GHQ-12 has not yet been tested with occupational groups such as the early childhood teachers. This significant gap in the literature must be addressed for prospective research initiatives. The Korean version of the GHQ-12 would facilitate the research process and allow direct comparison of studies focusing on the mental health of teachers in the Korean settings vis-à-vis investigations conducted across different cultures or settings.

### 2.2. Existing Factor Structures of the GHQ-12

A large number of studies have found the GHQ-12 to have favorable psychometric properties among various populations in different countries, including adults in the general population [19,20,22], older adults [23], primary care patients [18], out-patients with psychological disorders [24], pregnant women [25] and adolescents [15].

Despite the encouraging psychometric evaluation of GHQ-12, several studies have applied exploratory and/or confirmatory factor analyses (EFA, CFA) to query whether the GHQ-12 is dimensional or multidimensional and subsequently debated the validity of its underlying structure. Goldberg initially developed the GHQ-12 as a unidimensional construct; however, only a few scholars have supported the one-factor latent structure in subsequent empirical studies [26,27]. Conversely, different factor models emerged when

researchers investigated the dimensionality of the GHQ, suggesting that the instrument is multidimensional and that it contains two or three clinically meaningful factors. Several alternative multidimensional models have been proposed since then, mainly with two or three factors. The three-factor model proposed by Graetz [28] has received the most empirical support in this context and has later been replicated in confirmatory analyses [29–31]. This model comprises the factors of anxiety, social dysfunction and loss of confidence. Notably, the one-factor encompasses all six positively worded items and the six negatively worded items are divided into two separate factors. Simultaneously, other studies have also evidenced that the GHQ-12 comprises three dimensions but have termed the factors differently from Graetz. For example, Worsley and Gribbin's [32] EFA study produced three dimensions (anhedonia/sleep disturbance, social performance and loss of confidence) with several cross-loadings. Martin [33] used CFA and found support for a three-factor solution, labeling the dimensions as self-esteem, stress and successful coping.

Nonetheless, the instrument's two-factor model, which includes six negatively worded and six positively worded items grouped into two factors, has also been sustained by studies based on EFA [34–36]. However, one of the problems with both the two- and three-factor model involves the separation of negatively and positively worded items into separate factors. As such, the question remained whether these factors represented substantive meaning or whether they only denoted artifacts of a response style associated with the positive and negative wording of the items or the so-called method factor. Responding to this challenge, later studies have attempted to model wording effects for the negatively worded items in confirmatory factor models. Hankins' [37] pioneering work conducted on an English sample found that the unidimensional model, with correlated errors on the negatively worded items, was more apt than both the two-factor (positively and negatively worded items) and three-factor models. The studies conducted by Li [38] and Aguado et al. [39] also discovered that the unidimensional model, including its wording effects, was a better fit than Graetz's three-factor model. Nevertheless, no consensus has been achieved about the validity and utility of these multidimensional models, primarily Graetz's. These models have been questioned because of the high degree of correlation between factors.

Some studies in Korea have performed EFA and CFA on the Korean version of the GHQ-12 to examine the psychometric properties of this measure. The following results have been reported: (1) the GHQ-12 demonstrated adequate internal consistency [21,22]; (2) the EFA revealed a two-factor structure [22]; (3) comparisons of single-factor, two-factor and three-factor models using CFA have found that the three-factor model fit the structure of the scale [21]. However, no evidence currently exists to posit that GHQ-12 is suitable for use with early childhood teachers in Korea. Investigations conducted by Park et al. [22] and Lee et al. [21] evinced the good psychometric properties of the Korean version of the GHQ-12; however, at least two limitations currently prevent its use in the context of scholarship. First, Park et al.'s study included the general adult population, whereas Lee et al.'s examination encompassed university students; our teacher sample may differ in important ways from these two distinct samples. Second, the mean age in the university student sample studied by Lee et al. was 20.2 years (age range 18–28 years), a span that is much more limited than the age range represented by the teachers participating in our study. Notably, it would be erroneous to translate the psychometric findings attained from the general adult population and from university students to specific teacher populations that teach and nurture young children. These deficiencies signify that the extant studies do not sufficiently validate the applicability of the GHQ-12 to early childhood teachers. Psychometric properties of measures must be examined in new populations to ensure that they function in manners similar to the original instrument.

### *2.3. Significance of the Study*

Several gaps in the existing scholarship must be addressed when the available research on factor structures of the GHQ-12 is considered. The existing studies assessing the factor structure of the GHQ-12 have yielded inconclusive results. Hence, it is still neces-

sary to verify the factor structure of the GHQ-12. It is true that validated mental health measurements are required to screen and investigate the effects of interventions on early childhood teachers; however, the selection of the most appropriate measure for a specific application depends on several factors. Facets to consider could include study sample characteristics, practical issues such as respondent burden, mode of administration, the need for validated language translation and the psychometric properties of the instrument. Psychometric properties of instruments such as the GHQ-12 can vary among different populations and cultural groups [16]. Hence, a systematic assessment of the instrument's psychometric properties is mandated before the instrument is employed and widely used on a specific population. Further, it is important in practical terms to identify whether the psychometric properties of the Korean version of the GHQ-12 are apt for use with early childhood teachers for whom the identification of efficient measures of mental health is especially important.

As concerns increase about early childhood teachers' job-related stress, the need for brief instruments that efficiently evaluate symptoms of mental health disorders also increases. The GHQ-12, then, might be a particularly useful measure with this population. Distinguishing the psychometric properties of the GHQ-12 could inform health professionals with respect to the appropriate design of prevention programs pertaining to mental disorders that target early childhood teachers who potentially suffer, or are already suffering, from psychological distress. In Korea, no information is available on the GHQ-12's psychometric properties with early childhood teachers. Thus, far, the current study aims to examine the psychometric properties of the GHQ-12 among early childhood teachers through an evaluation of its measurement model validity by CFA and to demonstrate preliminary evidence of convergent and discriminate validity of the GHQ. To date, only one psychometric properties study has assessed the convergent validity of GHQ-12 [40] and evaluated its unique correlations with similar psychiatric instruments such as Patient Health Questionnaire-9 (PHQ-9) [41] and Becks Depression Inventory (BDI) [42]. The current study also used the same measures such as PHQ-9 and BDI to confirm whether the measuring symptoms of psychiatric disorders were complementary rather than duplicative.

### 3. Method

#### 3.1. Participants and Procedures

Ethical approval to conduct this study was obtained from the Institutional Review Board of Woosong University in Korea (Protocol Code: 1041549-201006-SB-103). Our study employed non-random purposive sampling. Thus, our participants comprised early childhood teachers charged with the nurture and instruction of children aged 0–5, who was recruited from daycare centers. Elements of this population were selected arbitrarily and in accordance with certain characteristics; thus, non-random sampling did not allow the estimation of sampling errors. There is no statistical method of assessing the validity of the results obtained from non-random samples.

After receiving ethical approval, the pilot study was conducted with four teachers from two childcare centers who agreed to participate in the study. First, the principal investigator visited the childcare centers in person to explain the purpose of the study, construction of the study scale and method of response. The responses were recorded using a survey tool provided by Google. Questions were added that concerned the time required to complete the survey, whether any questions were difficult to understand or ambiguous in meaning and whether there was any inconvenience in using the system. It was revealed that both understanding and recognizing the questions were not difficult and that the time required to complete the survey was approximately 20 min.

After the pilot study, the main survey was conducted with teachers working at childcare centers. Data were collected using a free survey tool provided by Google in the form of a web-based drive. In more detail, participants were recruited through online postings on the Korean national early childhood teachers' community website, on which only certified early childhood teachers can access. Postings described the study's purpose and directed

those interested to an online-survey link to complete the questionnaire. The participants were informed that participation was voluntary and return of the completed questionnaire was considered as the informed consent. The completed questionnaires were automatically submitted to the researcher. Owing to the possibility of duplicate respondents or the reduction of survey response rate, which is possible when the survey period is too long or too short, the survey period was set to 10 days from April 1, 2020 to April 10, 2020. A message that requested respondents to complete the survey sincerely and emphasized the advantages of anonymity and flexibility of response time was also sent. In total, 230 copies of the questionnaire were collected, of which 225 were utilized for the final analysis; five with unreliable responses were excluded.

This study's participants were 252 early childhood teachers (243 females; 9 males) who taught children aged zero to five at childcare centers. At the time of the survey, participants ranged from 21 to 59 years old ( $M = 33.5$ ,  $SD = 10.3$ ). On average, they had 13.08 years of teaching experience ( $SD = 2.34$ ; range 1 month–35 years). Participants were employed at various types of childcare centers: on-site (49.6%), private (21.0%), public (11.9%), corporate (6.7%), domestic childcare centers (5.6%) and others (5.2%).

### 3.2. Measures

The General Health Questionnaire (GHQ-12) [14] is a self-report measure for detecting psychiatric disorders in the general population within community and non-psychiatric clinical settings. The questionnaire contains 12 items, each scored on a four-point Likert scale from 0 to 3. Thus, the total score ranges from 0 to 36, with higher scores indicating worse conditions. The Korean version, translated and validated by Park et al. [22] was used in this study. Cronbach's  $\alpha$  coefficient was 0.86 for the overall GHQ-12 in this study.

The Patient Health Questionnaire (PHQ-9) [41] was used as a measure of depressive symptomatology. The PHQ-9 consists of nine items scored on a four-point Likert scale from 0 to 3, resulting in a total score from 0 to 27, with a higher score reflecting more severe symptoms of depression. The Korean version of the PHQ-9 has been validated and demonstrated to exhibit excellent psychometric properties in Korean adults [43]. In this study, the PHQ's Korean version demonstrated a Cronbach's  $\alpha$  coefficient of 0.81.

The Beck Depression Inventory (BDI) [42] was used to assess depressive symptomatology's presence and severity based on the past 2 weeks. It comprises 21 items scored on a four-point Likert scale ranging from 0 to 3. Items are summed to provide a total score ranging from 0 to 63, with higher scores indicating more severe depressive symptoms. This study used the BDI's Korean version, verified and validated by Lee and Song [44]. In the current study, its Cronbach's  $\alpha$  coefficient was 0.80.

### 3.3. Statistical Analyses

Collected data were analyzed using IBM SPSS Statistics for Windows (version 23) and AMOS (version 20) (IBM Corp., Armonk, NY, USA). CFA was conducted through structural equation modeling, using robust maximum likelihood estimation to assess varied latent structure models of the GHQ-12 because Mardia's test indicated that our data violated the multivariate normality assumption (Mardia's kurtosis = 104.70,  $p < 0.001$ ). Models examined were based on results from previous research on the GHQ-12's factor structures, specifically, five competing models. Model 1 is the original one-factor structure hypothesized by Goldberg [14], with all 12 items loaded onto a single factor. Proposed by Andrich and Van Schaubroeck [34], Model 2 is a correlated two-factor structure with six negatively worded items loaded onto one factor and six positively worded items loaded onto another. Model 3 is a unidimensional model with a method factor specifically for the negative items suggested by Hankins [37]. Suggested by Graetz [28], Model 4 is a correlated three-factor model consisting of anxiety and depression (4 items), anhedonia and social dysfunction (6 items) and loss of confidence (2 items). Postulated by Martin [33], Model 5 is also a correlated three-factor model in which three latent variables are represented by cope (4 items), stress (3 items) and depression (5 items). Finally, Model 6 was reported by

Worsley and Gribbin [32] who also proposed three factors: anhedonia and sleep disturbance (2 items), social performance (6 items) and loss of confidence (4 items).

To evaluate model fit, incremental fit indices, including the chi square ( $\chi^2$ ) and its subsequent ratio with the number of degrees of freedom ( $\chi^2/\text{df}$ ); comparative fit index (CFI); goodness-of-fit index (GFI); root mean square error of approximation (RMSEA) and its 90% confidence interval (90% CI); and standardized root mean square residual (SRMR) were used. Acceptable data fit to model is indicated by  $\chi^2/\text{df} < 3$  [45], CFI  $> 0.90$  [46], GFI  $> 0.90$  [47], RMSEA  $< 0.08$  [48] and SRMR  $< 0.08$  [47]. Akaike's information criterion (AIC) was used to compare alternative plausible models, with lower values signifying a better model fit.

To determine whether models differed significantly, chi-square difference tests were used. To evaluate convergent validity, Pearson's  $r$  was used to test the associations between the GHQ-12 and criteria instruments (i.e., PHQ-9 and BDI). Convergent validity was also assessed through an assessment of item factor loadings and their statistical significance, followed by an assessment of factor-related average variance extracted (AVE). Convergent validity was indicated by an item factor loading and AVE equal to or greater than 0.50 [49]. Discriminant validity was assessed by adhering to the procedures suggested by Fornell and Larcker [50]. Discriminant validity is assured if the square root of the AVE of each construct is greater than its correlations with any other composite construct in the assessed model. The internal consistency was computed using Cronbach's  $\alpha$  and composite reliability (CR) scores for each of the suggested factors of the model. Cronbach's  $\alpha$  value above 0.70 and above are generally considered acceptable [51]. CR values between 0.60 to 0.70 are deemed satisfactory; however, the value must be higher than 0.70 at more advanced stages [50].

## 4. Results

### 4.1. Descriptive Statistics

Table 1 displays the GHQ-12's overall and individual item scores. A mean score of 21.05 (SD = 5.03)—higher than the cutoff point of 12—was obtained. Items with the highest mean scores—more than 2.30—were 1 and 5. Item 5 was notable for the highest score, indicating that the majority of respondents felt they were under strain. Moreover, separate mean scores for males and females were not calculated because the sample included only nine males.

**Table 1.** Descriptive statistics for GHQ-12 items.

GHQ-12 Items	Mean	SD
1. Able to concentrate	2.31	0.87
2. Loss of sleep over worry	1.83	0.72
3. Playing a useful part	1.58	1.21
4. Capable of making decisions	1.52	1.01
5. Felt constantly under strain	2.39	1.64
6. Couldn't overcome difficulties	1.60	0.77
7. Able to enjoy day-to-day activities	2.13	1.83
8. Able to face problems	1.46	0.73
9. Feeling unhappy and depressed	1.56	0.67
10. Losing confidence	1.71	0.66
11. Thinking of self as worthless	1.33	0.66
12. Feeling reasonably happy	1.63	1.12
Mean overall score	21.05	5.03

### 4.2. Confirmatory Factor Analysis

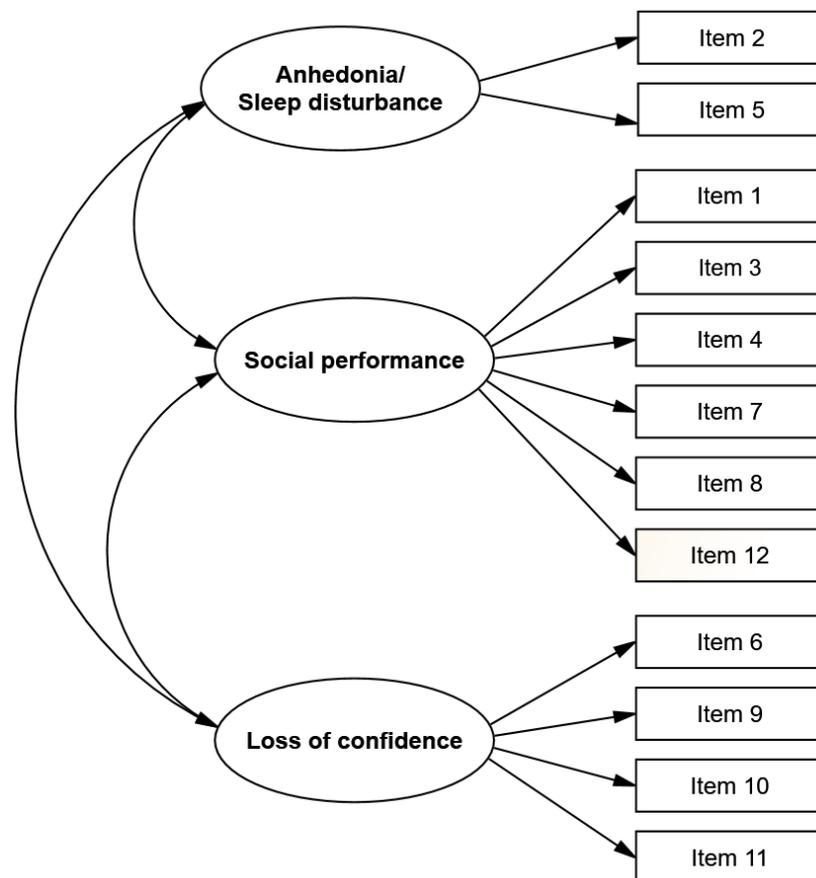
Table 2 shows competing models' goodness-of-fit indices. Across the whole sample, the overall fit indices of the six-factor models were examined across the entire sample using a variety of fit indices. The results revealed that all two- and three-factor models except for Martin's single-factor and three-factor models were acceptably apt. However, the evaluation accomplished using the stated model fit indices disclosed that Worsley

and Gribbin's three-factor model achieved the best fit, demonstrating highly satisfactory suitability across all model fit indices (Figure 1).

**Table 2.** Goodness-of-fit indices for GHQ-12 models in CFA.

Models	<i>k</i>	$\chi^2$	<i>df</i>	$\chi^2/df$	CFI	GFI	RMSEA (90% CI)	SRMR	AIC
Model 1	12	130.32 *	54	2.41	0.83	0.95	0.075 (0.059–0.092)	0.064	178.32
Model 2	12	97.05 *	53	1.83	0.90	0.97	0.058 (0.039–0.075)	0.054	147.05
Model 3	12	93.63	48	1.95	0.90	0.97	0.062 (0.043–0.080)	0.053	153.64
Model 4	12	95.10 *	51	1.87	0.90	0.97	0.059 (0.040–0.077)	0.054	149.10
Model 5	12	116.83 *	51	2.29	0.86	0.96	0.072 (0.055–0.089)	0.062	170.83
Model 6 †	12	54.0 *	32	1.69	0.93	0.99	0.052 (0.026–0.076)	0.048	100.04

Notes. \*  $p < 0.01$ ; *k* = number of items; *df* = degrees of freedom; CFI = comparative fit index; GFI = goodness of fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean residual; AIC = Akaike's information criterion; † represents a final model used in the study.



**Figure 1.** Three-factor model of GHQ-12 [ $\chi^2 = 54.0$ ,  $df = 32$ ;  $\chi^2/df = 1.69$ ; CFI = 0.93; GFI = 0.99; RMSEA = 0.052 (90% CI = 0.026–0.076)].

The AIC statistics further confirm the superior fit of Worsley and Gribbin's three-factor model (Model 6), as the AIC is 100.04, which is lower than the rest of the models tested in the study. Moreover,  $\chi^2$  difference tests revealed that Model 6 had significantly better fit to data than Model 1 ( $\chi^2(22) = 76.32$ ,  $p < 0.001$ ), Model 2 ( $\chi^2(21) = 43.05$ ,  $p < 0.001$ ), Model 3 ( $\chi^2(16) = 39.6$ ,  $p < 0.001$ ), Model 4 ( $\chi^2(19) = 41.10$ ,  $p < 0.001$ ) and Model 5 ( $\chi^2(19) = 62.83$ ,  $p < 0.001$ ). In addition, Model 6's three factors were weakly and moderately correlated: 0.30 between first and third factors, 0.43 between second and third factors and 0.22 between first and second factors.

Overall, results demonstrated that all two- and three-factor models, as well as a model with a method factor excepting those of Goldberg and Martin, have acceptable fit. Nevertheless, evaluation of indices revealed Worsley and Gribbin's three-factor model as the best because it demonstrated highly acceptable fit across all indices.

#### 4.3. Convergent Validity

The relationship between the three GHQ-12 subscales and total GHQ-12, PHQ-9 and BDI scores was obtained through Pearson's correlations. Table 3 displays moderate correlations among total scores of the GHQ-12, PHQ-9 and BDI. GHQ-12 subscales and total scores of the GHQ-12, PHQ-9 and BDI were also moderately correlated. Lastly, the GHQ-12 total score correlated moderately with subscales of anhedonia and sleep disturbance, social performance and loss of confidence. All correlation  $p$  values were less than 0.001. Further, convergent validity was satisfactory, with all factor loadings exceeding 0.50. The factor loading of all items was significant, given the range of 0.56–0.82. The AVE of all constructs also surpassed 0.50, indicating sufficient convergent validity (Table 4).

**Table 3.** Correlations between the GHQ-12, its subscales PHQ-9 and BDI.

	1	2	3	4	5	6
1. Total GHQ-12	-					
2. GHQ-12 Anhedonia/Sleep disturbance	0.51 *	-				
3. GHQ-12 Social performance	0.53 *	0.13	-			
4. GHQ-12 Loss of confidence	0.44 *	0.47 *	0.27 *	-		
5. PHQ-9	0.48 *	0.44 *	0.30 *	0.55 *	-	
6. BDI	0.56 *	0.45 *	0.37 *	0.66 *	0.75 *	-

Notes. \*  $p < 0.01$ ; GHQ-12: General Health Questionnaire-12. PHQ-9: Patient Health Questionnaire. BDI: Beck Depression Inventory.

**Table 4.** Convergent validity of the Korean version of the GHQ-12.

Construct	Items	$\beta^a$	$B^b$	SE <sup>c</sup>	Cronbach's $\alpha$	CR <sup>d</sup>	AVE <sup>e</sup>
Anhedonia/sleep disturbance	2	0.57	1.00	0.15	0.42	0.66	0.50
	5	0.82	1.13	0.01			
Social performance	1	0.56	0.69	0.06	0.85	0.92	0.51
	3	0.67	1.21	0.06			
	4	0.71	1.01	0.06			
	7	0.80	1.66	0.18			
	8	0.74	1.29	0.13			
Loss of confidence	12	0.78	0.72	0.01	0.80	0.88	0.50
	6	0.77	1.03	0.19			
	9	0.64	1.02	0.05			
	10	0.59	1.00	0.12			
	11	0.79	1.08	0.24			

Notes. <sup>a</sup> Standardized coefficient. <sup>b</sup> Unstandardized coefficient. <sup>c</sup> SE = Standard error. <sup>d</sup> CR = Composite Reliability. <sup>e</sup> AVE = Average Variance Extracted.

#### 4.4. Discriminant Validity

Table 5 exhibits the square roots of AVE indexes for all three subscales. Our results confirm that discriminant validity was achieved as all indexes (diagonal values in bold) were higher than the inner diagonal values representing the correlations among constructs. Hence, the results support the discriminatory validity of the instrument.

**Table 5.** Fornell–Larcker criterion.

Construct	Anhedonia/Sleep Disturbance	Social Performance	Loss of Confidence
Anhedonia/sleep disturbance	0.67		
Social performance	0.54	0.71	
Loss of confidence	0.43	0.55	0.73

#### 4.5. Internal Consistency

Cronbach's  $\alpha$  ranged between 0.42 (for anhedonia/sleep disturbance) and 0.85 (for social performance). The CR values exceeded the recommended computation of 0.60 for all three subscales (Table 4).

## 5. Discussion

This study examined the psychometric properties of the Korean version of the GHQ-12. To the authors' best knowledge, this is the first study in Korea that has attempted to examine the GHQ-12's factor structure using CFA and its psychometric properties with a sample of early childhood teachers, an occupational group particularly vulnerable to mental health problems. Thus, the current research contributes to past research by examining the structure of the GHQ in another vulnerable occupational group.

Our CFA findings suggested that all models exhibited an RMSEA of less than 0.08. However, our overall results revealed that Andrich and Van Schaubroeck's [34] two-factor model, Hankins' [37] model including an artifactual factor that encompassed all the negative items, Graetz's [28] three-factor model and another three-factor model proposed by Worsley and Gribbin [32] were the only models to evidence a good fit. These three models also revealed acceptable model fit across other indices. Although the GHQ-12 was originally developed as a unidimensional structure, numerous other two- and three-factor structures have been identified and, thus, researchers have reached no consensus regarding its dimensionality or factor structure [15]. In this study, Worsley and Gribbin's [32] three-factor model, which was initially described in a cross-sectional community sample of Australian adults, provided the best fit to data with three factors of anhedonia and sleep disturbance, social performance and loss of confidence. Aloba et al. [15] confirmed this finding with Nigerian adolescents. We also found low to moderate correlations among the three factors, reflecting a low amount of covariance, thus further supporting this three-factor model as best explaining psychological distress in our sample.

As for convergent validity, the Korean version of the GHQ-12 total score showed moderate correlations with the three subscales of anhedonia and sleep disturbance, social performance and loss of confidence, consistent with results from current CFA and implying that its total score can measure general distress in this population. Notably, moderate positive correlations of the three subscales and the PHQ-9 and BDI corroborate these subscales' associations with mental health problems. Although the correlational strength of the total GHQ-12 with its three subscales and the other similar measures were moderate, the directions were all as expected. Therefore, the convergent validity between the GHQ-12, PHQ-9 and BDI was moderate, confirming that GHQ-12 was designed to measure symptoms assessing mental distress and minor psychiatric morbidities. Similar findings have been reported by Martin et al., [40] who also found moderate to strong associations between the total GHQ-12, PHQ-9 and BDI in community-based samples in Germany. The convergent validity of the GHQ-12 was also indicated by adequate factor loadings and acceptable AVE values. For discriminant validity, AVE values for the subscales were higher in relation to  $r^2$  values. Hence, the discriminant validity for each subscale was confirmed.

Cronbach's  $\alpha$  coefficient values for this study indicated adequate total internal consistency for the GHQ-12 and for two of the subscales. Other studies involving non-clinical and clinical adult samples in Germany, China, India and Iran have similarly reported Cronbach's  $\alpha$  coefficients ranging from good to excellent [18–20,23]. In sum, these findings suggest that the GHQ-12 demonstrates satisfactory internal consistency across varied

populations and languages. However, only the anhedonia/sleep disturbance subscale showed an internal consistency lower than the recommended value. This finding is aligned with Aloba et al. [15] who also found and replicated the three-factor model developed by Gribben and Worsely. In Aloba et al.'s study, Cronbach's alpha values ranged from 0.60 to 0.69, a level that is deemed unsatisfactory. Another study that discovered a three-factor model, however, did not calculate the Cronbach's alpha of the different subscales because factors 2 and 3 each comprised only three items [29]. The low internal consistency of the neutralizing subscale probably resulted from the fact that the anhedonia and sleep disturbance factor is only comprised of two items. This possibility was tested by computing composite reliability indexes, an action that has been recommended for the generation of better estimates of true reliability in testing subscales than is possible through the coefficient alpha [52]. The estimates of true reliability obtained in our study through CR were, on average, better (larger) than corresponding coefficient alpha values for all the subscales.

Our study has an important implication for assessment and diagnosis of mental health problems among early childhood teachers in Korea. Effective preventive and promotional measures are essential to minimize mental disorders' impact on the individual. Therefore, a valid instrument such as the GHQ-12 can enable clinicians to identify those at increased risk of mental health issues, with early intervention appropriately planned, implemented and evaluated.

Limitations of this study need to be acknowledged. First, the sample size was somewhat small for CFA and that female teachers outnumbered male teachers. Gender imbalance in the early childhood workforce is a longstanding global phenomenon [53]. Extensive research conducted in this domain has discovered male early childhood educators represented only 1–3% of the aggregate of early childhood practitioners in most Western and non-Western countries [53,54]. Korea is no different, with males only denoting 1% of the early childhood teachers [55]. The reason for this gender imbalance may be attributed to the fact that early childcare and education have historically been seen as women's work. The widespread cultural belief that women are more nurturing and caring than men hinders men who may want to make a career in early childhood education [56]. Future evaluations of the GHQ-12 would benefit from larger samples with more male respondents even though women are considerably overrepresented in the teaching profession, especially in early childhood education settings.

Next, early childhood teachers belong to a highly stressed occupational group. Indeed, our participants scored above the cutoff threshold, indicating poor mental health, so these results might not be applicable to the general population or other occupational groups. Importantly, our findings regarding the CFA needs to be interpreted with caution. It should be noted that the alpha was too low for the "anhedonia and sleep" factor as it comprised of only two factors. The elimination of specific factors with low factor loadings or low alpha is a controversial issue because it implies the contemplation of both the positive and negative aspects of reducing the number of items of an established questionnaire [21]. In fact, the removal of items or factors could guarantee that a measure would become more robust and reliable. However, such exclusions could also mean that the newly validated scale cannot be compared to other previously published and currently used versions. Notably, the original 12-item GHQ is the most widely used across different studies and samples despite the potential weaknesses that could arise from the retention of all items of the scale. In addition, it seems apt and relevant for comparative purposes to sustain the same 12-item version. Moreover, the deletion of a specific factor may not represent the correct solution in the case at hand because the three-factor model provided the best fit to the data in reproducing the observed data, as all standardized factor loadings were significant at 0.50 and CR values were excellent for all GHQ-12 scale scores. We believe that there is no appropriate indication of item or factor removal.

## 6. Conclusions

The current study addressed a gap in the literature by employing CFA to examine the factorial structure of the Korean version of GHQ-12 with respect to early childhood teachers. The findings of this study confirm that the GHQ-12 is best conceived as a multi-dimensional tool that can assess several distinct aspects of distress rather than a unidimensional or a unitary screening measure. The results of CFA revealed that Worsley and Gribbin's three-factor model offered the best fit to the data. Further, the outcomes of the study confirm the reliability and validity of the Korean version of the GHQ-12 as a tool that can effectively be employed for the assessment of general symptoms of psychological distress in early childhood teachers. It is also worth mentioning the strengths of this study. The current study employed both classical (e.g., based on Cronbach's alpha) and modern (e.g., based on structural equation modeling and CR, an alternative preferred to Cronbach's alpha as a test of convergent validity in a reflective model [57]) methods to evaluate the psychometric properties of the Korean version of the GHQ-12. The convergent and discriminant validity of the GHQ-12 signifies that this scale accurately measured perceived psychological distress in the sample of early childhood teachers. Further, to the best of our knowledge, this study is the first to demonstrate that a three-factor model provided a conceptually acceptable fit to the data in a sample of early childhood teachers in Korea.

**Author Contributions:** Conceptualization, B.L.; methodology, B.L.; software, B.L.; validation, B.L.; formal analysis, B.L.; investigation, Y.-E.K.; resources, Y.-E.K.; data curation, Y.-E.K.; writing—original draft preparation, B.L.; writing—review and editing, B.L.; visualization, Y.-E.K.; supervision, B.L.; project administration, Y.-E.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Woosong University [Grant No. WLB 1928-2332-01].

**Institutional Review Board Statement:** The study was approved by the Ethics Committee of Woosong University (protocol code 1041549-201006-SB-103).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jeon, L.; Buettner, C.K.; Grant, A.A. Early childhood teachers' psychological well-being: Exploring potential predictors of depression, stress, and emotional exhaustion. *Educ. Dev.* **2017**, *29*, 1–17. [[CrossRef](#)]
2. Jeon, H.J.; Kwon, K.A.; Walsh, B.; Burnhama, M.M.; Choi, Y.J. Relations of early childhood education teachers' depressive symptoms, job-related stress, and professional Motivation to beliefs about children and teaching practices. *Educ. Dev.* **2019**, *30*, 131–144. [[CrossRef](#)]
3. Cho, J.J.; Kim, J.Y.; Chang, S.J.; Fiedler, N.; Koh, S.B.; Crabtree, F.B.; Kang, D.M.; Kim, Y.K.; Choi, Y.H. Occupational stress and depression in Korean employees. *Int. Arch. Occup. Environ. Health* **2008**, *82*, 47–57. [[CrossRef](#)] [[PubMed](#)]
4. Shin, S.; Hwang, E. Factors influencing depressive symptoms among Korean older adults with chronic illnesses: Using the 2014 National Survey on older adults. *Korean J. Adult. Nurs.* **2018**, *30*, 577–585. [[CrossRef](#)]
5. Kim, J. The effects of working conditions and job satisfaction on the mental health and presenteeism of early childhood teachers. *Korean J. Occup. Health. Nurs.* **2018**, *27*, 171–179.
6. Koo, E.M. A study on child care teacher's physical health, mental health and burnout. *J. Ear. Child. Educ. Educ. Admin.* **2011**, *15*, 119–139.
7. Bernard, M.E. Teacher beliefs and stress. *J. Ration Emot. Cogn. Behav. Ther.* **2016**, *34*, 209–224. [[CrossRef](#)]
8. Fernet, C.; Guay, F.; Senécal, C.; Austin, S. Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teach. Teach. Educ.* **2012**, *28*, 514–525. [[CrossRef](#)]
9. De Stasio, S.; Fiorilli, C.; Benevene, P.; Uusitalo-Malmivaara, L.; Chiacchio, C.D. Burnout in special needs teachers at kindergarten and primary school: Investigating the role of personal resources and work wellbeing. *Psychol. Sch.* **2017**, *54*, 472–486. [[CrossRef](#)]
10. Seo, Y.J.; Lee, D.K. Stress and coping skill of teachers in child-care institutions. *J. Future Ear. Child. Educ.* **2011**, *18*, 259–291.

11. Ahn, N.R.; Kim, H.S.; Ahn, S.H. The relationship between job stress and turnover intention of child care teachers and the moderating role of motivation for child care work. *J. Korean Home Manag. Assoc.* **2015**, *33*, 87–102. [[CrossRef](#)]
12. Maughan, A.; Cicchetti, D.; Toth, S.L.; Rogosch, A.F. Early-occurring maternal depression and maternal negativity in predicting young children's emotion regulation and socioemotional difficulties. *J. Abnorm. Child. Psychol.* **2007**, *35*, 685–703. [[CrossRef](#)] [[PubMed](#)]
13. Thakur, S.S. To study the relationship between burnout and effectiveness of primary school teachers. *Int. Index. Ref. J.* **2012**, *3*, 15–16.
14. Goldberg, D.P. *The Detection of Psychiatric Illness by Questionnaire; A Technique for the Identification and Assessment of Non-Psychotic Psychiatry Illness*; Oxford University Press: Oxford, UK, 1972.
15. Aloba, O.; Opakunle, T.; Ogunrinu, K. Alternative models examination and gender measurement invariance of the 12-item general health questionnaire among Nigerian adolescents. *Psychiatry Investig.* **2019**, *16*, 808–815. [[CrossRef](#)]
16. Hystad, S.W.; Johnsen, B.H. The dimensionality of the 12-item general health questionnaire (GHQ-12): Comparisons of factor structures and invariance across samples and time. *Front. Psychol.* **2020**, *11*, 1300. [[CrossRef](#)] [[PubMed](#)]
17. Goldberg, D.P.; Gater, R.; Sartorius, N.; Ustun, T.B.; Piccinelli, M.; Gureje, O.; Rutter, C. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol. Med.* **1997**, *27*, 191–197. [[CrossRef](#)] [[PubMed](#)]
18. Schmitz, N.; Kruse, J.; Tress, W. Psychometric properties of the general health questionnaire (GHQ-12) in a German primary care sample. *Acta Psychiatry Scand.* **1999**, *100*, 462–468. [[CrossRef](#)] [[PubMed](#)]
19. Guan, M.; Han, B. Factor structures of general health questionnaire-12 within the number of kins among the rural residents in China. *Front. Psychol.* **2019**, *10*, 1774. [[CrossRef](#)]
20. Endsley, P.; Weobong, B.; Nadkarni, A. The psychometric properties of GHQ for detecting common mental disorder among community dwelling men in Goa, India. *Asian J. Psychiatry* **2017**, *28*, 106–110. [[CrossRef](#)]
21. Lee, B.; Kim, Y.E. Factor structure of the 12-item general health questionnaire (GHQ-12) among Korean university students. *Psychiatry Clin. Psychopharmacol.* **2020**, *30*, 248–253. [[CrossRef](#)]
22. Park, J.I.; Kim, Y.J.; Cho, M.J. Factor structure of the 12-item general health questionnaire in the Korean general adult population. *J. Korean Neuropsychiatry Assoc.* **2012**, *51*, 178–184. [[CrossRef](#)]
23. Namjoo, S.; Shaghaghi, A.; Sarbaksh, P.; Allahverdipour, H.; Pakpour, A.H. Psychometric properties of the general health questionnaire (GHQ-12) to be applied for the Iranian elder population. *Aging Ment. Health* **2017**, *21*, 1047–1051. [[CrossRef](#)] [[PubMed](#)]
24. Gao, F.; Luo, N.; Thumboo, J.; Fones, C.; Li, S.C.; Cheung, Y.B. Does the 12-item general health questionnaire contain multiple factors and do we need them? *Health Qual. Life Outcomes* **2004**, *2*, 63. [[CrossRef](#)] [[PubMed](#)]
25. Ip, W.Y.; Martin, C.R. Psychometric properties of the 12-item general health questionnaire (GHQ-12) in Chinese women during pregnancy and in the postnatal period. *Psychol. Health Med.* **2006**, *11*, 60–69. [[CrossRef](#)] [[PubMed](#)]
26. Banks, M.H.; Clegg, C.W.; Jackson, P.R.; Kemp, N.J.; Stafford, E.M.; Wall, T.D. The use of the general health questionnaire as an indicator of mental health in occupational studies. *J. Occup. Psychol.* **1980**, *53*, 187–194. [[CrossRef](#)]
27. Winefield, H.R.; Goldney, R.D.; Winefield, A.H.; Tiggemann, M. The general health questionnaire: Reliability and validity for Australian youth. *Aust. N. Z. J. Psychiatry* **1989**, *23*, 53–58. [[CrossRef](#)]
28. Graetz, B. Multidimensional properties of the general health questionnaire. *Soc. Psychiatry Psychiatry Epidemiol.* **1991**, *26*, 132–138. [[CrossRef](#)]
29. Pardon, A.; Gelan, I.; Durban, M.; Gandarillas, A.; Rodriguez-Artalejo, F. Confirmatory factor analysis of the general health questionnaire (GHQ-12) in Spanish adolescents. *Qual. Life Res.* **2012**, *21*, 1291–1298.
30. Gelaye, B.; Tadesse, M.G.; Lohsoonthorn, V.; Lertmeharit, S.; Pensuksan, W.C.; Sanchez, S.E.; Lemma, S.; Berhane, Y.; Vélez, J.C.; Barbosa, C.; et al. Psychometric properties and factor structure of the general health questionnaire as a screening tool for anxiety and depressive symptoms in a multi-national study of young adults. *J. Affect. Disord.* **2015**, *187*, 197–202. [[CrossRef](#)] [[PubMed](#)]
31. Guan, M. Measuring the effects of socioeconomic factors on mental health among migrants in urban China: A multiple indicators multiple causes model. *Intern. J. Ment. Health Syst.* **2017**, *11*, 10. [[CrossRef](#)]
32. Worsley, A.; Gribbin, C.C. A factor analytic study of the twelve item general health questionnaire. *Aust. N. Z. J. Psychiatry* **1977**, *11*, 269–272. [[CrossRef](#)] [[PubMed](#)]
33. Martin, A.J. Assessing the multidimensionality of the 12-item general health questionnaire. *Psychol. Rep.* **1999**, *84*, 927–935. [[CrossRef](#)] [[PubMed](#)]
34. Andrich, D.; van Schoubroeck, L. The general health questionnaire: A psychometric analysis using latent trait theory. *Psychol. Med.* **1989**, *19*, 469–485. [[CrossRef](#)] [[PubMed](#)]
35. Gao, W.; Stark, D.; Bennett, M.I.; Siegert, R.J.; Murray, S.; Higginson, I.J. Using the 12-item general health questionnaire to screen psychological distress from survivorship to end-of-life care: Dimensionality and item quality. *Psychol. Oncol.* **2012**, *21*, 954–961. [[CrossRef](#)] [[PubMed](#)]
36. Politi, P.L.; Piccinelli, M.; Wilkinson, G. Reliability, validity and factor structure of the 12-item general health questionnaire among young males in Italy. *Acta Psychiatry Scand.* **1994**, *90*, 432–437. [[CrossRef](#)]
37. Hankins, M. The factor structure of the twelve item general health questionnaire (GHQ-12): The result of negative phrasing? *Clin. Pract. Epidemiol. Ment. Health* **2008**, *4*, 10. [[CrossRef](#)]

38. Li, W.H.; Chung, J.O.; Chui, M.M.; Chan, P.S. Factorial structure of the Chinese version of the 12-item general health questionnaire in adolescents. *J. Clin. Nurs.* **2009**, *18*, 3253–3261. [[CrossRef](#)] [[PubMed](#)]
39. Aguado, J.; Campbell, A.; Ascaso, C.; Navarro, P.; Garcia-Esteve, L.; Luciano, J.V. Examining the factor structure and discriminant validity of the 12-item general health questionnaire (ghq-12) among Spanish postpartum women. *Assessment* **2012**, *19*, 517–525. [[CrossRef](#)]
40. Martin, A.; Rief, W.; Klaiberg, A.; Braehler, E. Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. *Gen. Hosp. Psychiatry* **2006**, *28*, 71–77. [[CrossRef](#)]
41. Kroenke, K.; Spitzer, R.L.; Williams, J.B. The PHQ-9: Validity of a brief depression severity measure. *J. Gen. Intern. Med.* **2001**, *16*, 606–613. [[CrossRef](#)]
42. Beck, A.T.; Steer, R.A.; Garbin, M.G. Psychometric properties of the beck depression inventory: Twenty-five years of evaluation. *Clin. Psychol. Rev.* **1988**, *8*, 77–100. [[CrossRef](#)]
43. Park, S.J.; Choi, H.R.; Choi, J.H.; Kim, K.W.; Hong, J.P. Reliability and validity of the Korean version of the patient health questionnaire-9 (PHQ-9). *Anxiety Mood.* **2010**, *6*, 119–124.
44. Lee, Y.H.; Song, J.Y. A study of the reliability and the validity of the BDI, SDS, and MMPI-D scales. *J. Korean. Neuropsychiatr. Assoc.* **1991**, *10*, 98–113.
45. Kline, R.B. *Principles and Practice of Structural Equation Modeling*, 3rd ed.; Guilford Press: New York, NY, USA, 2011.
46. Marsh, H.W.; Hau, K.T.; Wen, Z. In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999) findings. *Struct. Equ. Model.* **2004**, *11*, 320–341.
47. Fan, X.; Sivo, S.A. Sensitivity of fit indices to model misspecification and model types. *Multivar. Behav. Res.* **2007**, *42*, 509–529. [[CrossRef](#)]
48. Browne, M.W.; Cudeck, R. Alternative ways of assessing model fit. *Soc. Methods Res.* **1992**, *21*, 230–258. [[CrossRef](#)]
49. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariate Data Analysis*, 7th ed.; Prentice Hall: Upper Saddle Rive, NJ, USA, 2010.
50. Fornell, C.; Larcker, D.F. Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* **1981**, *18*, 39–50. [[CrossRef](#)]
51. Nunnally, J.C. *Psychometric Theory*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1978.
52. Raykov, T. Coefficient alpha and composite reliability with interrelated nonhomogenous items. *Appl. Psychol. Meas.* **1998**, *22*, 375–385. [[CrossRef](#)]
53. Peeters, J.; Rohrmann, T.; Emilsen, K. Gender balance in ECEC: Why is there so little progress? *Eur. Early Child. Educ. Res. J.* **2015**, *23*, 302–314. [[CrossRef](#)]
54. Xu, Y.; Waniganayake, M. An exploratory study of gender and male teachers in early childhood education and care centres in China. *Compare* **2018**, *48*, 518–534.
55. Ministry of Health and Welfare. Early Childhood Education and Care Statistical Data. 2019. Available online: <https://www.mohw.go.kr/react/jb/sjb030301vw.jsp> (accessed on 29 April 2020).
56. Joseph, S.; Wright, Z. Men as early childhood educators: Experience and perspectives of two male prospective teachers. *J. Educ. Hum. Dev.* **2016**, *5*, 213–221. [[CrossRef](#)]
57. Peterson, R.A.; Kim, Y. On the relationship between coefficient alpha and composite reliability. *J. Appl. Psychol.* **2013**, *98*, 194–198. [[CrossRef](#)] [[PubMed](#)]