Sleep Medicine 16 (2015) 598–603

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Original Article

Validating the cross-cultural factor structure and invariance property of the Insomnia Severity Index: evidence based on ordinal EFA and CFA

Po-Yi Chen [a,b], Chien-Ming Yang [c,*], Charles M. Morin [d,e]

[a] Department of Psychology, University of Kansas, 1415 Jayhawk Blvd, Fraser Hall, Room 426, Lawrence, KS 66045-7556, USA
[b] Center of Research Method and Data Analysis, University of Kansas, 1415 Jayhawk Blvd, Waston Library 470, Lawrence, KS, USA
[c] Department of Psychology/The Research Center for Mind, Brain, & Learning, National Chengchi University, No. 64, Sec 2, ZhiNan Rd., Wenshan District, Taipei, Taiwan
[d] École de Psychologie, Université Laval, Québec, QC, Canada
[e] Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec, Québec, QC, Canada

ARTICLE INFO

Article history:
Received 25 July 2014
Received in revised form 8 November 2014
Accepted 13 November 2014
Available online 12 February 2015

Keywords:
Insomnia Severity Index
Measurement invariance
Factorial structure
Ordinal exploratory factor analysis
Multigroup structural equation model

ABSTRACT

Objective: The purpose of this study is to examine the factor structure of the Insomnia Severity Index (ISI) across samples recruited from different countries. We tried to identify the most appropriate factor model for the ISI and further examined the measurement invariance property of the ISI across samples from different countries.

Methods: Our analyses included one data set collected from a Taiwanese sample and two data sets obtained from samples in Hong Kong and Canada. The data set collected in Taiwan was analyzed with ordinal exploratory factor analysis (EFA) to obtain the appropriate factor model for the ISI. After that, we conducted a series of confirmatory factor analyses (CFAs), which is a special case of the structural equation model (SEM) that concerns the parameters in the measurement model, to the statistics collected in Canada and Hong Kong. The purposes of these CFA were to cross-validate the result obtained from EFA and further examine the cross-cultural measurement invariance of the ISI.

Results: The three-factor model outperforms other models in terms of global fit indices in Taiwan's population. Its external validity is also supported by confirmatory factor analyses. Furthermore, the measurement invariance analyses show that the strong invariance property between the samples from different cultures holds, providing evidence that the ISI results obtained in different cultures are comparable.

Conclusions: The factorial validity of the ISI is stable in different populations. More importantly, its invariance property across cultures suggests that the ISI is a valid measure of the insomnia severity construct across countries.

1. Introduction

The Insomnia Severity Index (ISI) is one of the most well-known instruments for assessing insomnia [1]. It contains seven items, all scored on a five-point scale, to measure the symptoms of insomnia. Previous studies have repeatedly demonstrated the efficiency and validity of the ISI as a clinical evaluation instrument [2,3]. Given this support, the ISI has been translated into different languages and has been increasingly used in both research and clinical settings [1,4–7].

Among the studies that investigated the psychometric properties of the ISI, some reported the factor structure underlying it. In their original study of the psychometric properties of the ISI, Morin and his colleagues proposed a three-factor model based on the results of their exploratory factor analysis (EFA) in a clinical English-speaking population [2]. The first factor was termed the impact component and defined by items referring to distress, interference, and noticeability of insomnia. The second factor was named the severity component, as it contained three items describing the primary symptoms of insomnia at different time points of the night (initial, middle, and terminal). The third factor was called the (dis)satisfaction component and was defined by a single item about overall satisfaction with sleep. According to Bastien et al. [2], these three factors captured the main diagnostic criteria for insomnia. The study by Fernandez-Mendoza et al. [8] supported this by using confirmatory factor analysis (CFA) in a nonclinical Spanish-speaking population. In fact, both of these studies considered the results of

* Corresponding author. Department of Psychology/The Research Center for Mind, Brain, & Learning, National Chengchi University, No. 64, Sec 2, ZhiNan Road, Wenshan District, Taipei, Taiwan. Tel.: +886 2 29387383; fax: +886 2 29390644.
E-mail address: yangcm@nccu.edu.tw (C.-M. Yang).

http://dx.doi.org/10.1016/j.sleep.2014.11.016
1389-9457/

their three-factor model to be evidence supporting the validity of the English and Spanish versions of the ISI, respectively.

However, except for these two studies, all other studies in different countries, according to our knowledge, have yielded different results. One-factor models or different types of two-factor models have been proposed for different versions of the ISI [1,5,7,9–11]. For example, to the best of our knowledge, the three-factor model of the ISI has never been generated in studies with a Chinese version of the ISI. Instead, the one-factor model and different types of two-factor models have been reported in different Chinese-speaking populations in Hong Kong and Taiwan [7,9,11].

There are two intuitive possible interpretations of these inconsistent results: (1) Some versions of the ISI lack factorial validity due to reasons such as inappropriate translations, and (2) as a construct measure, different versions of the ISI really have different factor structures due to the influences of sociocultural factors. However, if any of these reasons are true, the applicability of the ISI as a construct measure will be seriously hindered due to the lack of factorial validity and/or measurement invariance (construct comparability).

Factorial (structural) validity concerns how well a construct measure conforms to the construct's theoretical definition [12]. It is an important piece of evidence that researchers believe should be established before valid inferences interpretation from test scores could be made [13]. Measurement invariance, on the other hand, is the property that allows data sets collected from different groups or sociocultural contexts to be meaningfully compared (eg, results from male and female groups, or results from Chinese and Canadian populations).

Little [14] defined that for cross-cultural studies, the sets of data obtained by using the same construct measure (eg, English and Chinese versions of the ISI) in different cultures will not be comparable to each other unless the strong factorial invariance property holds true. Strong invariance requires that the two data sets have not only the same underlying factor structure but also identical factor loadings and intercepts across factor models. These constraints are necessary; otherwise, people with the same level of a latent trait will yield very different results on manifest variables due to the instrument-specific properties instead of true differences in the latent variables. For example, if the factor loadings of the Chinese version of the ISI (C-ISI) are much higher than those of the Canadian version, then a Chinese subject with any subjective insomnia can score much higher on the ISI than a Canadian, even if they have same level of subjective insomnia, due to the high factor loadings of the C-ISI. Given this condition, researchers will not be able to make meaningful comparisons of the ISI results obtained across countries, for it will be hard to exclude the potential confounding caused by the differences in factor loadings.

As the ISI has become a widely used instrument worldwide, the lack of measurement invariance data across different versions is likely to impede the cross-cultural studies of insomnia with the ISI. Given this reason, it is important to clarify the cause of inconsistent results of factor analyses (FA). With a more systemic review of previous studies on the ISI [1,5,7–11], an alternative explanation for the inconsistencies could be raised. Different researchers tend to use different methods to investigate the factorial structure of the ISI. Furthermore, many researchers have used FA methods that are not recommended by psychometricians and that may potentially lead to problematic results.

First, some previous researchers used principle component analysis (PCA) as an FA [1,5,9–11]. Using PCA is inappropriate because it is a technique for reducing the dimensions of manifest variables. It essentially does not contain the concept of latent variables (ie, psychological constructs or factors). Using PCA as an FA will result in a biased estimation of factor loadings and numbers of factors [15]. Instead, researchers suggest using maximum likelihood (ML) if the data are relatively normal because it can offer many global fit indices and results of statistical testing for researchers to refer to [16]. On the other hand, if the data are non-normal due to the ordinal nature of the scale (eg, a Likert-type scale of less than seven points), then the (robust) weighted least squared method (WLSMV) is recommended as a good alternative [17,18]. WLSMV is an estimator for EFA and CFA that can take the ordinality of items into account. It requires researchers to have the raw data, rather than only the traditional summarized statistics for FA such as correlations, means, and standard deviations. Second, some studies used the Kaiser criterion to decide the number of factors [1]. Although Kaiser's rule is the default method in some statistical programs, it has been considered highly problematic because it can either underestimate or overestimate the number of factors. Instead, jointly considering the information from global fit indices or scree plots can offer researchers more accurate results [16]. Last, some researchers used orthogonal rotation methods [1,9–11]. Latent variables are generally correlated to each other and make the assumption of orthogonality inappropriate. Instead, it has been suggested that an oblique rotation method should be used [19].

In summary, previous studies of the ISI have yielded various FA results. Different factor models, from one-factor model to various types of two- or three-factor models, have been proposed. These inconsistencies will hinder the applicability of the ISI. As many of these studies used some potentially problematic FA methods, it is possible that inconsistencies might be resolved using appropriate methods. Therefore, in the current study, we referred to the results of previous studies (eg, the one-, two-, and three-factor models they proposed) and applied additional analyses to examine the factorial structure of the ISI. Specifically, we first used ordinal EFA and CFA to compare the different factor models of the ISI proposed before, and then examined the measurement invariance properties of the ISI by applying structural equation modeling (SEM) to cross-country data sets. The methods we employed will be elaborated in the method section.

2. Method

A total of three data sets were analyzed in the current study. The first data set contained data from 345 individuals recruited from schools, communities, and a hospital in Taiwan. This data set was collected to address some psychometric issues of the C-ISI. All of the participants filled in the C-ISI. The descriptive statistics of these data are listed in Table 1. The other two sets of descriptive statistics (means, SD, correlation matrixes of the ISI), calculated from two other data sets from Hong Kong and Canada (offered by the original authors), are also included in our study. The descriptive statistics of the Hong Kong data set, offered by Chung and his colleagues, were from a nonclinical sample ($n = 1516$), and the FA results were published in 2011 [9]. The statistics from the Canada data set, provided by Morin and his colleagues, were from a community sample ($n = 959$) in their 2011 study [3]. Readers can refer to the original papers for further details of these two data sets.

Based on the data sets mentioned above, we designed a three-stage analytical procedure. The stages were to: (1) find out the most

Table 1
Descriptive statistics based on the Taiwanese sample ($n = 345$).

Variable	Mean	SD	Skewness	Kurtosis
ISI_1a	0.97	1.08	1.05	0.45
ISI_1b	1.06	1.14	0.85	0.18
ISI_1c	0.88	1.04	0.93	−0.06
ISI_2	2.01	1.05	−0.03	−0.70
ISI_3	1.67	1.04	0.19	−0.67
ISI_4	1.02	1.04	0.71	−0.43
ISI_5	1.46	1.15	0.47	−0.64

Table 2
The fit indices of different EFA models based on the Taiwanese sample (the rotation method used is quartimin, $n = 345$).

Estimator	Model	CFI	TLI	RMSEA	SRMR	AIC[d]	BIC[d]
WLSMV[a]	One-factor	0.906	0.859	0.243	0.110	NA	NA
	Two-factor	0.983	0.955	0.138	0.032	NA	NA
	Three-factor	0.999	0.995	0.047	0.010	NA	NA
MLR[b]	One-factor	0.813	0.720	0.181	0.079	6307.281	6387.996
	Two-factor	0.968	0.916	0.099	0.025	6136.818	6240.593
	Three-factor	0.997	0.982	0.046	0.010	6116.708	6239.701
ML[c]	One-factor	0.819	0.729	0.205	0.079	6307.281	6387.996
	Two-factor	0.976	0.937	0.099	0.025	6136.818	6240.593
	Three-factor	0.998	0.987	0.045	0.010	6116.708	6239.701

[a] ML denotes the maximum likelihood estimator.
[b] MLR denotes the robust maximum likelihood estimator.
[c] WLSMV denotes the *weighted least squares* means and variance adjusted estimator.
[d] AIC and BIC are the model fit indices that will take the model's parsimony into account. They are not available when the WLSMV is used as the estimator.

appropriate factor model by conducting the ordinal EFA on a Taiwanese sample and comparing the global fit indices among different models; (2) cross-validate the EFA results by reanalyzing the statistics obtained from Hong Kong data set with single-group CFA; and (3) examine the cross-cultural measurement invariance property of the ISI by analyzing the statistics from Hong Kong and Canada with multigroup CFA. All analyses were performed with Mplus 7.0.

In the first stage, referring to the literature related to FA [17,18,20], we employed the oblique quartimin rotation method and the robust WLSMV in Mplus to conduct the ordinal EFA on the Taiwanese sample ($n = 345$). The purpose of this analysis was to compare the fitness of all the models that have been proposed before for ISI. The strength of using an ordinal FA estimator such as WLSMV in FA is that it can accurately estimate the factor loadings [17,18]. We believe it is an important property for EFA, as researchers will then need to use the estimated loadings to decide the relation between the items and factors. However, considering that WLSMV cannot offer fit indices such as AIC or BIC, which can take the model parsimony into account, we also conducted the EFA for the Taiwanese sample with ML and robust ML estimators to obtain these two fit indices.

In the second stage, we fitted the FA model that we obtained in stage one to the statistics obtained from Hong Kong with a single-group CFA model and ML estimator. These statistics are offered by Chung and his colleagues. They calculated these statistics from their studies in 2011 [9]. Due to the fact that we have only summarized statistics (ie, correlations, means, and SD) in the data sets from Hong Kong (and from Canada as well), all other factorial analyses in our studies were conducted with ML estimator, except for the EFA analyses in stage one. This set of data originally demonstrated a two-factor structure when it was analyzed with PCA. What we tried to do in this stage was to cross-validate the results we obtained in stage one and further examine our hypothesis about the cause of the inconsistent FA results among previous ISI studies. It must be noted that CFA is a special case of SEM that specifically concerns the parameters in the measurement model (ie, the relations between factors and indicators). Unlike EFA, CFA is highly theory-driven. Researchers should have a clear prior sense of the number of factors and their relationships with indicators in the model [21]. As we had already selected the model, which needed to be further validated from the results of EFA in stage one (ie, the factor model presented in Table 3), the CFA was a reasonable methodological choice for the purposes of this analytical stage.

In stage three, we extended the single-group CFA model in stage two to the scenario of multigroup CFA with ML estimator by further including a set of the ISI statistics obtained from a Canadian sample. This set of statistics was calculated from the community sample in an article from 2011 [3]. In this analytical stage, we tried to answer the research questions on the cross-cultural measurement invariance properties of the ISI. Technically, in order to establish a strong invariance property between two groups, we had to first make sure the configural and weak invariance models were accepted [14]. The configural model is the multigroup CFA model that constrains two groups to the exact same factor structure. The fitness of the configural model can be evaluated by using the fit indices. On the other hand, after establishing the configural model, one can further constrain all the corresponding factor loadings across groups to be equal to establish the weak invariance model. Researchers can accept the weak invariance if the ΔCFI value between configural and weak invariance models is no more than 0.01 [22]. Last, once the configural and weak invariance properties have been established, researchers can then further constrain all the corresponding intercepts across groups to be equal to establish the strong invariance model. Again, the ΔCFI rule can be used as the criterion for accepting the strong invariance model.

The free baseline approach for MI mentioned above (ie, from configural MI to strong MI) was used as the procedure to examine the MI of the ISI. In an ideal situation, cross-cultural comparability can be claimed if the fully strong invariance can be established through this approach [14]. In cases that the full weak or full strong invariance model does not fall in the accepted range (eg, $\Delta CFI > 0.01$), which indicates that at least one of the intercepts or loadings is not identical across groups, the constraint that causes the misfit can be identified by referring to the modification indices and can be allowed to be freely estimated [23]. That is, instead of establishing a fully strong or weak invariance model, a few parameters are allowed to be non-invariant across groups to establish a "partial" invariance model, given that the majority of the items are still invariant across groups [22].

3. Results

First, regarding the number of factors, we compared all the models that had been proposed in previous studies due to the exploratory nature of EFA. Therefore, we compared the one-, two-, and

Table 3
The estimated Factor loadings of the three-factor model in Taiwanese sample ($n = 345$, with quartimin rotation method and WLSMV estimator in Mplus).

Item	Factor 1	Factor 2	Factor 3
ISI_1a (sleep onset)		0.327	0.325
ISI_1b (sleep maintenance)		1.028	
ISI_1c (early morning)		0.752	
ISI_2 (satisfaction)			1.018
ISI_3 (interference)	0.848		
ISI_4 (noticeability)	0.783		
ISI_5 (distress)	0.573		0.347

three-factor models in ordinal EFA. The global fit indices of these models are presented in Table 2. The table shows that the three-factor model fit better than the other two in all indices. It is noticeable that the AIC and BIC values in the three-factor model were also lower than those in the one- or two-factor models. These mean that even after taking the model parsimony into account, it is still reasonable for researchers to adopt the three-factor model due to its sufficient fitness.

After deciding on the number of factors, we used the factor loadings estimate from WLSMV to investigate which item belonged to which factor. We obtained the factor structure presented in Table 3 after ignoring the factor loadings lower than 0.32, as per the rule of thumb suggested in Tabachnick and Fidell [24]. This model is identical with the results that Bastien et al. [2] and Fernandez-Mendoza et al. [8] found by EFA and CFA, respectively (ie, factor 1→impact, factor 2→(dis)satisfication, factor 3→severity). The reliability coefficients of these factors are all around or above 0.8 (Cronbach's α for factor 1 – factor 3: 0.795, 0.792, and 0.821).

In order to test our hypothesis about the cause of inconsistencies in previous studies and to examine the external validity of our results, we reanalyzed the summarized statistics (means, SD, and correlation matrix) offered by Chung and his colleagues (ie, the statistics of the Hong Kong sample) with CFA. These data were originally reported to support a two-factor model in previous research, when PCA and the Kaiser criterion were applied [9]. We fitted the three-factor model we obtained previously to their data by using ML as the estimator. The fit indices showed that the model fit well ($\chi^2 = 29.794$, df = 9, root mean square error of approximation (RMSEA) = 0.039, CFI = 0.994, TLI = 0.987, standardized root mean residual (SRMR) = 0.015), according to the rules of thumb of fit indices summarized by Little [22]. The results of this CFA model are presented in Fig. 1.

To further examine the measurement invariance properties of the ISI on cross-cultural samples, we extended the model in Fig. 1 to the scenario of a multiple-group CFA model by including the statistics offered by Morin and his colleagues into analyses (ie, the statistics of the Canadian sample). Before the multiple-group analysis was conducted, it was necessary to validate the factor structure of each group alone [21]. Therefore, we first fitted the three-factor model we obtained to the Canadian sample with EFA and CFA. The EFA related results are presented in Table 4. Again, all indices indicated that the three-factor model fit the data best. Similar results were obtained in CFA; the three-factor model fit the Canadian data well (estimator: ML, CFI = 0.993, TLI = 0.985, RMSEA = 0.056, SRMR = 0.011).

After confirming the factorial structure of each group, we then used the multiple-group CFA to examine the invariance property between the Hong Kong and Canadian samples. The results are presented in Table 5. All of the invariance models demonstrated sufficient fitness to the data by themselves in the terms of all alternative fit indices. Furthermore, according to Little's [22] suggestion (ΔCFI of no more than 0.01 for weak and strong measurement invariance tests), the strong invariance between the two populations was supported, indicating the assumption that all of corresponding intercept and loadings between the Canadian and Hong Kong sample were identical.

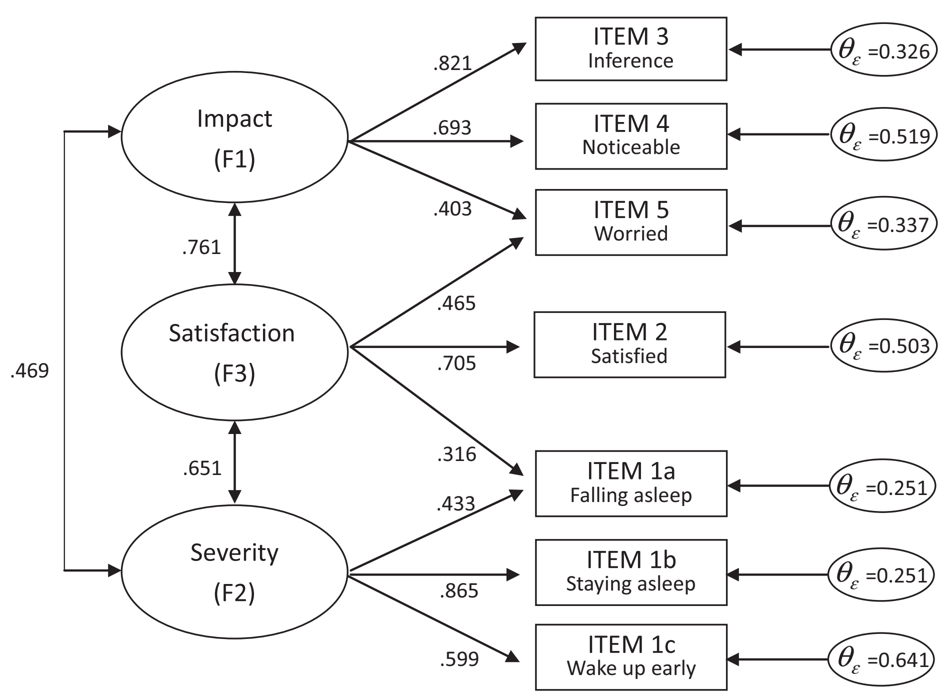

Fig. 1. The path diagram of the three-factor model based on the statistics calculated from the Hong Kong sample (i.e., Chung et al. [9]).

Table 4
The fit indices of different EFA models based on the Canadian sample (the rotation method used is quartimin, $n = 953$).

Estimator	Model	CFI	TLI	RMSEA	SRMR	AIC	BIC
ML[a]	One-factor	0.936	0.904	0.141	0.047	15,055.939	15,123.973
	Two-factor	0.990	0.975	0.072	0.014	14,836.987	14,934.179
	Three-factor	1.000	1.000	0.000	0.003	14,800.591	14,922.081

[a] ML denotes the maximum likelihood estimator.

Table 5
Results of measurement invariance between Hong Kong and Canadian samples.

	χ^2	Df	CFI	ΔCFI	SRMR	RMSEA
Configural	66.052	18	0.994		0.013	0.047
Weak	151.848	24	0.984	–0.010	0.039	0.066
Strong	193.555	28	0.979	–0.005	0.043	0.069

The ML is used as the estimator.

We further included the Taiwanese data into the above analyses. The results of the three-group invariance tests are presented in Table 6. It is noticeable that although the weak invariance model in Table 6 fit well to the data in terms of its own fit indices, ΔCFI between the configural and weak invariance models was slightly greater than 0.01 (ΔCFI = 0.012). Therefore, we referred to the modification indices and allowed the factor loading between factor 1 and ISI_5 in the Taiwanese sample to be freely estimated (i.e., we established the partial weak invariance model in Table 6). All fit indices (including ΔCFI) were found to be satisfied, and the fitness between this model and the data was supported. We then further constrained all the corresponding intercepts across groups to be equal, based on this partial weak invariance model. Again, as shown by the fit indices presented in Table 6, the fitness of this partial strong invariance model was supported by our data. The results in Table 7 indicated that all other 41 parameters (loadings and intercepts) in the model were identical to their corresponding parameters across groups, except for one factor loading in the Taiwanese sample.

4. Discussion

The results of this study show that the three-factor model, which Morin and his colleagues proposed in 2001 (ie, the original factor model for the ISI) [2], is supported in Chinese-speaking populations. Furthermore, we also demonstrated that this model can be successfully replicated in a set of statistics that originally demonstrated a two-factor model with PCA. Jointly considering the reproducibility and the substantial interpretability of this model (ie, corresponding to the clinical diagnosis criteria), we think the validity of this model is adequately supported.

The validation of the factorial structure and MI property of the ISI has important implications that can help future research in several ways. First, our results offer an explanation for the inconsistent results of previous studies. Other researchers can employ the procedure we propose to reevaluate the factorial structural of their versions of the ISI, or even other insomnia-related questionnaires to extend our understanding of insomnia. Second, clarifying the factor structure of the ISI allows future researchers to include the ISI in their study as a “latent construct” (ie, subjective severity of insomnia) and to analyze it with the latent variable techniques like SEM. SEM has some statistical advantages over analysis of variance (ANOVA) and regression, for it can eliminate the influence of measure error on statistical testing if the factor model of the questionnaire is correctly specified. Therefore, based on the three-factor model we have verified, future research can analyze the ISI via SEM to obtain more accurate group mean comparisons without worrying about the impact of measurement error, such as underestimation of the treatment efficacy or the relation among psychological factors.

Last, the invariance property of the ISI on cross-country data sets supports that the ISI can be used as a measure for international insomnia studies. Strong invariance is a necessary property that allows meaningful interpretation of the mean differences across groups (eg, countries and cultures), a common research question that researchers like to address [25]. Little [14] even defined strong invariance as the prerequisite of a construct’s comparability for cross-cultural study. It is noticeable that the three-group CFA in our study indicates that one factor loading in the Taiwanese group is non-invariant. However, referring to the rule of thumb about partial MI suggested by Little [22] (ie, two of three items within a factor are still invariant) and the findings in most recent studies (ie, unequal factor loading has only a small effect on cross-group comparisons) [26], a single non-invariant factor loading in the whole model should not have a profound effect on the cross-group comparison for latent means. Therefore, we think it is reasonable to state that our results support the cross-cultural comparability of the ISI and provide evidence that the ISI is a measure with an internationally valid construct (subjective insomnia).

Table 7
The unstandardized loadings and intercepts of the three-factor model in the three-group CFA (with strong invariance constrains, except the loading between ISI_5 and factor 1 in the Taiwanese sample is freely estimated).

Item	Loadings			Intercept
	Factor 1	Factor 2	Factor 3	
ISI_1a		0.232	0.382	0.806
ISI_1b		0.774		0.592
ISI_1c		0.537		0.648
ISI_2			0.784	1.696
ISI_3	0.778			1.490
ISI_4	0.635			1.087
ISI_5	0.349 (0.464)[a]		0.471	0.966

[a] This is the loading between item ISI_5 and factor 1 in the Taiwanese sample in the three-group CFA.

Table 6
Results of measurement invariance across all samples.

	χ^2	Df	CFI	ΔCFI	SRMR	RMSEA
Configural (M1)	81.765	27	0.994		0.014	0.047
Weak (M2)	195.309	39	0.982	0.012	0.042	0.065
Partial weak[a] (M3)	179.164	38	0.984	0.010[d]	0.040	0.063
Strong with one free loading[b] (M4)	254.399	45	0.976	0.008[e]	0.042	0.070
Strong[c] (M5)	259.136	46	0.976		0.042	0.070

The ML is used as the estimator. M1–M5: model 1–model 5.

[a] This is a partial weak invariance model that allows the loading between factor 1 and ISI_5 in the Taiwanese sample to be different from the other two groups, while all other corresponding loadings across three groups are still constrained to equal.

[b] This is a partial strong invariance model extended from model 3. Except for the loading between factor 1 and ISI_5 in the Taiwanese sample being freely estimated, all other corresponding loadings and intercepts are constrained to be equal across all groups.

[c] This is the typical strong invariance model, which has all corresponding intercepts and loadings across groups constrained to be equal. From the perspective of the free baseline approach, we do not have to demonstrate this model because from the model comparison in M1 versus M2 and M3, we already know that one loading in the Taiwanese group should be freely estimated. However, as some researchers tend to directly evaluate the strong invariance model based on their global fitness, we still added the information of this model to the table for reference.

[d] ΔCFI between M3 and M1.

[e] ΔCFI between M4 and M3.

As far as we know, few insomnia or sleep-related questionnaires have demonstrated cross-cultural measurement invariance before. Considering the fact that insomnia is a common problem worldwide, it is important for researchers to have research instruments that have been validated cross-culturally and can allow meaningful comparison across countries. The measurement invariance property of the ISI that we found fills just this gap, and it should be very useful for future researchers who address subjective perceived insomnia as a common, worldwide construct.

To sum up, in the current study, we have attempted to clarify the factor structure underlying the ISI by making several adjustments according to psychometric principles. Our results not only confirm the three-factor model of the ISI from data collected in different languages but also provide evidence that the ISI could be an internationally valid construct measure. Future studies can employ the ISI as a valid construct measure of subjective insomnia to study cultural aspects of insomnia.

In light of the significant implications of our findings, some limitations should be kept in mind while interpreting the results. The first limitation is about the estimators used in our factor analyses. The ordinal estimator (WLSMV) was conducted only for Taiwanese data because we do not have the raw data for the Hong Kong and Canadian samples.

According to the previous literature [27,28], an ordinal estimator such as WLSMV can take the ordinality of Likert-type scales into account and provide more precise estimation of the fit indices and factor loadings. Although methodological studies have shown that once the Likert-type scales have at least five points (eg, the items in the ISI), researchers can treat them with ML because the biases will become negligible [29,30]. If raw data of the ISI from different samples are available, future studies will be able to select other appropriate estimators to further confirm our results. For example, one could use WLSMV for all samples to obtain precise estimations for factor loadings or use either robust ML or ML to obtain AIC and BIC for nonnested model comparison. Researchers can even cross-validate the MI results obtained from different estimators to provide further support for their conclusions. Furthermore, considering that most of the participants in our study were a nonclinical sample, future studies can try to replicate our results to more clinical populations.

Conflict of interest

The ICMJE Uniform Disclosure Form for Potential Conflicts of Interest associated with this article can be viewed by clicking on the following link: http://dx.doi.org/10.1016/j.sleep.2014.11.016.

Acknowledgments

The authors gratefully acknowledge Dr. Hans Ivers, Dr. Ka Fai Chung, and Dr. Wing Fai Yeung for providing us with the valuable statistics from their studies. Our thanks also go to Dr. Shih-Chieh Hsu and the members of the sleep laboratory at National Chengchi University for their help in the process of data collection. Last, we also wish to thank all of our participants and the Department of Psychology at the National Chengchi University for their support.

References

[1] Sadeghniiat-Haghighi K, Montazeri A, Khajeh-Mehrizi S, Nedjat C, Aminian O. The Insomnia Severity Index: cross-cultural adaptation and psychometric evaluation of a Persian version. Qual Life Res 2013;23:533–7. doi:10.1007/s11136-013-0489-3.
[2] Bastien CH, Vallieres A, Morin CM. Validation of the insomnia severity index as a clinical outcome measure for insomnia research. Sleep Med 2001;2:297–307.
[3] Morin CM, Belleville G, Belanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011;34:601–8.
[4] Gupta R, Lahan V. Translation and validation of insomnia severity index in hindi language. Indian J Psychol Med 2011;33:172–6.
[5] Sierra JC, Guillen-Serrano V, Santos-Iglesias P. Insomnia Severity Index: some indicators about its reliability and validity on an older adult sample. Rev Neurol 2008;47:566–70.
[6] Suleiman KH, Yates BC. Translating the insomnia severity index into Arabic. J Nurs Scholarsh 2001;43:49–53.
[7] Yang CM, Hsu SC, Lin SC, Chou YY, Chen YM. Reliability and validity of the Chinese version of Insomnia Severity Index. Arch Clin Psychol 2009;4(2):95–104 [in Chinese].
[8] Fernandez-Mendoza J, Rodriguez-Muñoz A, Vela-Bueno A, Olavarrieta-Bernardino S, Calhoun SL, Bixler EO. The Spanish version of the Insomnia Severity Index: a confirmatory factor analysis. Sleep Med 2012;13(2):207–10.
[9] Chung KF, Kan KK, Yeung WF. Assessing insomnia in adolescents: comparison of Insomnia Severity Index, Athens Insomnia Scale and Sleep Quality Index. Sleep Med 2011;12(5):463–70.
[10] Savard MH, Savard J, Simard S, Ivers H. Empirical validation of the Insomnia Severity Index in cancer patients. Psychooncology 2005;14:429–41.
[11] Yu DS. Insomnia Severity Index: psychometric properties with Chinese community-dwelling older people. J Adv Nurs 2010;66:2350–9.
[12] Hoyle RH, Smith GT. Formulating clinical research hypotheses as structural equation models: a conceptual overview. J Consult Clin Psychol 1994;62:429–40.
[13] Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychologist 1995;50:741–9.
[14] Little TD. Mean and covariance structures (MACS): analyses of cross-cultural data. Multivariate Behav Res 1997;32:53–76.
[15] Widaman KF. Common factor-analysis versus principle component analysis – differential bias in representing model parameters. Multivariate Behav Res 1993;28(3):263–311.
[16] Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods 1999;4:272–99.
[17] Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods 2004;9:466–91.
[18] Hogado-Tello F, Chacon-Moscoso S, Barbero-Garcia I, Vila-Abad E. Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. Qual Quant 2010;44:153–66.
[19] Costello AB, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval 2005;10(7):1–9.
[20] Browne M. An overview of analytic rotation in exploratory factor analysis. Multivariate Behav Res 2001;36:111–50.
[21] Brown TA. Confirmatory factor analysis for applied research. New York: Guilford Press; 2006.
[22] Little TD. Longitudinal structural equation modeling. New York: Guilford; 2013.
[23] Yoon M, Kim ES. A comparison of sequential and nonsequential specification searches in testing factorial invariance. Behav Res 2014;46:1199–206. doi:10.3758/s13428-013-0430-2.
[24] Tabachnick BG, Fidell LS. Using multivariate statistics. Boston: Allyn and Bacon; 2001.
[25] Widaman KF, Reise SP. Exploring the measurement invariance of psychological instruments: applications in the substance abuse domain. In: Bryant KJ, Windle M, editors. The science of prevention: methodological advance from alcohol and substance abuse research. Washington, DC: American Psychological Association; 1997. p. 281–324.
[26] Steinmetz H. Analyzing observed composite differences across groups is partial measurement invariance enough? Methodology 2013;9:1–12.
[27] Green SB, Akey TM, Fleming KK, Hershberger SL, Marquis JG. Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. Struct Equ Model 1997;4:108–20.
[28] DiStefano C. Effects of ordered categorical data with confirmatory factor analysis. Struct Equ Model 2002;9:327–46.
[29] Beauducel A, Herzberg PY. On the Performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. Struct Equ Model 2006;13(2):186–203.
[30] Rhemtulla M, Brosseau-Liard PÉ, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. Psychol Methods 2011;17(3):354–73.