Psychological Testing
June 2023, Vol.70 No.2, pp.109-132

Measurement Invariance of a Developed Leisure Motivation Scale With Regard to the Background Variables of Physical Education Teachers

Wei-Ching Lee

Abstract

In this study, a reliable and valid leisure motivation scale was constructed for physical education teachers in Taiwanese junior high and elementary schools. The developed scale was evaluated for invariances in the background variables of physical education teachers. In total, 878 valid samples were collected through stratified random sampling and cluster sampling. The samples were randomly divided into two sets. The first sample set was examined through item analysis, confirmatory factor analysis, and multigroup analysis. The second sample set was tested through cross-validation. The results indicated that the items of the developed scale are appropriate, and the collected data fit the developed theoretical model, which indicated that this model is stable and predictable. Therefore, the proposed scale is reliable and valid. Invariance was identified in two variables: teaching stage and graduation department. The partial invariance in gender and teaching location is related to competence mastery and intellectual categories, respectively.

Keywords: Confirmatory factor analysis, cross-validation, multigroup analysis

Wei-Ching Lee, Associate Professor, Department of Adapted Physical Education, National Taiwan Sport University
Received: 2022.09.16; Revised: 2022.12.30; Accepted: 2023.01.12
Email: lwceeh@gmail.com
Address: No. 250, Wenhua 1st Rd., Guishan Dist., Taoyuan City 333325, Taiwan
Department of Adapted Physical Education, National Taiwan Sport University

《測驗學刊》
第七十輯第二期　2023 年 6 月 109～132 頁

「休閒動機量表」在體育教師背景變項之測驗恆等性檢定

李偉清

摘要

本研究旨在針對國中小體育教師編製一份具有信效度之「休閒動機量表」，並比較不同背景的體育教師在此工具上是否具有組間不變性。研究者透過分層隨機與叢集取樣共獲得 878 份有效樣本，並將此樣本隨機分為二等分。先以第一組樣本作項目分析、驗證性因素分析和多群組分析，然後再用第二組樣本進行複核效化研究。結果發現，本量表題目非常適切，且蒐集的資料適配本研究所提出之理論模式，顯示此模式具有穩定與預測性。因此，本研究的量表是一份具有信效度的量表，且體育教師的「任教階段」及「畢業科系」變項在該量表上均具有組間不變性，但「性別」與「任教地點」變項僅具有部分之測量不變性，主要差異則分別源於「勝任熟練性」和「智力性」構面題目之影響。

關鍵詞：多群組分析、複核效化、驗證性因素分析

李偉清，國立體育大學適應體育學系副教授
收件日期：2022.09.16；完成修改：2022.12.30；正式接受：2023.01.12
Email：lwceeh@gmail.com
地址：333325 桃園市龜山區文化一路 250 號 國立體育大學適應體育學系

Introduction

Motivation is among the most researched themes in the field of leisure research (Luo et al., 2022). However, most studies on motivations for leisure have focused on young people (Caldwell et al., 2010; Chen et al., 2013; Kueh et al., 2018, 2019; Padhy et al., 2015; Santos-Labrador et al., 2021; Tsai & Lin, 2008; Wang & Chiang, 2003; Weybright et al., 2020), college students (Chien, 2016; Fehmi et al., 2014; Kim et al., 2019; Kueh et al., 2020; Wu, 2015), tourists (Huang et al., 2017; Sharma et al., 2014; Tu & Li, 2016), silver-haired people (Beggs et al., 2014; Gadiraju et al., 2022; Huang, 2017; Tsai & Kuo, 2016), and athletes (Ayyildiz Durhan, 2019; Chen et al., 2017; Jia, 2018; Koç et al., 2019; Kuo et al., 2016; Lin et al., 2007; Qiu et al., 2020; Van Lankveld et al., 2021). Less attention has been given to leisure motivations among teachers. Empirical research on this aspect in Taiwan has centered on grade-school-level teachers (Huang et al., 2012, 2013; Wang et al., 2009), university professors (Chen, 2013; Chen & Chen, 2004; Lee et al., 2008, Li et al., 2018), and retired teachers (Tsai, 2012; Tsai & Lee, 2011). However, no study has specifically addressed physical education (PE) teachers. Therefore, in the present study, a reliable and valid leisure motivation scale was created for PE teachers in Taiwanese junior high and elementary schools. The developed scale was then used to evaluate the background of PE teachers in compulsory education. Finally, a norm was established to obtain reference results for future scale measurement studies.

Definition and measurement of leisure motivation

Motivation is defined as an energetic behavior directed towards a specific purpose, and one of the most distinguishing characteristics of motivation is its individuality. The human being is at the heart of motivation. Motivation is a subjective experience that changes from person to person (Dias et al., 2022). Leisure motivation is a driving force for leisure activities participation (Belošević & Ferić, 2022). Dimensions of leisure motivation are often varied and inconsistent. Studies have divided leisure motivation into 3-17 categories (Crandall, 1980; Lin, 1978; Tinsley & Tinsley, 1986) on the basis of various arguments (Lee, 2011). Hsieh (2004) considered the

four categories of Beard and Ragheb (1983) to be the most intuitive because they cover a wide range of content and are presented clearly. Beard and Ragheb provided psychological and social reasons to explain why people participate in leisurely behavior. Their four leisure motivation categories are as follows:

1. Intellectual (IN) motivation: Under intellectual motivation, people perform mental activities such as discovering new ideas, learning new skills, exploring, expanding knowledge, satisfying curiosities, creating, and imagining.

2. Competence mastery (CM): Under this motivation, people perform physical activities such as demonstrating athletic talents, improving sports skills and abilities, strengthening physical fitness, and maintaining body shape.

3. Stimulus avoidance (SA): Under this motivation, people perform activities such as relaxing, slowing down, resting, and avoiding crowds and noises. The individual attempts to escape from excessive stimulation and social contact and pursues a more peaceful environment.

4. Social (SO) motivation: Under this motivation, people perform activities such as interacting with other people, establishing friendships, seeking confidants, attempting to experience a sense of belonging, and attempting to be recognized by peers. Under social motivation, two basic needs are sought: interpersonal relationships and the respect of others (Beard & Ragheb, 1983; Lee, 2011).

On the basis of the leisure motivation theory of Beard and Ragheb (1983), several scales have been created for evaluating the leisure motivations among the general public and college students (Hsieh, 2004; Hsieh, 1998; Kim et al., 2019; Lounsbury & Franz, 1990), teenage offenders (Reddon et al., 1996), teenagers in Taipei (Wang & Chiang, 2003), gifted students (Lee, 2011), virtual-reality gamers (Isaacs et al., 2021), swimmers (Koç et al., 2019), managers (Murray & Nakajima, 1999), everyday tourists (Ryan & Glendon, 1998), and farm tourists (Lin, 2001). Currently, only Huang et al. (2012) has created a leisure motivation scale for teachers in compulsory education. However, they only evaluated junior high school teachers in Tainan, Taiwan, and could not assess teachers across the country, and did not focus on PE teachers. Therefore, in the present study, a leisure motivation scale was created to evaluate PE teachers in junior high schools and elementary schools in Taiwan in terms of the four leisure m otivation categories of Beard and Ragheb (1983). By using the de-

veloped scale, the leisure motivation of PE teachers in Taiwan was defined and measured.

Methodology

Research participants and survey samples

The participants of this study were junior high and elementary school PE teachers in 19 counties and cities in Taiwan. According to the Education Statistics Inquiry Network of the Department of Statistics of the Ministry of Education of Taiwan, the numbers of classes in junior high schools and elementary schools in Taiwanese counties and cities were 22,525 (31%) and 51,119 (69%), respectively, on May 4, 2020, with the total number of classes being 73,644. Therefore, the maximum study population was approximately 73,644.

A sampling method that involved combining stratified random sampling and cluster sampling was adopted. If 1/100 of the population was sampled, 736 questionnaires were distributed. However, the recovery rate of the questionnaire could be poor, and some questionnaires could be invalid. Consequently, the number of questionnaires sent out in this study was 900. If three PE teachers were selected from each school, 300 schools were selected. The proportion of the number of classes was then multiplied by the total number of schools (Table 1), and the rounding method was used to obtain the number of required school samples.

The research sample was created by randomly sampling elementary and junior high schools listed on the education department websites of each county and city government in Taiwan. This sample was divided into two equal groups by using IBM SPSS Statistics 23 software program. For the first set of samples, the IBM SPSS Statistics 23 and IBM SPSS Amos 23 software programs were used to perform item analysis and confirmatory factor analysis (CFA) to determine the best theoretical construction and the most appropriate items. On the basis of the results of CFA and item analysis, the structural equation modeling (SEM) goodness-of-fit test and multigroup analysis were performed to establish the theoretical model and compare its parameters with those of the cross-sample model. The second set of samples was used to ver-

ify whether the developed theoretical model was stable and predictable.

Table 1. *Number of schools that were sampled.*

County and city	Number of junior high schools	Number of elementary schools
Keelung City	2	3
Taipei City	10	19
New Taipei City	15	31
Taoyuan City	10	19
Hsinchu City	2	4
Hsinchu County	2	6
Miaoli County	3	6
Taichung City	12	25
Changhua County	5	12
Nantou County	2	6
Yunlin County	3	7
Chiayi City	1	2
Chiayi County	2	5
Tainan City	8	15
Kaohsiung City	12	22
Pingtung County	4	8
Yilan County	2	5
Hualien County	2	4
Taitung County	1	3
Total	99	201

A total of 881 questionnaires were collected. After eliminating three invalid questionnaires, 878 valid questionnaires remained, and the recovery rate of valid questionnaires was 97.56%. Therefore, 439 (878/2) participants were included in each sample group.

Research tools

The research objects of this study included junior high school and elementary school PE teachers, some of these teachers graduated from sports-related departments and some graduated from general departments, some teachers taught in urban schools

and some taught in rural schools, as well as referring to the leisure motivation scales of Lee (2011) and Huang et al. (2012), both explored gender variables, so the developed scale was used to determine the demographics: gender, teaching stage, graduation department, and school location of the PE teachers participating in this study.

Since the scale of Huang et al. (2012) was not very rigorous in the process of compilation. He only used exploratory factor analysis (EFA) to eliminate test items. It was not as good as Lee (2011) who used EFA, item analysis, and CFA of SEM to delete items, also verified the theoretical framework of Beard and Ragheb (1983), and conducted reliability analysis. Therefore, the scale of Lee had better reliability and validity, so this research adopted all items from the Leisure Motivation Scale for Gifted Students, which was developed by Lee. This scale contains relatively few items and has high reliability and validity. The developed scale contained items on the four leisure motivation categories of Beard and Ragheb, with the IN, CM, SA, and SO categories containing six, seven, three, and three items, respectively; thus, the developed scale comprised 19 items.

When completing the questionnaire, teachers were asked to select the most suitable answers on a 5-point Likert scale. The higher the total score, the stronger was the leisure motivation.

An alpha coefficient analysis of the four subscales was conducted to determine reliability. During CFA, the observed variables were analyzed for individual reliability, and the latent variables were analyzed for construct reliability. The content validity, construct validity, cross-validation, convergence validity, and discriminant validity were evaluated to verify the applicability of the theoretical framework of Beard and Ragheb (1983) to the sample of this study.

Data processing analysis

Means and standard deviations were calculated, and norm was built in addition to the aforementioned item and reliability analyses, CFA, SEM, and multigroup analysis. Moreover, the backgrounds of the PE teachers were evaluated, and the invariance of the developed scale was measured. Finally, the stability of the developed model was tested for validity.

Results

Item analysis and CFA

The 19 items of the developed scale were selected sequentially for item analysis and CFA. The criterion for eliminating items was a nonsignificant ($p < 0.05$) critical ratio (CR), nonsignificant correlations or correlations of less than 0.30 with the total score of the subscale, and a factor loading of less than 0.70 in CFA. None of the items met the aforementioned criteria; thus, all the 19 items were retained and analyzed.

Observed variable examination and multivariate normality verification

The means of the scale items were between 3.85 and 4.14, and the standard deviations of the items ranged from 0.67 to 0.89. Therefore, the indicators of the developed scale were appropriate. The skewness values of the observed variables were between -0.49 and -0.13, and the kurtosis values were between -1.01 and 0.33. According to Kline (2005), the absolute skewness values must be greater than 2 to be considered extreme, and the absolute kurtosis values must be greater than 7 for problems to arise. Therefore, the variables of the developed scale were consistent with a univariate normal distribution. To determine whether variables exhibited a multivariate normal distribution, the bootstrap method was adopted to re-estimate the bias and measure the accuracy of the maximum likelihood method.

After conducting bootstrapping 5,000 times, the results indicated that the differences between the estimated standard error and the standard error estimated using the maximum likelihood method were extremely small. The maximum error was less than 0.001, which indicated that the differences between the parameters estimated using the two methods were small. Therefore, even if the data deviated from the multivariate normal distribution, bias did not occur when estimating parameters.

Convergence validity and discriminant validity analysis

The research model comprised four constructs, and the results of CFA for each construct are present in Table 2. The R^2 (squared multiple correlation) values of all

Table 2. *Reliability and AVE of the observed and latent variables.*

Latent variable	Observed variable	R^2	Composite reliability	AVE
IN			.94	.72
	1. To learn and expand knowledge.	.72		
	2. To stimulate imagination or cultivate creativity.	.71		
	3. To undertake more mentally demanding tasks.	.61		
	4. To acquire more life experiences.	.85		
	5. To explore and discover new things.	.78		
	6. To try something new.	.62		
CM			.95	.71
	7. To develop athletic ability.	.62		
	8. To challenge athletic skills.	.61		
	9. To improve sports skills.	.77		
	10. To improve physical fitness.	.79		
	11. To gain self-confidence.	.68		
	12. To feel accomplished.	.78		
	13. To improve spiritual vitality.	.75		
SA			.90	.74
	14. To get away from crowds and noise.	.81		
	15. To be in a quiet environment.	.72		
	16. To slow down the pace of life.	.68		
SO			.95	.85
	17. To establish and maintain good friendships.	.86		
	18. To expand interpersonal relationships or connections.	.84		
	19. To meet new friends or different groups of people.	.86		

the observed indicators were greater than 0.50. The results of the individual reliability evaluation were satisfactory. Moreover, the construct reliability of the latent variables was considerably greater than 0.70, which indicated that the observation indicators provided credible construct measurements for the four latent variables. The four average variances extracted (AVE) were higher than 0.50, which revealed that the variance of the observed variables was rarely affected by the measurement error. All the aforementioned results were consistent with the convergent validity standards pro-

posed by Fornell and Larcker (1981) and Hair et al. (2009). Cronbach's α coefficient was between 0.89 and 0.95, and all coefficients reached reliability values above 0.70. Thus, the constructs of the research model had su fficient convergent validity.

In the discriminant validity analysis, the confidence interval (CI) method was used to examine the correlation coefficients between the constructs. Torkzadeh et al. (2003) noted that if the CI does not include 1, discriminant validity is present between the constructs. The bootstrap method was used to repeat the estimation 5,000 times. At a 95% CI, the results revealed no case in which the CI included 1 (Table 3). Therefore, the constructs demonstrated discriminant validity.

Table 3. *Estimation of Cis.*

Parameter	Point estimate	Bias-corrected		Percentile method		Point estimate ± 2 standard error	
		Lower	Upper	Lower	Upper	Lower	Upper
CM <--> IN	.43	.34	.53	.33	.53	.33	.53
SO <--> SA	.64	.55	.72	.54	.72	.54	.72
SO <--> CM	.77	.71	.83	.70	.82	.70	.83
I N<--> SO	.33	.22	.43	.22	.43	.22	.43
SA <--> CM	.78	.72	.84	.71	.83	.71	.84
SA <--> IN	.42	.32	.51	.32	.50	.32	.51

Checking of offending estimates

Before model evaluation, ensuring that the estimated parameters did not violate the acceptable range of statistics was essential. First, the estimated parameters did not have any negative error variance, and the error variances were significant (0.07-0.25). Second, the standardized errors were not high (0.01-0.06), and the standardized regression weights were less than 0.95 (0.33-0.93). Therefore, no offending estimates were present in this study, which met the criteria proposed by Hair et al. (2009). Consequently, the overall model fit test and validity test of the individual variables were credible and valid.

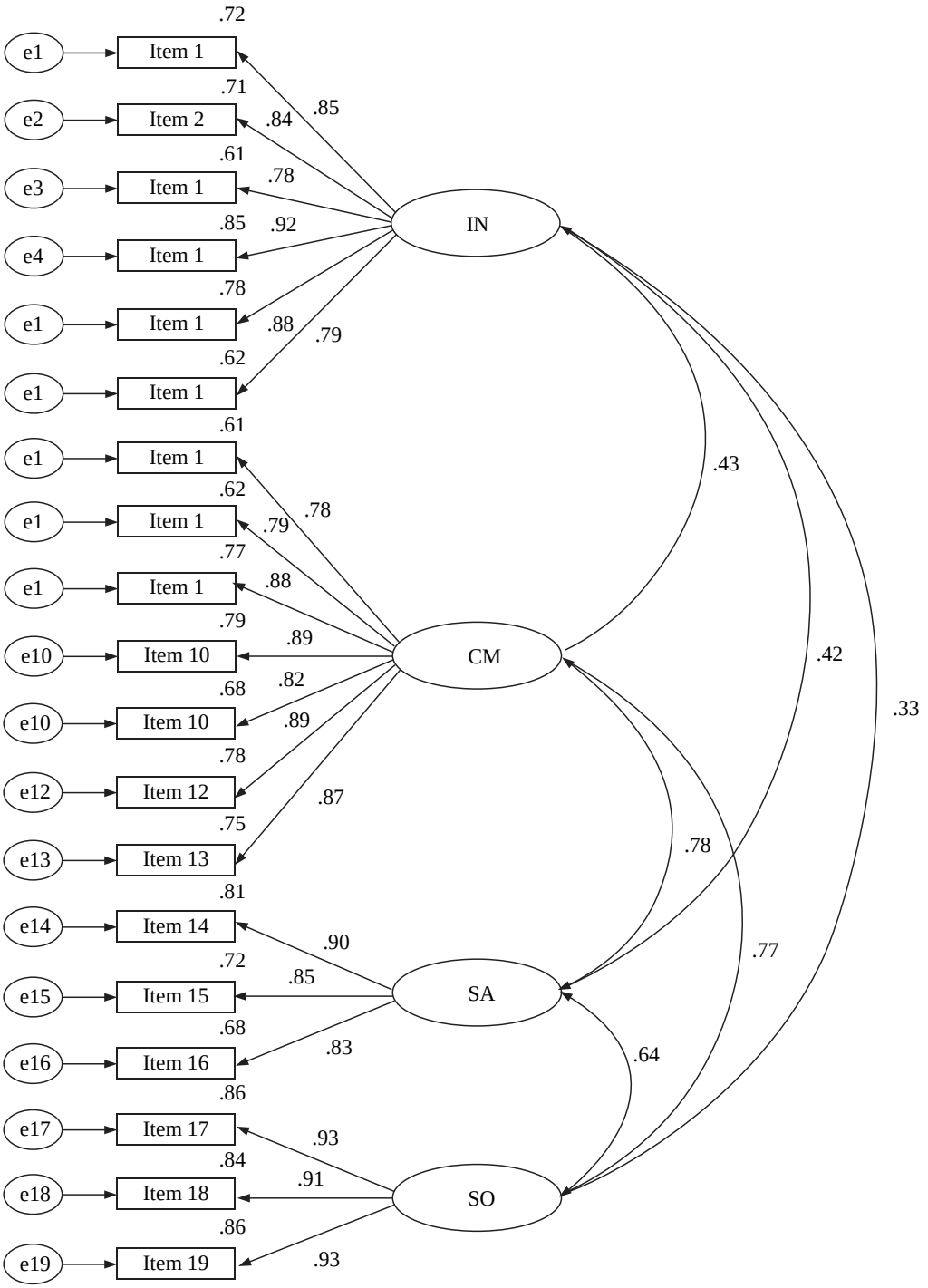
Overall model fit

The analysis results of the measurement model (Table 4) indicated that the chi-squared value was too high, and the *p* value (0.000) was significant because of the large sample or poor model fit. Bollen and Stine (1992) suggested that the Bollen-Stine *p* value correction can be used to correct the chi-squared value. After bootstrapping the model 5,000 times, 5,000 well-fitted models were produced. No model had a poor fit, and the probability of another poor-fit model was low ($p = 0.0002$), which indicated that the chi-squared value of the bootstrap sample model differed from that of the original sample. The chi-squared values of the 5,000 models were better than that of the original sample. Thus, the *p* value of the original model was significant because of the large sample size and not the model setting. The average chi-squared value estimated using the bootstrap method was used to recalculate all the model fit indicators. All the recalculated indicators met the evaluation standards, and all the models had good fit (Table 4). It was true that the large sample size caused the chi-squared value to increase, and the *p* value was significant. The standardized coefficients estimated through CFA are presented in Figure 1.

Table 4. *Results of the testing of the hypothesized model for examining the overall model fit index.*

Evaluation index	Evaluation criteria	Analysis result	Analysis result correction
Absolute fit measures	χ^2 is not significant	1704.10 ($p = .000$)	227.37
	GFI > .9	.76	.98
	AGFI > .9	.69	.97
	RMSEA < .05	.16	.04
Incremental fit measures	NFI > .9	.82	.98
	RFI > .9	.78	.97
	IFI > .9	.83	.99
	TLI > .9	.80	.99
	CFI > .9	.83	.99
Parsimonious fit measures	χ^2/df between 1 to 3	11.67	1.56
	PNFI > .5	.70	.83
	PCFI > .5	.71	.83

Figure 1. *Standardization coefficients of the hypothesized model.*



Multigroup analysis

Byrne (2001) suggested that multigroup analysis can be conducted according to Joreskog's traditional approach to test whether the factor loading of each observed variable is equal among multiple groups. If some observed factor loadings have measurement invariances among groups, then when the new set of parameters are tested for equivalency, the parameters can be restricted. Therefore, the group invariance of the research model was tested through a series of gradually severe statistical tests, which is called a nested model comparison.

Regardless of the background variable, the value of the unconstrained model reached a significance value of 0.05 (Table 5), which indicated that the developed scale differed for PE teachers of various backgrounds. The configural invariance (i.e., the number of factors and the number of items) might have been different. However, because the null hypothesis test of the unconstrained model is often rejected on account of the lack of baseline model comparisons, the model fit can be confirmed by examining other indicators of the unconstrained model (Li, 2006). After evaluating other indicators of the overall model fit test (such as gender, teaching stage, graduation department, and teaching location; IFI = 0.82, 0.82, 0.82, and 0.83 for these variables, respectively; CFI = 0.82, 0.82, 0.82, and 0.83 for these variables, respectively; PNFI = 0.68, 0.68, 0.68, and 0.68 for these variables, respectively; and PCFI = 0.70, 0.70, 0.70, and 0.71 for these variables, respectively) and after the comprehensive evaluation, the model fit was acceptable. Therefore, the configural invariances of PE teachers can be regarded as equivalent.

According to the nested model comparison, only teaching stage and graduation department exhibited group invariance ($p > 0.05$) for measurement weights. For structural covariances, all variables except gender were invariant among groups ($p > 0.05$). However, all variables were significantly different among groups ($p < 0.05$) for measurement residuals. Li (2006) argued that $\Delta\chi^2$ is affected by sample size. When the number of samples is high, the p value can easily be significant even if the $\Delta\chi^2$ value is small. Cheung and Rensvold (2002) noted that if $\Delta\text{CFI} \leq |0.01|$, ΔCFI is not practically significant. Thus, items that are not exactly equal have little effect on the overall model fit. Little (1997) proposed that if $\Delta\text{TLI} \leq 0.05$, no difference exists am-

ong the nested structure models, that is, the comparison among groups indicates no difference. Tabachnick and Fidell (2001) emphasized that when parameters su ch as factor loading and factor covariances are equivalent among groups, the measurement model is invariant.

The comparative results of the group variables in the present study indicated that some *p* values were greater than 0.05, some CFI differences were significantly less than |0.01|, and all TLI differences were less than 0.05. Therefore, the developed scale had measurement invariance for teaching stage and graduation department but only partial invariance for gender and teaching location.

Table 5. *Results of multigroup analysis.*

Variable	Model	χ^2	df	$\Delta\chi^2$	Δdf	*p*	ΔCFI	ΔTLI
Gender	Unconstrained	1973.88	292	--	--	.000	--	--
	Measurement weights	2004.57	307	3	15	.010	.001	-.009
	Structural residuals	2032.69	317	2	10	.002	.002	-.004
	Measurement residuals	2084.12	336	5	19	.000	.004	-.008
Teaching stage	Unconstrained	1962.91	292	--	--	.000	--	--
	Measurement weights	1983.44	307	2	15	.152	.000	-.010
	Structural residuals	1994.83	317	1	10	.328	.001	-.006
	Measurement residuals	2216.49	336	2	19	.000	.021	.011
Graduation department	Unconstrained	1983.34	292	--	--	.000	--	--
	Measurement weights	1993.63	307	1	15	.801	-.000	-.011
	Structural residuals	2009.83	317	1	10	.094	.001	-.006
	Measurement residuals	2237.36	336	2	19	.000	.022	.012
Teaching location	Unconstrained	1899.53	292	--	--	.000	--	--
	Measurement weights	1960.93	307	6	15	.000	.005	-.004
	Structural residuals	1971.87	317	1	10	.362	.001	-.006
	Measurement residuals	2079.79	336	1	19	.000	.009	-.001

A posteriori test of invariance for gender and teaching location was conducted to evaluate the difference in the measurement weights among group variables. Checked that the absolute value of the CR among parameters was greater than the critical value of 1.96 ($p < 0.05$). According to the parameter pairing comparison, differences between male and female PE teachers were primarily observed in the factor loadings

of items 9 to 13 in the CM construct. In teaching location, differences between urban and rural schools were primarily observed in the factor loadings of items 3 and 6 in the IN construct, and the constrained paired parameters of the remaining items did not exceed the critical value.

Cross-validation

To avoid idiosyncratic samples and for further verifying the stability of the developed theoretical model, the second set of independent samples was used as a validity sample to conduct an Amos cross-validation analysis. When the 19 factor loadings of the calibration and validity samples were equal, the chi-squared value increased by 10.71, and the test result was $p = 0.773$ (Table 6), which did not reach the significant level of 0.05. Thus, the 19 equal factor loadings were acceptable. Under the constraints of maintaining the measurement weight model, 10 parameters were set to be equal, with four variances and six covariances being observed. Consequently, the chi-squared value increased by 10.89 ($p = 0.366$), which was nonsignificant. Therefore, the 10 equal parameters were also acceptable. Under the constraints of maintaining the structural covariance model, 19 measurement residuals were added to the model and made equal. Consequently, the chi-square value increased by 66.64 ($p = 0.000$), which was significant. Thus, at least one of the 19 measurement residuals was not equal. Byrne (2010) argued that if the equivalence of the structural covariance model can be achieved, the cross-validity can be verified even if the measurement residuals are unequal because the equivalence of the measurement residuals is too strict. The ΔCFI value between the two models in the present study was within 0.01, and the ΔTLI value was considerably less than 0.05, which supported cross-validation of the calibration and validity samples. Thus, no practical difference was noted between the two models. Therefore, the developed model is effective and can be used to evaluate samples of various groups in the same population.

Table 6. *Evaluation of the model stability.*

Model	χ^2	df	$\Delta\chi^2$	Δdf	*p*	ΔCFI	ΔTLI
Unconstrained	2665.61	292	--	--	.000	--	--
Measurement weights	2676.32	307	10.71	15	.773	-.000	-.008
Structural residuals	2687.21	317	10.89	10	.366	.000	-.005
Measurement residuals	2753.84	336	66.64	19	.000	.003	-.005

Norm establishment

Gender, teaching stage, graduation department, and teaching location exhibited no significant differences in their total scores on the developed scale. The sequence results were t = -0.91, -0.15, 1.58, and -0.29 (p = 0.364, 0.88, 0.114, and 0.772, respectively). None of these results indicated significant differences in leisure motivation among the PE teachers. Therefore, the developed scale did not require a norm for the different strata of each background variable. Consequently, only comparisons of raw scores, percentile ranks, and t scores were conducted to obtain reference results for future studies on scale measurement.

Discussion

In this study, SEM, which is a mainstream statistics method, was used to not only analyze the potential structure and simplify the data but also verify the developed model and estimate the measurement error. Moreover, the reliability and validity of the developed scale and the group invariance were tested. This was the biggest difference of the compilation between the developed scale and previously proposed leisure motivation scales.

The CFA model of the developed scale exhibited an ideal model fit. This model was supported by observational data, and its cross-validation results were suitable, which indicates that the factor structure of the developed scale is stable. In addition to fully echoing the scale structure of other studies (Hsieh, 2004; Isaacs et al., 2021; Kim et al., 2019; Koç et al., 2019; Tsai, 2012), the developed scale reinforces the feasibility of the theoretical model of Beard and Ragheb (1983). Their leisure motivation theory can also be applied to PE teachers because of the validity, stability, and predic-

tability of their model. Finally, the developed model can explain the motivations for individuals to participate in leisurely activities.

According to the standardized regression weights and the values of R^2 in Figure 1, Item 4 is more relevant for IN than are the other five items. For CM, Item 10 is more prominent than are the other six items. Item 14 is the most dominant item for SA. For SO, Item 19 reflects a higher validity than do the other two items. Therefore, Items 4, 10, 14, and 19 are the most relevant items within the corresponding constructs and have the strongest influences. Thus, these items are the most suitable for evaluating potential factors of the developed scale.

The four categories of leisure motivation had a moderate correlation (the coefficient was between 0.33 and 0.78), and the correlation did not exceed 0.85 (Figure 1); thus, the developed model should have not another higher-level factor structure (Hwang, 2007; Li, 2006). To be cautious, the standard errors of the correlation coefficients between the factors were checked again, and the approximate CI of the true correlation was 0.22-0.84 (Table 3). Because the CI did not include 1, the developed model can be a four-construct measurement model.

The mean scores of the items of the developed scale were between 3.85 and 4.14, which was higher than the midpoint three scores of each item. Therefore, all the Taiwanese PE teachers recruited in this study exhibited strong leisure motivation, which is consistent with the findings of other studies (Chen, 2013; Chen & Chen, 2004; Huang et al., 2012; Tsai, 2012; Tsai & Lee, 2011; Wang et al., 2009). Their strong leisure motivation might be associated with the pressure of daily teaching. Teaching involves long working hours, and high energy expenditure, regardless of PE teacher background. Thus, strong leisure motivation can be expected among PE teachers.

Conclusion

Several conclusions can be drawn from this study. When the overall model fit of the theoretical model was tested, the fit between the model and the observed data was satisfactory, and the reliability and validity of the model were ideal. Thus, the categorical constructs are stable, reliable, and valid. Moreover, the hypothetical model

has excellent internal and external qualities. To inspect the cross-validation, a different sample with the same number of participants as the original sample was used to test the validity. The results indicate that the proposed leisure motivation model is stable, inferential, valid, and predictable, and this model exhibits good fit.

To compare whether PE teachers with various backgrounds differ among groups, a nested model comparison was conducted in multigroup analysis. The results indicated that PE teachers of various backgrounds exhibited measurement invariances in teaching stage and graduation department; however, only the equivalent comparison of factor configuration in terms of "gender" and "teaching location" can be regarded as equal. The factor loading comparison among groups indicated obvious differences. After the posteriori test of the scale invariance of item levels, the differences in the CM construct were correlated with the differences in the gender groups. Moreover, the differences in the IN construct were correlated with the differences in the teaching location groups. Therefore, the partial measurement invariance of the developed model was only observed in the gender and teaching location of the PE teachers.

References

Ayyildiz Durhan, T. (2019). A study on the effect of personal values on leisure motivation. *International Journal of Recreation and Sports Science, 3*(1), 25-38. https://doi.org/10.46463/ijrss.455766

Beard, J. G., & Ragheb, M. G. (1983). Measuring leisure motivation. *Journal of Leisure Research, 15*(3), 219-228. https://doi.org/10.1080/00222216.1983.11969557

Beggs, B., Kleparski, T., Elkins, D., & Hurd, A. (2014). Leisure motivation of older adults in relation to other adult life stages. *Activities, Adaptation & Aging, 38*(3), 175-187. https://doi.org/10.1080/01924788.2014.935910

Belošević, M., & Ferić, M. (2022). Contribution of frequency and structured leisure activities features to the alcohol use: Perception of adolescents from the northwest county of Croatia. *International Journal of Adolescence and Youth, 27*(1), 249-263. https://doi.org/10.1080/02673843.2022.2081088

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods and Research, 21*(2), 205-229. https://doi.org/10.1177%2F0049124192021002004

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Lawrence Erlbaum Associates.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.

Caldwell, L. L., Patrick, M. E., Smith, E. A., Palen, L., & Wegner, L. (2010). Influencing adolescent leisure motivation: Intervention effects of healthwise South Africa. *Journal of Leisure Research, 42*(2), 203-220. https://doi.org/10.1080/00222216.2010.11950202

Chen, I.-F., Wang, C.-W., Shin, K.-Y., & Wang, W.-C. (2017). A study on leisure motivation and leisure benefits of running race participants: A study of southern Taiwan road races. *NPUST Humanities and Social Sciences Research: Pedagogy, 11*(2), 38-51. https://doi.org/10.6618%2fHSSRP.2017.11(2)3

Chen, L.-L., & Chen, T.-L. (2004). The effects of college faculties' job satisfaction, leisure motivation, selected demographic variables and department chairpersons' leadership behavior on faculties' leisure participation. *Hungkuang Academic Review, 43*, 191-209. https://doi.org/10.6615/HAR.200405.43.20

Chen, W.-C. (2013). Department chairpersons' leadership behavior and faculty members'

recreational behavior among departments related to recreation, sport, and health of southern Taiwan colleges. *Journal of Sport and Recreation Management, 10*(2), 1-20. https://doi.org/10.6214%2fJSRM.201306_10(2).0001

Chen, Y.-C., Li, R.-H., & Chen, S.-H. (2013). Relationships among adolescents' leisure motivation, leisure involvement, and leisure satisfaction: A Structural equation model. *Social Indicators Research, 110*(3), 1187-1199. https://psycnet.apa.org/doi/10.1007/s11205-011-9979-2

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Chien, M.-K. (2016). The study of college students' leisure motivation, flow experience and leisure satisfaction: For example a college in southern Taiwan. *Leisure Industry Research, 14*(2), 1-11. https://doi.org/10.6746%2fLIR.201606_14(2).0001

Crandall, R. (1980). Motivations for leisure. *Journal of Leisure Research, 12*(1), 45-54. https://doi.org/10.1080/00222216.1980.11969418

Dias, M. V., Braganza, S. J., Kaur, D., & Marwah, K. (2022). Motivation: A key factor impelling student participation in sports. *Journal of Positive School Psychology*, 10172-10177.

Fehmi, C., Mahmut, G., Kursad, S., Ihsan, S., & Erdi, K. (2014). Analysis and examination of the relationship between leisure motivation and academic motivation of students of schools of physical education and sports. *Journal of Physical Education & Sport, 14* (1), 127-133. https://doi.org/10.13140/RG.2.1.3369.7126

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50. https://doi.org/10.2307/3151312

Gadiraju, P., Thomas, S., & Vanlalhruaii, C. (2022). Psychological states, self-esteem, leisure motivation and quality of life among institutionalized and non-institutionalized older adults. *Indian Journal of Gerontology, 36* (1), 68-80.

Hair, J. F. J., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Prentice-Hall.

Hsieh, C. M. (1998). *Leisure attitudes, motivation, participation, and satisfaction: Test of a model of leisure behavior* [Unpublished doctoral dissertation]. Indiana University.

Hsieh, C.-H. (2004). *The study of leisure motivation, participation, and satisfaction of college tennis players* [Unpublished master's thesis]. National Taiwan Sport College.

Huang, C.-H., Lu, P.-C., & Zeng, J.-H. (2013). A study of the relations models on leisure

motivation, free time management, and leisure satisfaction of slow pitch of the elementary school teachers. *NTU Journal of Physical Education, 25*, 35-45. https://doi.org/10.6569%2fNTUJPE.201309_(25).0004

Huang, C.-Y. (2017). Empirical study on leisure sports participation, motivation and happiness of seniors in Tainan area. *Journal of Physical Education Fu Jen Catholic University, 16*, 87-111.

Huang, J.-M., Lu, C., Lee, S.-C., & Chen, C.-Y. (2017). A study of leisure motivation and leisure benefit in Matsu tourist anglers. *Leisure Industry Research, 15*(1), 19-39. https://doi.org/10.6746%2fLIR.201703_15(1).0002

Huang, M.-L., Tang, T.-W., & Ho, C.-Y. (2012). An effect of leisure motivation and leisure lifestyle and on self-reported health of junior high school teachers. *Journal of Sport and Recreation Management, 9*(1), 73-99. https://doi.org/10.6214% 2fJSRM.201206_9(1).0005

Hwang, F.-M. (2007). *Structural equation modeling* (5th ed.). Wunan.

Isaacs, L. L., Nelson, R., & Trapp, S. (2021). Recreational therapists consider leisure motivation when evaluating virtual reality games. *Therapeutic Recreation Journal, 55* (4), 399-413. https://doi.org/10.18666/TRJ-2021-V55-I4-10976

Jia, S. (2018). Leisure motivation and satisfaction: A text mining of yoga centres, yoga consumers, and their interactions. *Sustainability, 10*(12), 4458. https://doi.org/10.3390/su10124458

Kim, J.-H., Brown, S. L., & Yang, H. (2019). Types of leisure, leisure motivation, and well-being in university students. *World Leisure Journal, 61*(1), 43-57. https://doi.org/10.1080/16078055.2018.1545691

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.

Koç, M. C., Ayar, H., Gümüş, H., & çimen, K. (2019). Leisure motivation and leisure satisfaction: A practice on open water swimming competitions. *OPUS International Journal of Society Researches, 13*(19), 1914-1929. https://doi.org/10.26466/opus.596597

Kueh, Y. C., Abdullah, N., Kuan, G., Morris, T., & Naing, N. N. (2018). Testing measurement and factor structure invariance of the physical activity and leisure motivation scale for youth across gender. *Front. Psychol., 9*, 1096. https://doi.org/10.3389/fpsyg.2018.01096

Kueh, Y. C., Bhasavanija, T., Abdullah, N., Morris, T., & Kuan, G. (2020). The validation of the Thai version of the physical activity and leisure motivation scale using confirm-

atory factor analysis. *Gazzetta Medica Italiana, 179*(6), 404-411. https://doi.org/10.23736/S0393-3660.19.04117-2

Kueh, Y. C., Kuan, G., & Morris, T. (2019). The physical activity and leisure motivation scale: A confirmatory study of the Malay language version. *International Journal of Sport and Exercise Psychology, 17*(3), 250-265. https://doi.org/10.1080/1612197X.2017.1321029

Kuo, C.-J., Hou, J.-S., Lee, C.-J., & Hsu, Y.-H. (2016). A study on the relationships among badminton players' leisure motivation, leisure involvement, and leisure satisfaction in Yunlin County. *Journal of National Pingtung University: Liberal Arts and Social Sciences, 1*, 27-60. https://doi.org/10.3390/land11081224

Lee, L.-C., Lin, M.-C., & Lee, S.-W. (2008). The differences of staffs' leisure attitude between leisure activity types, leisure motivation and leisure satisfaction: The Lee-Ming Institute of Technology case study. *Archives of University Education and Sports, 97year*, 253-259. https://doi.org/10.6695%2fAUES.200805_97.0035

Lee, W.-C. (2011). The development of leisure motivation scale for gifted students. *Psychological Testing, 58*(3), 501-530. https://doi.org/10.7108%2fPT.201109.0501

Li, C.-W., Li, H.-C., & Hsu, C.-W. (2018). The study of personal attributes differences in leisure activity motivation. *Journal of Global Sport and Recreation Management, 1*(2), 18-25. https://doi.org/10.6616/GSRM.201812_1(2).0002

Li, M.-N. (2006). *An introduction to Amos and its uses in scale development: Graphics & Basic*. Psychological Publishing Company.

Lin, J. Y., Wang, E. S., & Cheng, J. M. (2007). The relationship between extroversion and leisure motivation: Evidence from fitness center participation. *Social Behavior & Personality, 35*(10), 1317-1322. https://doi.org/10.2224/sbp.2007.35.10.1317

Lin, W.-C. (2001). *The research on tourists' behavior on holidays in leisure farms in Taiwan* [Unpublished master's thesis]. National Sun Yat-sen University.

Lin, W.-Z. (1978). Exploring the classification and scope of leisure activities. *National Sports Quarterly, 7*, 18-21.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32*(1), 53-76. https://doi.org/10.1207/s15327906mbr3201_3

Lounsbury, J. W., & Franz, C. P. G. (1990). Vacation discrepancy: A leisure motivation approach. *Psychological Reports, 66*(2), 699-702. https://doi.org/10.2466%2Fpr0.1990.66.2.699

Luo, Y., Deng, J., Pierskalla, C., Lee, J. H., & Tang, J. (2022). New ecological paradigm,

leisure motivation, and wellbeing satisfaction: A comparative analysis of recreational use of urban parks before and after the COVID-19 outbreak. *Land, 11*(8), 1224.

Murray, C., & Nakajima, I. (1999). The leisure motivation of Japanese managers: A research note on scale development. *Leisure Studies, 18*(1), 57-65. https://doi.org/10.1080/026143699375050

Padhy, M., Valli, S. K., Pienyu, R., Padiri, R. A., & Chelli, K. (2015). Leisure motivation and well-being among adolescents and young adults. *Psychological Studies, 60*(3), 314-320. https://doi.org/10.1007%2Fs12646-015-0327-5

Qiu, Y., Tian, H., Zhou, W., Lin, Y., & Gao, J. (2020). Why do people commit to long distance running: Serious leisure qualities and leisure motivation of marathon runners. *Sport in Society, 23*(7), 1256-1272. https://doi.org/10.1080/17430437.2020.1720655

Reddon, J. R., Pope, G. A., Friel, J. P., & Sinha, B. K. (1996). Leisure motivation in relation to psychosocial adjustment and personality in young offender and high school samples. *Journal of Clinical Psychology, 52*(6), 679-685. https://doi.org/10.1002/(sici)1097-4679(199611)52:6%3C679::aid-jclp10%3E3.0.co;2-j

Ryan, C., & Glendon, I. (1998). Application of leisure motivation scale to tourism. *Annals of Tourism Research, 25*(1), 169-184. https://doi.org/10.1016/S0160-7383(97)00066-2

Santos-Labrador, R. M., Melero-Ventola, A. R., Cortés-Rodríguez, M., Sánchez-Barba, M., & Arroyo-Anlló, E. M. (2021). Validation of the Physical Activity and Leisure Motivation Scale in adolescent school children in Spain (PALMS-e). *Sustainability, 13*, 7714. https://doi.org/10.3390/su13147714

Sharma, S., Abraham, A., & Masih, P. (2014). Leisure motivation amongst tourists. *Indian Journal of Positive Psychology, 5*(4), 481-483.

Tabachnick, B. G., & Fidell, L. S. F. (2001). *Using multivariate statistics* (4th ed.). Allyn & Bacon.

Tinsley, H. E. A., & Tinsley, D. J. (1986). A theory of the attributes, benefits, and causes of leisure experience. *Leisure Sciences, 8*(1), 1-45. https://doi.org/10.1080/01490408609513056

Torkzadeh, G., Koufteros, X., & Pflughoeft, K. (2003). Confirmatory analysis of computer self-efficacy. *Structural Equation Modeling, 10*(2), 263-275. https://doi.org/10.1207/S15328007SEM1002_6

Tsai, F.-C., & Lin, H.-L. (2008). Talking about the phenomenon of teenagers' internet addiction from the theory of leisure behavior. *Leisure & Exercise, 7*, 129-134. https://doi.org/10.29909%2fIWCMAT.200806.0012

Tsai, F.-H. (2012). Exploring the influence of leisure motivation of senior citizens on leisure satisfaction-taking retired teachers in Kaohsiung and Pingtung area as an example. *Asia's Holistic Health and Industrial Development for Senior Citizens Journal, 2*, 15-25. https://doi.org/10.6263%2fASHLE.201212_(2).0002

Tsai, F.-H., & Lee, C.-H. (2011). The investigation of retired teachers' leisure attitude and leisure motivation in Kaohsiung and Pingtung area. *Journal of Meiho University, 30* (2), 171-181. https://doi.org/10.7067%2fJMU.201109.0171

Tsai, M.-H., & Kuo, Y.-K. (2016). The relationship among leisure motivation, leisure involvement and satisfaction of elders. *Journal of Tourism and Leisure Management, 4* (1), 119-128. https://doi.org/10.6510%2fJTLM.4(1).11

Tu, Y.-R., & Li, H.-Z. (2016). Study of tourists' leisure motivations on action intentions: Taking the Hou-Fon bikeway as an example. *Journal of Leisure & Holistic Wellness, 16*, 240-248.

Van Lankveld, W., Linskens, F., & Stolwijk, N. (2021). Motivation for physical activity: Validation of the Dutch version of the Physical Activity and Leisure Motivation Scale (PALMS). *International Journal of Environmental Research and Public Health, 18* (10), 5328. https://doi.org/10.3390/ijerph18105328

Wang, M.-S., & Chiang, T.-C. (2003). The relationship between leisure motivation and leisure participation among adolescents in Taipei City. *TPEC Press, 11*, 203-216. https://doi.org/10.6167%2fTPEC%2f2003.11.19

Wang, Y.-W., Hsien, D.-T., & Pai, C.-L. (2009). An exploration of leisure participation motivations, leisure benefits, and the relationship between participation motivations and perception of leisure benefits the secondary schools' teachers at Hsin-Chu City. *Leisure Industry Research, 7*(4), 40-58. https://doi.org/10.6746%2fLIR.200912_7(4).0004

Weybright, E. H., Palen, L.-A., Caldwell, L. L., Wegner, L., Jacobs, J. J., & Smith, E. (2020). Profiles of adolescent leisure motivation and substance use in the healthwise South Africa research trial: A person-centered approach. *Leisure Sciences, 42*(5-6), 482-501. https://doi.org/10.1080/01490400.2018.1499055

Wu, C.-H. (2015). Study on leisure motivation, leisure participation and happiness of undergraduates in Ping-Tung region. *Journal of Sport and Recreation Management, 12*(4), 89-104. https://doi.org/10.6214%2fJSRM.1204.006