Development of Marine Science Affect Scale for Junior High School Students in Taiwan: Testing for Measurement Invariance

Liang-Ting Tsai [1], Cheng-Chieh Chang [1*], Chin-Kuo Wu [1]

[1] National Taiwan Ocean University, Keelung, TAIWAN

ABSTRACT

This study constructed a marine science affect scale to understand junior high school students' emotions toward the ocean. This work comprised three stages. First, the researcher compiled factors and items associated with marine perception through an extensive literature review. Second, the compiled factors, the items of each factor, and their content validity were examined by eight experts, and a scale was constructed containing 7 items with 2 factors. Third, this was tested on a sample of 1,683 Taiwanese junior high students. The results from a series of multigroup confirmatory factor analyses supported the reliability, content and construct validity, and gender invariance of the questionnaire.

Keywords: marine science affect scale, marine education, negative affect, positive affect

INTRODUCTION

Marine education is part of science education. Many studies have focused on the development of affective questionnaires for science leaning, but few have developed affective questionnaires for understanding students' emotions toward the ocean. Although the affective dimensions of science learning are recognized as important, they have received much less attention from researchers than the cognitive dimensions, especially marine education. Thus, this study constructed a short marine science affect scale for junior high school students, and tested the reliability and validity of the questionnaire. In addition, to improve the impartiality of the questionnaire, the study examined whether the measurements of the scale varied across gender.

Emotional responses play a crucial role in learning because they influence learners' judgements and performance. Learners' emotional responses to their learning tasks, outcomes, and abilities often affect their learning performance (Cherng & Lin, 2000; Chang & Cherng, 2010). As indicated by researchers in this field, emotions can be positive or negative. Positive emotions occur when learners enjoy acquiring new knowledge and are satisfied with their learning performance, and negative emotions occur when learners are upset over their learning, are concerned about their learning abilities, or are disappointed at their learning performance. Positive and negative emotions can consequently influence peoples' decisions and judgements, and emotional responses can influence learners' learning processes and performances.

Positive and negative emotions may be considered incompatible; however, studies have indicated that experiences of positive and negative affections are distinct (Bradbum, 1969). Watson and Tellegen (1985) indicated that affections can be positive or negative, and Watson, Clark, and Tellegen (1988) reported that positive and negative affections are independent of each other. In addition, positive and negative affections are crucial assessment indicators for subjective well-being (Diener, Lucas, & Oishi, 2005). Because affections are two-dimensional, positive affections should not be considered as the reverse of negative affections; the factors that influence positive affections differ from those that influence negative affections. For example, a person who actively participates in social activities often experiences positive affections; however, this does not exclude this person from experiencing negative affections (Bradbum, 1969). Therefore, some experiences or events can enhance positive affections without influencing negative affections, whereas some experiences or events can amplify negative affections without influencing positive affections. Zautra et al. (2003) proposed a two-factor model to explain the effects of positive and negative traits, according to which positive personal traits can increase the likelihood of

Contribution of this paper to the literature

- This initial investigation demonstrated favorable psychomentric properties for the short version of a marine affect scale for a sample of Taiwanese junior high school students.
- Marine science educators can use the marine science affect scale to enhance teaching efficacy by addressing students' emotions toward the ocean, as associated with their marine literacy or marine science knowledge.
- The developed scale may enhance marine science learning both in classrooms and in family life.

positive social conditions occurring and enhance positive affections, but they do not increase the likelihood of negative social conditions occurring, nor do they influence negative affections. Similarly, negative personal traits can increase the likelihood of negative social conditions occurring and amplify negative affections, but they do not influence positive social conditions or positive affections. In other words, positive and negative personal traits differ in how they influence positive and negative social conditions and affections.

According to the aforementioned assertions, knowledge of the positive domain explains the negative domain but to an extremely limited degree, and vice versa. Therefore, the influences of positive and negative affections warrant investigation.

Many researchers have developed and discussed the psychometric properties of the positive and negative affect scale (PANAS) (Chin, 2009; Chiu, Hung, & Chou, 2013; Pires, Filgueiras, Ribas, & Santana, 2013; Ebesutani, Okamura, Higa-McMillan, & Chorpita, 2011). Chin (2009) developed an instrument for assessing students' affect toward science writing and found that the instrument had the psychometric properties of reliability and validity in a sample of Taiwanese university students. Chiu et al. (2013) also developed a physical education affect scale for college students and demonstrated that it could measure emotional response and motivation. Moreover, several short forms of the PANAS have been proposed (Ebesutani, Regan, Smith, Reise, Higa-McMillan, & Chorpita, 2012; Karim, Weisz, & Rehman, 2011; Thompson, 2007). Ebesutani et al. (2012) and Karim et al. (2011) have both developed a shortened 5-item positive affect (PA) scale and a 5-item negative affect (NA) scale. Thompson (2007) developed an international PANAS short form in English, and the cross-sample stability, internal reliability, temporal stability, cross-cultural factorial invariance, and convergent and criterion-related validities were found to be psychometrically acceptable.

Many studies have demonstrated that the PANAS scale has strong psychometric properties. In addition to the reliability and validity of the development scale, measurement invariance must be established. Assessing measurement invariance is valuable when developing a measurement or questionnaire, especially in cross-cultural or group investigations, because it allows the researcher to determine whether participants from different groups or cultures attribute the same meanings to questionnaire items (Chen, 2008; Cheung & Rensvold, 2002; Cheung & Rensvold, 2000; Milfont & Fischer, 2010; Tsai, Yang, & Chang, 2015; Tsai & Yang, 2012). Measurement invariance does not mean that all subjects are equal with respect to the variable in question, rather that they are equal with respect to the instrument in question (Tsai et al., 2015). Thus, measurement invariance is an important property developed to compare the mean level of a certain construct or trait among different groups, because interpretation of the mean differences may be problematic unless the underlying constructs are the same across groups (Wu & Yao, 2006). Therefore, the conclusions of investigations that do not demonstrate invariance are likely to be biased, with results that are consequently difficult to decipher.

The aim of this study was to construct a short marine science affect scale for junior high school students after establishing measurement invariance across gender, reliability, content validity, and construct validity.

METHODS

Participants

Three stages of complex sampling design were employed to collect the data. First, two cohorts were randomly selected from four regions of Taiwan (northern, central, southern, and eastern). In the second and third stages, random sampling was employed to select schools and students in each selected region. A total of 1,675 junior high school students (442, 430, 439, and 364 from the northern, central, southern, and eastern regions, respectively, totaling 882 boys and 793 girls) participated in this study. All participants completed the marine science affect scale in groups.

Measure

The development of the marine science affect scale comprised three steps. First, the researcher compiled factors and items associated with the PANAS scale and marine perception scale through a comprehensive literature review (e.g., Watson, Clark, & Tellegen, 1988; Crawford & Henry, 2004). Second, three focus group meetings were conducted with eight experts to revise the marine science affect scale to cater culturally and linguistically to middle-school students in Taiwan, and thus measure their emotions about oceans more accurately. The experts revised the scale and its items, which were to be rated on a 5-point Likert scale. The first draft (version) of the scale consisted of two dimensions: 7 items were related to PA about marine education and 4 items to NA. Second, the factors, items of each factor, and content validity were examined by eight experts. After the focus group discussions, the second draft (version) of the questionnaire was developed, containing 5 items related to PA ("I like to learn about the ocean," "I am satisfied when I learn about the ocean," "I like to participate in ocean-related activities," "I am satisfied when I participate in ocean-related activities," and "I like to participate in ocean-related activities with my family on holidays") and 2 items related to NA ("I feel stressed when learning about the ocean;" "I feel stressed participating in ocean-related activities").

Finally, the revised version of the marine science affect scale was tested for conduct reliability and construct validity. This final step involved a study sample comprising ninth graders from northern, central, southern, and eastern Taiwan. The questionnaire employed a 5-point Likert scale with response categories *totally disagree*, *disagree*, *indifferent*, *agree*, and *totally agree*; however, the response scoring for the NA items was reversed. The average score was calculated to represent a level of PA and NA about oceans and marine education ranging from 0 to 5. Higher values on the PA dimension corresponded to a higher degree of positive affect, whereas higher values on the NA dimension corresponded to a lower degree of negative affect.

Data Analysis

The current study developed a short marine science affect scale for junior high school students, and examined whether the scale was measurement invariant across gender. Reliability analyses were conducted to evaluate the internal consistency of the scale. Confirmatory factor analysis (CFA) was also used in this study to evaluate construct validity. A series of multigroup CFA (MGCFA) analyses using Mplus 7 (Muthén & Muthén, 1998-2012), based on the MLR (maximum likelihood estimation with robust standard errors) estimation procedure (Muthén & Muthén, 1998-2012; Tsai & Yang, 2012, 2013; Tsai, Yang, & Chang, 2015) for conducting measurement invariance across gender, were also conducted.

In the first step, the reliability of the different dimensions and total scale were recognized, and CFA was used to identify the latent variables estimated from the observed items. If the goodness-of-fit of the CFA model fit well, then the construct validity and appropriateness of hypothetical latent variables were also established (Tsai et al., 2015). In the second step, multigroup analysis was conducted to test for gender measurement invariance. This procedure was used to test for the equivalence of relationships among variables in the hypothesized model across gender. Several nested models with increasing parameter constraints were examined hierarchically. If the model with more constrained parameters could yield a good fit, this increased confidence that the hypothesized model is stable and valid. In other words, the measurement invariance of the hypothesized model across gender frequently reflected the invariance of relations among items and latent variables in the model.

In this study, three goodness-of-fit indices, including comparative fit index (CFI), root mean squared error of approximation (RMSEA), and standardized root mean square residuals (SRMR), were utilized for model assessment. The chi-square statistic indicates sensitivity to sample size (Bollen, 1989; Schermelleh-Engel, Moosbrugger, & Müller, 2003); therefore, other substitutive indices were used in this current study. According to recommendations in the literature (Beauducel & Wittmann, 2005; Fan & Sivo, 2005; Tsai et al., 2015), CFI, RMSEA, and SRMR can be used as indicators for model assessment. The chi-square statistic was also used to examine the results.

In addition, multigroup analysis was conducted to test for the invariance of model parameters across gender, and models with sequential-added constraints to the less restricted model were tested hierarchically. The fit and comparison of the nested models can be assessed by the goodness-of-fit indices or chi-square difference test (Byrne & Stewart, 2006; Satorra, 2000; Cheung & Rensvold, 2002). This study evaluated relative fit using both ΔCFI and ΔRMSEA.

Table 1. Descriptive statistics of all items

Item / Observed Variable		Male (N=888)				Female (N=795)			
		Mean	Std.	Skewness	Kurtosis	Mean	Std.	Skewness	Kurtosis
PA	PA_item1	2.55	1.295	.434	-.793	2.69	1.202	.300	-.648
	PA_item2	2.54	1.320	.498	-.776	2.59	1.221	.393	-.623
	PA_item3	2.74	1.278	.199	-.862	2.80	1.203	.122	-.706
	PA_item4	2.56	1.323	.429	-.865	2.66	1.246	.316	-.723
	PA_item5	2.88	1.359	.111	-1.106	2.83	1.273	.091	-.910
NA	NA_item1	2.58	1.357	.408	-.913	2.57	1.308	.412	-.806
	NA_item2	2.55	1.366	.468	-.898	2.49	1.349	.515	-.829

Note: Std. = Standard Deviation

Table 2. Cronbach's alpha analysis

Item / Observed Variable	Male (N=888)	Female (N=795)	Total sample (N=1,683)
PA	.880	.905	.889
NA	.890	.893	.888
Total scale	.875	.908	.890

RESULTS

Descriptive Analysis

Descriptive statistics, including the mean, standard deviation, skewness, and kurtosis for each item of the two dimensions by gender, are listed in **Table 1**. The means for each item were generally in the range of 2.55–2.88 for males and 2.49–2.83 for females. The means for each item do not indicate statistically significant differences between female and male students ($p > .05$). The values of skewness for the males and females ranged from 0.11 to 0.498 and 0.091 to 0.515, respectively, and the corresponding values of kurtosis ranged from −1.106 to −0.776 and −0.910 to −0.623. No items showed absolute values of skewness or kurtosis greater than the cutoffs of 3 or 8 recommend by Kline (2005) and Tsai et al. (2015), respectively. This indicates that the values fell within acceptable ranges of normal distribution.

Reliability

The reliability of two factors and that of the total scale were analyzed. The reliability was evaluated by Cronbach's alpha coefficient. The Cronbach's alpha of the two dimensions and total scale are .880, .890, and .875 for the male sample, and .905, .893, and .908 for the female sample (**Table 2**). The female sample exhibited higher reliability for the two dimensions and total scale than did the male sample. The Cronbach's alpha of the two dimensions and total scale are .889, .888, and .890 for the total sample. Analytical results indicated that the developed questionnaire exhibited appropriate internal consistency and acceptable reliability.

Factorial Validity

The hypothesized model is shown in **Figure 1**. The hypothesized model was identified and adequately fit the data. The goodness-of-fit indices were CFI = .985; TLI =.971; RMSEA = .080; SRMR = .028. All the estimated parameters shown in **Figure 1** are statistically significant ($p < .05$). CFI equal to or greater than .97 is preferred for a good model–data fit, and CFI greater than .90 is an acceptable lower bound (Schermelleh-Engel et al., 2003). The values of .08 for RMSEA and .05 for SRMR are also used as the upper bound for good model–data fit (Kline, 2005). These results indicate that the goodness-of-fit indices were all acceptable. This also indicated that the developed questionnaire had two dimensions and exhibited construct validity.
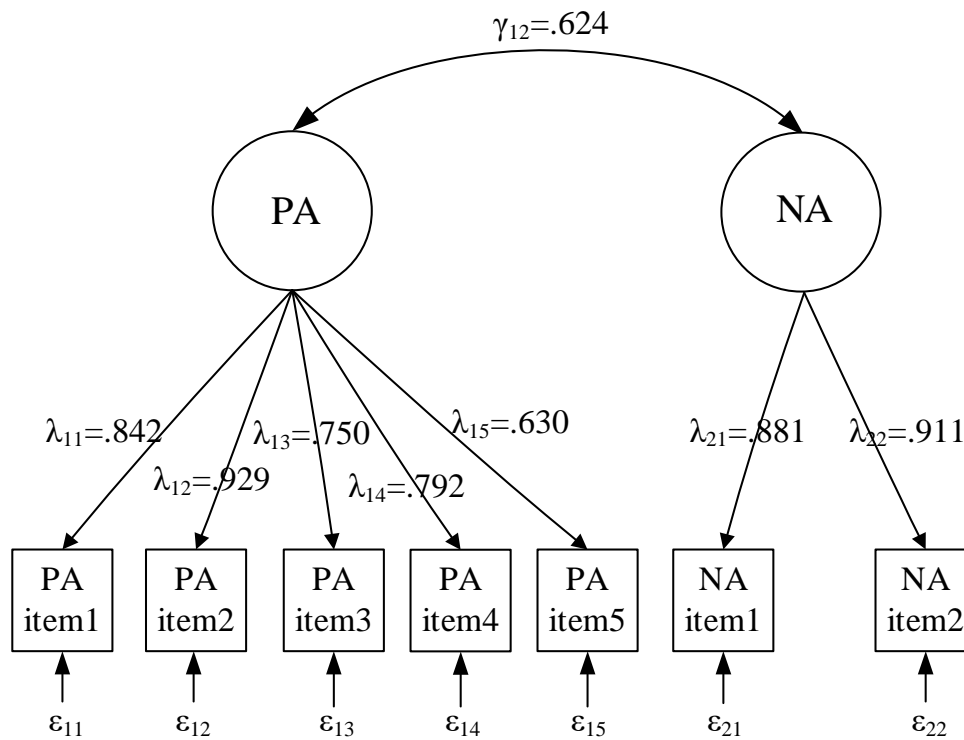
Figure 1. Hypothesized model of latent variables
Note: PA = positive affect; NA = negative affect

Measurement Invariance across Gender

To test whether the multiple-group analysis is valid and the invariance equivalent statistically, measurement invariance was tested through a series of nested models detection processes across groups (Byrne, 2008; Byrne & Stewart, 2006; Cheung & Rensvold, 2002; Tsai & Yang, 2012; Tsai et al., 2015). The first level involved the configural invariance (Model 1), also called the baseline model, with the least restrictions of the parameters. It was specified that each group had the same structure and pattern of estimated parameters. The second level (Model 2) tested the model with the equality constraints on the factor loadings across gender. The third level (Model 3) assessed whether the intercepts across gender were the same. Model 4 imposed constraints to test the invariance of disturbance of latent variables across gender. Finally, Model 5 had the most restriction; it added the residual variance equality constraints of the measured variables across gender (Byrne, 2008; Byrne & Stewart, 2006; Cheung & Rensvold, 2002). The models significantly differ if ΔCFI is greater than .01 (Cheung & Rensvold, 2002), or if ΔRMSEA is greater than .007 (Meade, Johnson, & Braddy, 2008). In other words, when the criteria of ΔCFI and ΔRMSEA are met, all specified constraints for parameters in the more restricted model are equal across gender.

Multigroup analyses for students of different genders were conducted sequentially with more additional constraints. The results of the goodness-of-fit indices are presented in **Table 3**. First, the configural invariance (Model 1) showed a good model–data fit across gender. Second, the goodness-of-fit for the invariance of factor loadings across gender (Model 2) indicated that it had an acceptable fit to the data. The ΔCFI and ΔRMSEA between Model 2 and Model 1 were less than the criteria; therefore, the invariance of factor loadings of measured variables was acceptable for both groups. Third, the goodness-of-fit of Model 3 was similarly acceptable, and ΔCFI and ΔRMSEA were also less than the criteria, indicating that the invariance of intercepts of measured variables was acceptable across gender. Subsequently, Model 5 constrained the disturbances of the latent variables to be equal across gender. The goodness-of-fit indices revealed that Model 5 had a good model–data fit and the changes in the goodness-of-fit of CFI and RMSEA were respectively less than the criteria. Finally, Model 5 constrained the invariance of the residual variance of the measured variables across gender. The goodness-of-fit indices showed that Model 5 fit the data well, and that the ΔCFI and ΔRMSEA were still less than the criteria. The results concluded that invariance of the residual variance of the measured variables across gender was supported.

Table 3. Fit indices for multigroup analysis across gender

Model	χ²	*df*	CFI	TLI	NFI	RMSEA	SRMR	Model Comparison	ΔCFI	ΔRMSEA
Model 1 Configural invariance (unconstrained, baseline model)	136.64	22	.985	.971	.982	.056	.034	-	-	-
Model 2 Invariance of factor loadings of measured variables (measurement weights)	151.34	27	.984	.975	.980	.052	.037	2 vs. 1	0.001	0.004
Model 3 Invariance of intercepts of measured variables (measurement intercepts)	166.73	34	.983	.979	.978	.048	.036	3 vs. 2	0.001	0.004
Model 4 Invariance of disturbances of latent variables (structural covariance)	187.62	37	.980	.978	.976	.048	.045	4 vs. 3	0.003	0.000
Model 5 Invariance of residuals variance of measured variables (measurement residuals)	247.87	46	.972	.985	.967	.052	.050	5 vs. 4	0.008	0.004

Note: Female group *N* = 795, male group *N* = 888; df = degree of freedom, CFI = comparative fit index, TLI = Tucker-Lewis index, NFI = normed fit index, RMSEA = root mean squared error of approximation, SRMR = standardized root mean square residuals.

Table 4. Parameter estimates for the different groups

Parameter	Male		Female		Total sample	
	Standardized estimates	Standard Deviation	Standardized estimates	Standard Deviation	Standardized estimates	Standard Deviation
λ_{11}	0.852	.011	0.850	.011	.842	.009
λ_{12}	0.910	.009	0.910	.009	.929	.007
λ_{13}	0.739	.017	0.776	.017	.750	.014
λ_{14}	0.785	.013	0.931	.013	.792	.011
λ_{15}	0.574	.019	0.607	.021	.630	.017
λ_{21}	0.872	.014	0.885	.013	.881	.012
λ_{21}	0.914	.014	0.913	.013	.911	.011
γ_{12}	0.582	.026	0.700	.022	.624	.018

CONCLUSION

This initial investigation demonstrated favorable psychometric properties for the short version of a marine affect scale for a sample of Taiwanese junior high school students. NA and PA dimensions as well as the overall scale evinced high internal consistency. The content validity of the scale was confirmed by eight experts who examined the factors and items of each factor, and CFA supported that both factors in the structure of the scale had construct validity. Lastly, MGCFA confirmed the measurement invariance across gender for the sample.

The finalized scale comprised constructs for positive and negative emotions. The construct of positive emotions encompassed five items, which were: "I like to learn about the ocean," "I am satisfied when I learn about the ocean," "I like to participate in ocean-related activities," "I am satisfied when I participate in ocean-related activities," and "I like to participate in ocean-related activities with my family on holidays." Key words in the items such as "like" and "satisfied" indicated positive emotions, and consequently the participants' liking for and interest in marine knowledge. Moreover, the participants' willingness to partake in marine activities or actual experiences of these activities were associated with mental relaxation and led to positive emotions that were articulated by the items in the construct. This finding agrees with both positive psychology and the argument of Cherng and Lin (2000) that learners with positive emotions are fond of and satisfied with what they learn, enjoy learning it, and feel happy, excited, and satisfied about their learning outcomes. It also corresponded with the view shared by Light (2003) and Pringle (2010) that positive emotional experiences strengthen learner willingness to participate in further activities. Female participants in this study had a higher average score in the construct of positive emotions than male participants, which corresponded with the findings of Tseng (2007). The construct of negative emotions consisted of two items "I feel stressed by learning about the ocean" and "I feel stressed when participating in ocean-related

activities." Male participants had a higher average score in this construct, suggesting that male students exhibit stronger negative emotions toward oceans than do their female counterparts. The average score was higher for positive rather than negative emotions, probably because acquiring marine knowledge and participating in marine activities, in contrast to attending lessons in math or literacy, allows the participants to unwind and enhance their interpersonal interactions, thereby experiencing more positive than negative emotions.

Marine science educators can use the marine science affect scale to enhance teaching efficacy by addressing students' emotions toward the ocean, as associated with their marine literacy or marine science knowledge. We also invited parents to use this instrument to quantitatively test their children's emotions toward the ocean so that they can suitably arrange marine leisure activities during vacations. Therefore, the developed scale may enhance marine science learning both in classrooms and in family life.

Although the present analyses provide evidence that the shortened Chinese version of the marine science affect scale could be used for Taiwan junior high students, this research has several limitations. First, all participants were Taiwanese ninth graders from the northern, central, southern, and eastern regions, but these were not a representative sample of Taiwanese junior high school students, limiting the generalizability of the results. Second, this study focused on the psychometric properties of the developed scale, and students were not separated into different groups (for example, those from coastal versus noncoastal states) for comparison. Third, indicators of sample characteristics which could influence students' positive and negative affects toward marine education were not measured, and thus could not be included in the analysis models to investigate the relationship between characteristics and scale scores. Despite these limitations, this study provides preliminary evidence that the marine science affect scale has adequate internal consistency, content validity, construct validity, and measurement invariance across gender. The scale appears sufficient to measure Taiwanese junior high school students' positive and negative affect toward marine science.

ACKNOWLEDGEMENT

REFERENCES

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*(1), 41-75.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: JohnWiley & Sons.

Bradburn, N. M. (1969). *The structure of psychological well-being*. Chicago: Aldine.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, *20*, 872-882.

Byrne, B. M., & Stewart, S. M. (2006). The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*(2), 287-321.

Chang, H. C., & Cherng, B. L. (2010). The interaction effects between action control orientation and success/failure conditions on cognition, motivation, and emotion. *Bulletin of Educational Psychology*, *41*(3), 605-634。

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*, 1005-1018.

Cherng, B. L., & Lin, C. S. (2000). The effects of action control training course. *Bulletin of Educational Psychology, 31*(2), 1-21.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, *31*, 188-213.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Chin, C. C. (2009). Developing an instrument for assessing college students' affection toward science writing. *Curriculum & Instruction Quarterly*, *12*(4), 113-140.

Chiu, Y. C., Hung, S. C., & Chou, H. S. (2013). Development of physical education class affection scale for college students. *Sports & Exercise Research*, *15*(2), 182-194.

Crawford, J. R. & Henry, J. D. (2004). The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology*, *43*, 245-265.

Diener, E., Lucas, R. E., & Oishi, S. (2005). Subjective well-being: The science of happiness and life satisfaction. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 63-73). New York: Oxford University Press

Ebesutani, C., Okamura, K., Higa-McMillan, K., & Chorpita, B. F. (2011). A psychometric analysis of the positive and negative affect schedule for children–parent version in a school sample. *Psychological Assessment*, *23*(2), 406-416.

Ebesutani, C., Regan, J., Smith, A., Reise, S., Higa-McMillan, C., & Chorpita, B. F. (2012). The 10-item positive and negative affect schedule for children, child and parent shortened versions: Application of item response theory for more efficient assessment. *Journal of Psychopathology and Behavioral Assessment*, *34*, 191-203.

Fan, X., & Sivo, S. (2005). Evaluating the sensitivity and generalizability of SEM fit indices while controlling for severity of model misspecification. *Structural Equation Modeling*, *12*(3), 343-367.

Karim, J., Weisz, R., & Rehman, S. U. (2011). International positive and negative affect schedule short-form (I-PANAS-SF): Testing for factorial invariance across cultures. *Procedia Social and Behavioral Sciences*, *15*, 2016-2022.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: The Guilford Press.

Light, R. (2003). The joy of learning: Emotion and learning in games through TGfU. *Journal of Physical Education New Zealand*, *36*(1), 93-108.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in the tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568-592.

Milfont, T. L., & Fischer, R., (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111-121.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User´s Guide*. Los Angeles, CA: Muthén & Muthén.

Pires, P., Filgueiras, A., Ribas, R., Santana, C. (2013). Positive and negative affect schedule: Psychometric properties for the Brazilizn Portuguese version. *Spanish Journal of Psychology*, *16*, 1-9.

Pringle, R. (2010). Finding pleasure in physical education: A critical examination of the educative value of positive movement affects. *Quest*, *62*(2), 119-134.

Satorra, A. (2000) Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In R. D. H. Heijmans, D. S. G. Pollock & A. Satorra (Eds.), *Innovations in multivariate statistical analyses. A festschrift for Heinz Neudecker* (pp. 233–247). London: Kluwer Academic.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23-74.

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross Cultural Psychology*, *38*(2), 227-242.

Tsai, L. T., & Yang, C. C. (2012). Improving measurement invariance assessments in survey research with missing data by novel artificial neural network. *Expert Systems with Application*. *39*, 10456-10464.

Tsai, L. T., & Yang, C. C. (2013). Evaluating latent interaction effects by model-based Sattora-Bentler scaled chi-square test and path-based Z-test. Chinese *Journal of Psychology*, *55*(2), 41-54.

Tsai, L. T., Yang, C. C., & Chang, Y. J. (2015). Gender differences in factors affecting science performance of eighth grade Taiwan students. *Asia-Pacific Education Research*, *24*(2), 445-456.

Tseng, W. C. (2007). The study of folk concepts of the good life and subjective well-being among college students. *Bulletin of Educational Psychology, 38*(4), 417-441.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070.

Wu, C. H., & Yao, G. (2006). Analysis of factorial invariance across gender in the Taiwan version of the Satisfaction with Life Scale. *Personality and Individual Differences, 40*, 1259-1268.

http://www.ejmste.com