Title	Validity and reliability of teachers' perceived beliefs about the purposes of physical education questionnaire
Author(s)	Pyun, Do Young, Wang, Chee Keng John, Kwon, Hyungil Harry and Chatzisarantis, Nikos
Source	*Asian Journal of Exercise and Sports Science, 5*(1), 1-7
Published by	Asian Council of Exercise and Sports Science

Original Article

Validity and Reliability of Teachers' Perceived Beliefs About the Purposes of Physical Education Questionnaire

Pyun, Do Young[1], Wang, Chee Keng John[1], Kwon, Hyungil Harry[2] & Chatzisarantis, Nikos[1]

[1] *National Institute of Education, Nanyang Technological University, Singapore*
[2] *Chung-Ang University, Korea*

The purpose of this study was to examine the psychometric properties of the Teachers' Perceived Beliefs about the Purposes of Physical Education Questionnaire. A total of 473 school physical education teachers in Singapore took part in the study. The sample was randomly split into two sets. An initial measurement model was assessed using a confirmatory factor analysis (CFA) based on the calibration sample (n = 235). Next, a modified measurement model was proposed based on the first CFA results and tested with the validation sample (n = 238). Finally, a two-stage factorial invariance analyses (Jöreskog & Sörbom, 1993) was employed to ensure concurrent validation for the modified model with both sample sets. The results showed that the cross validation of the modified measurement was not supported by the data of the two samples. Specifically, the factors of *being a good citizen* and *self-esteem* lacked convergent validity.
Asian Journal of Exercise & Sports Science, 2008, 5(1): 1-7

Keywords: Confirmatory factor analysis, factorial validity, validity, reliability,

Introduction

Physical Education (PE) in schools lays the foundations of physical health and lifelong physical activity participation for children and young people (Hardman & Marshall, 2000). It has very distinctive features in the school curriculum which provides opportunities for: (a) physical development; (b) integrated development of mind and body; (c) enhancement of self-esteem and self-confidence; (d) enhancement of social and cognitive development and academic achievement; and (e) development of respect of the body-self and others. In addition, a quality PE program in schools can also contribute to the main purposes of education. Talbot (1999) highlighted some of the possible outcomes of good quality PE programs, which include:

1. Teach children how to cope with failures and success, competition and cooperation;
2. Develop character, social skills, and aesthetic skills;
3. Provide skills and knowledge for leading a healthy lifestyle.

There is strong justification for the inclusion of PE in school curriculum from the perspectives of stakeholders, such as policy makers, educators, exercise leaders, and teachers. The outcomes of a quality PE will depend very much on the values and beliefs that PE teachers hold. However, the inquiry into PE teachers' perspectives of the purposes of PE in schools has not been extensively studied. Research in the educational domain showed that pupils' beliefs about what the aims of education should be were closely related to their goal perspectives and motivation (Nicholls, Patashnick, & Nolen, 1985). Nicholls and his colleagues showed that when students perceived that the role of education was to develop responsible and knowledgeable citizens, learning tended to be construed as an end in itself. This belief is positively related task orientation, where people focus on increasing their competence. When the students viewed the purpose of education was to gain status or wealth, learning tended to be construed as a means to an end. The view of the purpose of education is associated with ego orientation, in which people focus on demonstrating high normative abilities. Subsequently, causal attributions for success and failure are also consistent with the views of the purposes of education. Therefore, adolescents' theories of education are closely linked to their motivational orientation.

There is a lack of study in the values and beliefs of PE teachers and the impact of these beliefs on their teaching. Siedentop and Tannehill (2000) suggested that these beliefs and values might determine the way PE teachers delivered their lessons in class. Wang and Koh (2006) found that PE teachers could have varied beliefs in terms of the purposes of PE and these beliefs were associated with their motivation in teaching PE and autonomy/supportive behaviours in teaching. Therefore, an understanding of PE teachers' beliefs about the purposes of PE is an area worthy of study.

Adopting the Purposes of Schooling Questionnaire (Nicholls *et al.*, 1985) into the sport domain, Duda (1989) developed the Purpose of Sport Questionnaire. She conducted an exploratory factor analysis and found seven factors from the original 46 items. The seven factors are (a) high status career, (b) enhance self-esteem, (c) mastery and cooperation,

Corresponding Author
CK John Wang
Physical Education and Sports Science
National Institute of Education
Blk 5 #03-20, 1 Nanyang Walk
Singapore 637616
E-mail: john.wang@nie.edu.sg

(d) active lifestyle, (e) good citizen, (f) competitiveness, and (g) enhance social status.

Biddle and his colleagues (Biddle, Wang, Kavussanu, & Spray, 2003) conducted a systematic review and meta-analysis in the correlates of achievement goal orientations across 98 studies and 110 independent samples (N = 21,076). This review found that eleven studies conducted, examined the beliefs about the purposes of PE and sport and achievement goals.

In general, the systematic review showed that findings in the sport and physical activity domains are consistent with research in the academic domain (e.g., Nicholls, Cheung, Lauer, & Patashnick, 1989; Thorkildsen, 1988) in supporting the predictions of achievement goal theory. That is, task orientation is associated with the belief that the purpose of sport is to promote mastery and the values of effort exertion, enhance social responsibility, as well as encourage lifetime participation (e.g., White, Duda, & Keller, 1998). Ego orientation has been linked to the belief that sport is a means of enhancing one's status and recognition (e.g., Treasure & Roberts, 1994). In terms of the purposes of school PE, the findings of Papaioannou and McDonald (1993) and Walling and Duda (1995) have also been consistent with those from the competitive sport domain. Specifically, it has been found that high task-oriented students tend to believe that improvement, hard work, and cooperation are important functions of PE, whereas high ego-oriented students are more likely to believe that PE should enhance one's social status, provide them with an easy class, and teach them to be more competitive. Surprisingly, only these two studies were done in the PE context.

Recently, McNeill and Wang (2005) examined the cluster profile of 121 Singapore elite school sport players and the differences in their beliefs about the purposes of sport. This was the first study with Asian samples, and the Purpose of Sport Questionnaire did not show strong psychometric properties. In this study, the authors conducted an exploratory factor analysis using a principal component analysis and found 30 items loaded on four factors: (a) social status (10 items); (b) mastery/active lifestyle (seven items); (c) being a good citizen (eight items); and (d) competition (five items). The alpha coefficients ranged from .66 to .84 (social status, .84 mastery/active lifestyle, .74; being a good citizen, .72; and competition, .66).

In a separate study, Wang and his colleagues (Wang et al., 2008) used a 24-item Perceived Purposes of PE Questionnaire with five subscales which include social status (eight items), health and fitness (four items), self-esteem (three items), becoming a good citizen (four items), and mastery (five items). The internal consistency coefficients were satisfactory for the five subscales (social status, α = .86; health and fitness, α = .86; self-esteem, α = .69; becoming a good citizen, α = .78; mastery, α = .87). Wang and Koh (2006)'s work was the only study that examined pre-service PE teachers' perceptions of purposes of PE. A 30-item questionnaire from McNeill and Wang (2005) was used with four subscales including: (a) social status (10 items); (b) health and fitness (seven items); (c) competition (five items); and (e) becoming a good citizen (eight items). The alpha coefficients were adequate, but the factor validity was not tested (social status, α = .76; health and fitness, α = .90; competition, α = .77; becoming a good citizen, α = .72).

There are too many variations of the Perceived Purposes of PE Questionnaire and the factorial structure of the measurement tool has been inconsistent and requires attention. Poor measurement technology may hamper the identification of important constructs and links. The internal consistency coefficients of the two recent studies (McNeill & Wang, 2005; Wang *et al.*, 2008) were also not impressive. From an empirical perspective, researchers have consistently reported low internal reliability for the some subscales. There remains a need to examine the psychometric properties of the Perceived Purposes of PE Questionnaire. Therefore, the main purpose of this first study was to investigate the validity and reliability of the teachers' Perceived Purposes of PE Questionnaire, adapted from the Perceived Purposes of Sport Questionnaire (Wang *et al.*, 2008).

Methods

Participants and Procedure

Four hundred and ninety three PE teachers (n = 222 males, n = 262 females, 9 missing) from 90 primary schools and 98 secondary schools in Singapore took part in this study. In the data analysis, 20 cases deleted due to missing data. Overall, 33.2% of them were aged between 20 to 29 years, 51.1% aged between 30 to 39 years, 10.6% aged between 40 to 49 years, and 5.1% aged 50 years old and above. Permission for the data collection was sought through the principals and heads of PE departments. The teachers were told that participation in the survey was voluntary and they were free to withdraw at any time. Participants were informed that there were no right or wrong answers, assured of the confidentiality of their responses, and encouraged to ask questions if necessary. The procedure for conducting this study was cleared by the Ethical Review Committee in the university.

Measures

Teachers' perceptions of purposes of physical education were measured by the Perceived Purposes of Sport Questionnaire (PPSQ) used in Wang and Koh's (2006) study. The stem of the questions was "A very important thing that physical education lesson should do is to ...". There are five subscales including: (a) social status (eight items; e.g.,... makes me popular among my friends'); (b) health and fitness (four items; e.g., '... teach me how to keep myself fit'); (c) self-esteem (three items; e.g., '...increase my self-confidence'); (d) becoming a good citizen (four items; e.g., '... make us responsible law-abiding citizens'); and (e) mastery (five items; e.g., '... show me that success means always trying my best'). Responses were made on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree).

Data Analysis

For the preliminary test, descriptive statistics was performed to check missing and invalid values, outliers, and the assumption of the normality using SPSS 15.0. For the development of the measurement model, the data set (N = 473) was randomly split into two sets). With the first split (i.e., calibration sample, n = 235), an initial measurement model was assessed using a confirmatory factor analysis (CFA); a modified measurement model was proposed based on the results. With the second split (i.e., validation sample, n = 238), a validation test was employed to confirm the accuracy of the modified measurement model. Then, with the both sample sets, a two-stage factorial invariance analyses (Jöreskog & Sörbom, 1993) were employed to ensure concurrent validation for the modified model. All analyses in the measurement model development were performed using LISREL 8.80.

Results

Preliminary Test

The two split data (i.e., calibration and validation sets) were screened for the purposes of inspections of problematic observations and normality of measures. Descriptive statistics including mean, standard deviation, range, and frequency showed that there was no outlier or invalid data. Several missing values were detected and imputed with means of items in each construct. Skewness and kurtosis statistics of 24 indicators showed the univariate normality of the data.

Confirmatory Factor Analysis for the Initial Measurement Model

The initial measurement model included 24 indicators representing the five latent variables and was tested with a CFA using the calibration data (n = 235). For testing overall model fit, while most comparative goodness-of-fit indexes (e.g., NFI = .90; CFI = .92; IFI = .93) represented the overall adequacy of the model, absolute goodness-of-fit indexes (e.g., $\chi^2[242] = 783.61$, $p < .05$; $\chi^2/df = 3.24$; RMSEA = .098; GFI = .78) showed room for improvement (see Kelloway, 1998; Klein, 1998). For testing internal structures of the model, factor loadings and average variance extracted (AVE) scores were calculated (see Table 1). All loadings were significant at an alpha level of .05, indicating each loading was greater than twice its standard error (Anderson & Gerbing, 1988). However, 13 items, particularly in the social status, becoming a good citizen, and self-esteem constructs, were less than .707 (i.e., $R^2 = .50$), representing those items had more unique variance than common variance. Accordingly, the AVEs of the three constructs were less than .50 because the amount of common variance explained by each construct was less than the amount of variance due to measurement errors (Fornell & Larcker, 1981). According to the results from the overall and internal model fit tests, the initial measurement model needed to be modified to provide the best fit to the data.

Table 1. Factor Loadings and AVE in the Initial Model

Belief item	Loading	AVE
A very important thin the PE lesson should do is to		
Social status		
(SS1)	.33*	.33
(SS2)	.62*	
(SS3)	.75*	
(SS4)	.73*	
(SS5)	.39*	
(SS6)	.60*	
(SS7)	.46*	
(SS8)	.56*	
Becoming a good citizen		
(BGC1)	.71*	.44
(BGC2)	.71*	
(BGC3)	.61*	
(BGC4)	.60*	
Mastery		
(M1)	.67*	.56
(M2)	.84*	
(M3)	.61*	
(M4)	.82*	
(M5)	.77*	
Self-esteem		
(SE1)	.58*	.41
(SE2)	.62*	
(SE3)	.72*	
Health and fitness		
(HF1)	.89*	.66
(HF2)	.81*	
(HF3)	.68*	
(HF4)	.84*	

** Significant at an alpha level of .05.*

Two basic criteria for the model respecification were established. The researchers eliminated indicators which had their loadings less than .707 and retained at least three indicators per construct (see Bollen, 1989). Based on the criteria, five items (SS1, SS5, SS6, SS7, and SS8) in social status, one item (BGC4) in becoming a good citizen, two items (M1 and M3) in mastery, and one item (HF3) in health and fitness were removed from the initial item pool.

Confirmatory Factor Analysis for the Modified Measurement Model

After testing the initial measurement, a total number of indicators were reduced from 24 to 15. The modified measurement model with 15 indicators was reassessed by CFA. The overall goodness-of-fit indexes showed a better fit to the data: $\chi^2(80) = 277.20$, $p < .05$; $\chi^2/df = 3.47$; RMSEA = .10; GFI = .86; NFI = .93; CFI = .95; IFI = .95.

The internal model fit in terms of reliability and validity also improved. Only five indicators' loadings were less than .707, and AVEs of the five construct ranged from .41 to .71 (see Table 2). For discriminant validity, since

no correlation coefficient among the constructs exceeded .90 (see Table 3), all constructs seemed to be independent (Hair, Anderson, Tatham & Black, 1998). However, another method concerning relationships between AVEs and the squared correlations among the constructs revealed lack of the discriminant validity. Fornell and Larcker (1981) indicated that when an AVE value for a construct is smaller than squared correlations between the construct and other constructs, the construct lacks discriminant validity. The modified model test showed that self-esteem failed to be discriminant from being a good citizen, mastery, and health and fitness (see Tables 2 and 3).

The modified model was developed through two consecutive CFAs using the calibration sample. However, the results from the modified model could capitalize on the characteristics inherent within the same sample. The current findings derived from the calibration sample would not have the concurrent validity for further replications. Accordingly, the modified measurement model needed cross validation procedures using another sample set to show the robustness of the model.

Cross Validation Tests

Confirmatory factor analysis for the modified model. A replication of CFA for the modified model was conducted with the validation data (n = 238) to take a simple inspection of the parameters across the two samples. The results from the validation sample were showed in the parenthesis in Tables 2 and 3. The goodness-of-fit of the model in the validation sample was inconsistent with that in the calibration sample: $\chi^2(80) = 322.38$, $p < .05$; $\chi^2/df = 4.03$; RMSEA = .13; GFI = .82; NFI = .88; CFI = .91; IFI = .91.

Table 2. Factor Loadings, AVE, and Factorial Invariancea of Loadings in the Modified Model

Latent variable	Observed variable	Loading	Alpha	AVE	Test of invariance X2difference with the null model	p–value
Social Status	SS2	.60*(.59*)	.76(.72)	.53(.47)	$\chi^2_{difference}(1) = .16$	>.05
	SS3	.82* (.77*)			$\chi^2_{difference}(1) = .17$	>.05
	SS4	.73* (.69*)			$\chi^2_{difference}(1) = .92$	>.05
Becoming a good citizen	BGC1	.68* (.54*)	.72(.59)	.46(.33)	$\chi^2_{difference}(1) = 2.60$	>.05
	BGC2	.72* (.67*)			$\chi^2_{difference}(1) = .05$	>.05
	BGC3	.64* (.52*)			$\chi^2_{difference}(1) = 2.07$	>.05
Mastery	M2	.82* (.71*)	.86(.80)	.68(.57)	$\chi^2_{difference}(1) = .44$	>.05
	M4	.87* (.72*)			$\chi^2_{difference}(1) = .01$	>.05
	M5	.79* (.82*)			$\chi^2_{difference}(1) = .21$	>.05
Self-esteem	SE1	.57* (.59*)	.68(.66)	.41(.41)	$\chi^2_{difference}(1) = .99$	>.05
	SE2	.60* (.59*)			$\chi^2_{difference}(1) = .18$	>.05
	SE3	.74* (.72*)			$\chi^2_{difference}(1) = .16$	>.05
Health and fitness	HF1	.90* (.69*)	.88(.79)	.71(.58)	$\chi^2_{difference}(1) = 8.03$*	<.05
	HF2	.77* (.74*)			$\chi^2_{difference}(1) = .04$	>.05
	HF4	.85* (.84*)			$\chi^2_{difference}(1) = .38$	>.05

** Significant at an alpha level of .05.*
[a] The critical value of one degree of freedom is 3.84 at an alpha level of .05.
Note. The results from the validation test were shown in parentheses.

Table 3. Correlations among Latent Variables and Their Factorial Invariancea in the Modified Model

	Social status	Being a good citizen	Mastery	Self-esteem	Health and fitness
Social status	1.00				
Being a good citizen	.45*(.44*) $\chi^2_{difference}(1) = .01$	1.00			
Mastery	.01(-.02) $\chi^2_{difference}(1) = .02$	.64*(.71*) $\chi^2_{difference}(1) = 2.48$	1.00		
Self-esteem	.29*(.34*) $\chi^2_{difference}(1) = .15$	.88*(.75*) $\chi^2_{difference}(1) = 4.41$*	.88*(.75*) $\chi^2_{difference}(1) = 3.18$	1.00	
Health and fitness	.15*(.13) $\chi^2_{difference}(1) = .07$	.66*(.68*) $\chi^2_{difference}(1) = .00$	.77*(.82*) $\chi^2_{difference}(1) = .56$	.79*(.67*) $\chi^2_{difference}(1) = 1.26$	1.00

** Significant at an alpha level of .05..*
[a] The critical value of one degree of freedom is 3.84 at an alpha level of .05.
Note. The results from the validation tests were shown in parentheses.

During the internal model fit tests, factor loadings ranged from .52 to .84; eight indicators' loadings were less than .707 (see Table 2). The factor loadings seemed nearly consistent with the previous loadings observed in the calibration sample. The AVEs of the five constructs ranged from .33 to .58. The AVEs of becoming a good citizen and self-esteem were lower than .50 in the both samples and cross-validated using the two samples (see Table 2). For discriminant validity, the correlation among the construct in the validation sample seemed comparable to those in the calibration sample. All correlations in the validation sample met the .90 criterion suggested by Hair *et al.* (1988) and showed evidence of discriminant validity. Using the Fornell and Lacker's (1981) method (see Tables 2 and 3), Becoming a good citizen was not differentiated from mastery, self-esteem, and health and fitness due to its low AVE value in the validation sample. The cross validation test confirmed the lack of discriminant validity of self-esteem in the modified measurement model.

Factorial invariance test for the modified model

A two-stage of factorial invariance analyses were conducted with both the calibration and validation samples. In the first stage, the modified model developed in the calibration sample was applied to the data in the validation sample to simultaneously test whether the modified model was invariant across the two samples. For the test, two hypotheses were developed. The null hypothesis (Ho) stated that all parameters (i.e., factor loadings, factor covariance and variance, and measurement error variances) of the modified model were identical across the two samples. The alternative hypothesis HA stated that at least two parameters of the modified model were not identical across the two samples. The results from the χ^2 difference test between the Ho, $\chi^2(200) = 677.09$, p = .00, and the HA, $\chi^2(160) = 599.58$, $p = .00$ revealed that the difference of χ^2 statistics was 77.51 with 40 degrees of freedom. Based on the 55.76 critical value for _ with 40 degrees of freedom at an alpha level of .05, the cross validation of the modified measurement was not supported by the data of the two samples. Further scrutinization for the individual parameters was needed to pinpoint what was causing the variance within the model.

In the second stage, a series of invariance tests was followed to assess which parameters were non-invariant across the two samples. First, tests of the invariance of the individual factor loadings were conducted. The same Ho model that the researchers constrained all parameters to be equal across the two samples was used ($\chi^2 = 677.09$, $df = 200$). In addition, 15 HA models that the researchers constrained each single loading was not identical across the two samples were newly conducted. The HA models generated chi-square statistics ranging from 669.06 to 677.08 with df of 199. Table 2 summarizes the χ^2 differences with the null model for each loading. The tests showed that only one factor loading (i.e., HF1) was non-invariant across the samples.

Second, tests of the invariance of the individual factor correlations were conducted. Ten HA models that stated each factor correlation was not identical across the two samples were conducted for the tests. The HA models calculated chi-square statistics ranging from 672.68 to 677.09 with df of 199. Table 2 shows the _ differences with the null model ($\chi^2 = 677.09$, $df = 200$) for each correlation. The tests revealed that the _ difference test for the correlation between being a good citizen and self-esteem was rejected at an alpha level of .05.

Discussion

This study examined psychometric properties of Perceived Purposes of PE Questionnaire, which included 24 items representing five different subscales. The subscales included social status (eight items), health and fitness (four items), self-esteem (three items), becoming a good citizen (four items), and mastery (five items). Although the five subscales of the questionnaire significantly improved after the revision, the scale still showed lack of internal consistency and construct validity. The following sections discussed each sub-scale in relations to its reliability and validity and provided recommendations for future study.

Social Status

In the initial model, a total of eight items was included. Using the CFA with the calibration data set, the factor loadings were calculated. The loadings ranged from .33 to .75. Although all eight items were statistically significant, only the items, SS3 and SS4, exceeded the recommended value of .707. Due to the low factor loadings, the AVE of the initial social status construct was lower (AVE = .33) than recommended value of .50 (Hair *et al.*, 1998). The subscale of social status in the initial model lacked internal consistency and convergent validity as well. Thus, a revision was required. The revision was made primarily by eliminating problematic items out of the each scale. Out of the eight items, five items with low factor loadings were eliminated. Three items (SS2, SS3, and SS4) were retained for the next step. Although five items were eliminated, the three items still measure the concepts of 'being a winner', 'being a better person', and 'getting a better job'. After the revision, the subscale showed a significant psychometric improvement. The AVE was greater than the recommended value of .50 (Hair *et al.*, 1998). The Cronbach's alpha was .76, indicating that the subscale was internally consistent. However, the second CFA with validation data set produced mixed findings on AVE. The AVE using validation data was .47 and the factor loadings of SS2 ($\chi^2 = .59$) and SS4 ($\chi^2 = .69$) were lower than .707. Although the factor loading of item SS4 was near the cut-off value of .707, the item SSC2 consistently showed a low convergent validity in calibration and validation data as well. Future study needs to examine how this item works with the other two items to measure social status construct. One recommendation is to revise SS2 so that it can have more common variance with other two items. Whereas SS3 and SS4 are about "how to be a better person" and "how to win", SS2 is directly related to money. It seemed that some respondents did not juxtapose having more money with a better social status.

Health and Fitness

Health and fitness subscale initially included four items and all were statistically significant. All items except HF3 exceeded the recommended value of .707 and the AVE for the subscale was well over .50, which indicated a good reliability. However, since HF3 (χ^2 = .68) had a factor loading lower than .707, it was excluded for the next analyses. After the elimination, the subscale showed improvement in terms of factor loadings and AVE as well. The factor loadings ranged from .77 to .90 and the AVE was .71. The AVE of the health and fitness subscale was found to be the highest among five subscales in the calibration data analyses and validation data analyses as well. Although the factor loading of the first item (HF1) in validation data (.69) was lower than .707, it was high enough to ensure a convergent validity of the subscale. One concern of this item was that it caused factorial variance between the calibration and validation data (see Table 2). The factor loadings of the calibration data and the validation data showed a significant difference (.90 and .69) and this was rejected ($p < .05$) at an alpha level of .05. In general, all the factor loadings of the validation data were lower than calibration even though the original data set was divided randomly into two. There is no explanation available on this. A future study needs to examine this with different data sets and establish a consistency on this subject matter.

Self-esteem

All the items in the self-esteem subscale were retained in the modified model since it only had three items from the first. Originally, in the Purpose of Sport Questionnaire (Duda 1989), there were four items included. However, the second item "help us to reach the top in our jobs" was excluded because it seems to measure status rather than self-esteem. There was item cross-loading as well.

Among the three items chosen for Purposes of PE Questionnaire, two items (i.e., SE 1 and SE 2) showed lack of convergent validity across three data analyses (i.e., initial data analysis, modified model with calibration data, and modified model with validation data). Due to the low factor loadings of SE 1 and SE 2, AVEs of the subscale were all lower than recommended value of .50. In addition to the AVE, the Cronbach alphas of the subscale were .68 and .66 respectively for the calibration and validation data sets. The correlations among the items indicated that the correlation between SE2 and SE3 was the lowest (.37) when other correlations were .44s. The same pattern was found in the calibration data set. The correlation between SE2 and SE3 was .29 when the others were .33s. Since the correlation between SE1 and SE2 was one of the highest, one recommendation is that SE2 needs to be modified so that it can increase a correlation with SE3.

Becoming a Good Citizen

Becoming a good citizen subscale has been included from the original version of Duda's (1989) Purpose of Sport Questionnaire. In the initial version of the questionnaire, four items were included. At the initial model, two items (i.e., BGC3 and BGC 4) showed lower factor loadings: .61 and .60 respectively. Although they were statistically significant, the low factor loadings attributed to the low AVE of the subscale. Since the researchers attempted to keep at least three items in each construct, BGC4 was eliminated. However, in the later analyses with the calibration and validation data sets, the items still showed low convergent validity and internal consistency. In the two CFAs, all the items fell below .707 except the BGC2 in the calibration data. The AVEs were lower than .50 cut-off value. The same pattern between the calibration data and the validation data was found. The factor loadings and Cronbach's alpha were better with the calibration than validation data sets. Again, no specific explanation is available on this aspect. Future study needs to examine with different data sets to improve external validity of this subscale.

Mastery and Cooperation

Mastery and cooperation subscale was found to be the one of the two best subscales in terms of reliability and convergent validity in the questionnaire. In the initial model, three items (i.e., M2, M4, and M5) showed factor loadings greater than .707 whereas M1 and M3 were .67 and .61 respectively. The AVE of the initial model was also greater than .50. After the reduction of M1 and M3, the subscale showed better reliability and validity. In the original questionnaire, this subscale was labelled as Mastery and cooperation. Accordingly, the subscale included items that measured mastery of work (M1, M2, and M3) and team work (M4 and M5). The retention of items M2, M4 and M5 makes this subscale still able to capture the two different aspects (i.e., mastery and team work).

Conclusion

PE teachers could have varied beliefs in terms of the purposes of PE and these beliefs were associated with their motivation in teaching PE and autonomy/supportive behaviours in teaching. This study has made a case for understanding of PE teachers' beliefs about the purposes of PE. The psychometric properties of Perceived Purposes of PE Questionnaire were tested in this study and it appears that the measurement tool still showed lack of internal consistency and construct validity. The current study has highlighted the problematic items and future studies are required to improve external validity of this measurement tool.

References

Anderson JC & Gerbing DW. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin*, 103, 411-423.

Biddle SJH, Wang CKJ, Kavussanu M & Spray CM. (2003). Correlates of achievement goal orientations in physical activity: A systematic review of research. *Eur J Sport Sci*, 3, 1-20.

Bollen KA. (1989). Structural equations with latent variables. New York: John Wiley & Sons, Inc.

Duda JL. (1989). Relationship between task and ego orientations and the perceived purpose of sport among high school athletes. *J Sport & Exerc Psychol*, 11, 318-335.

Fornell C & Larcker D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *J Marketing Research*, 18, 39-50.

Hair JFJ, Anderson RE, Tatham RL & Black WC. (1998). Multivariate data analysis (5th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc..

Hardman K & Marshall JJ. (2000). The state and status of physical education in schools in international context. *Eur Physi Educ Rev*, 6(3), 203-229.

Jöreskog K & Sörbom D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Chicago: Scientific Software International.

Kelloway EK. (1998). Using LISREL for structural equation modeling: A researcher's guide. Newbury Park, CA: Sage Publication, Inc.

Klein RB. (1998). Principles and practice of structural equation modeling. New York: Guilford.

McNeill MC & Wang CKJ. (2005). Psychological profiles of elite school sports players in Singapore. *Psychol Sport & Exerc*, 6(1), 117-128.

Nicholls JG, Cheung PC, Lauer J & Patashnick M. (1989). Individual differences in academic motivation: Perceived ability, goals, beliefs, and values. *Learning & Individual Differences*, 1(1), 63-84.

Nicholls JG, Patashnick M & Nolen SB. (1985). Adolescents' theories of education. *J Educ Psychol*, 77, 683-692.

Papaioannou A & Macdonald AI. (1993). Goal perspectives and purposes of physical education as perceived by Greek adolescents. *Physic Educ Rev*, 16, 41-48.

Siedentop D & Tannehill D. (2000). Developing teaching skills in physical education. Mountain View, CA: Mayfield.

Talbot M. (1999). The case for physical education. Paper presented at the World summit on physical education Berlin, Germany.

Thorkildsen T A. (1988). Theories of education among academically able adolescents. *Contemporary Educ Psychol*, 13, 323-330.

Treasure DC & Roberts GC. (1994). Cognitive and affective concomitants of task and ego orientations during the middle school years. *J Sport & Exerc Psychol*, 16, 15-28.

Walling MD & Duda JL. (1995). Goals and their associations with beliefs about success in and perceptions of the purposes of physical education. *J Teaching in Physical Educ*, 14(2), 140-156.

Wang CKJ & Koh THM. (2006). Sport ability beliefs, self determination and beliefs about the purposes of physical education among Singaporean potential preservice teachers. *Asian J Exerc & Sports Sci*, 3(1), 25-34.

Wang CKJ, Lim BSC, Aplin NG, Chia M, McNeill M & Tan W KC. (2008). Students' Perceived Purposes of Physical Education in Singapore: Perspectives from a 2 x 2 Achievement Goals Framework. *Eur PE Rev*, 14, 51-70.

White SA, Duda JL & Keller MR. (1998). The relationship between goal orientation and perceived purposes of sport among youth sport participants. *J Sport Behavior*, 20(4), 474-483.