Measurement invariance of expectancy-value questionnaire in physical education

By: Xihe Zhu, Haichun Sun, Ang Chen, and Catherine Ennis

Zhu, X., Sun, H., Chen, A., & Ennis, C. D. (2012). Measurement invariance of expectancy-value questionnaire in physical education. *Measurement in Physical Education and Exercise Science, 16*(1), 41-54.

This is an Accepted Manuscript of an article published by Taylor & Francis in *Measurement in Physical Education and Exercise Science* on 24 January 2012, available online: http://www.tandfonline.com/10.1080/1091367X.2012.639629

Abstract:

Expectancy-Value Questionnaire (EVQ) measures student expectancy beliefs and task values of the domain content (Eccles & Wigfield, 1995). In this study the authors examine measurement invariance of EVQ in the domain of physical education between elementary and middle-school students. Participants included 811 students (3rd–5th grades) from 13 elementary and 903 (6th–8th grades) from 13 middle schools. Students completed the EVQ in their physical education classes within the same semester. A series of hierarchical steps for testing the measurement invariance were conducted based on means and covariance structures. The results suggest that the questionnaire possesses configural and metric invariance, but noninvariant item intercepts between these two groups of students ($|\Delta CFI| > .05$). EVQ can be used to measure expectancy-value constructs in physical education for both middle-school and elementary students. Yet the noninvariant item intercepts posted questions on how the measured construct difference should be interpreted.

Keywords: measurement invariance | expectancy belief | task value | physical education

Article:

Eccles and colleagues (Eccles, 1983; Eccles & Wigfield, 2002) proposed and tested the expectancy-value theory in multiple domains. According to the theory, one's achievement motivation derives from the individual's expectancy beliefs and task values about the domain. In educational research, expectancy beliefs and task values are considered to be important predictors of student academic performance and behavior choices (Eccles, 1983; Eccles & Wigfield, 1995; Wigfield & Guthrie, 1997). In physical education, students' expectancy-value motivation is also found to be associated with their performances and participation intentions (Xiang, McBride, & Bruene, 2004; Xiang, McBride, & Bruene, 2006). Therefore, expectancy-value theory is considered to be an important theoretical lens to examine student learning and motivation in physical education (Chen & Ennis, 2004; 2009).

Eccles and colleagues developed questionnaires to measure children and adolescents' expectancy beliefs and task values in different academic domains (Eccles, 1983). Particularly, the Expectancy-Value Questionnaire (EVQ; Eccles & Wigfield, 1995) was the primary measure that displayed sound psychometric properties and was commonly used in different domains. In physical education, Xiang, McBride, Guan, and Solmon (2003) had reported construct validity and sufficient reliability (.63–.87) for EVQ in elementary students. Likewise, EVQ has been reported with convincing reliability (.66–.89) in middle-school students in physical education (Zhu & Chen, 2010). Despite the sufficient internal reliability coefficients that were prevalent in these studies, it remains unknown whether EVQ possesses measurement invariance in these two different groups of students. Measurement invariance is considered an important step to ensure that the instrument can be used for cross-sectional and longitudinal research (Vandenberg & Lance, 2000). Therefore, the purpose of this study was to examine measurement invariance of EVQ (Eccles & Wigfield, 1995) in the domain of physical education between elementary and middle-school students.

Expectancy-Value Theory

The expectancy-value theory postulates that students' achievement-related choices and performances in a domain are directly influenced by their expectancy beliefs and task values (Eccles, 1983; Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). Expectancy beliefs are defined as students' beliefs about how well they will perform on upcoming tasks/activities. Unlike self-efficacy which ties to a specific task (Bandura, 1986), expectancy beliefs are conceived of as broad beliefs about one's competence in a given domain (Gao, Lee, & Harrison, 2008; Wigfield, Tonks, & Klauda, 2009). Eccles (1983) elaborated that there are four major task values attached to a certain domain: (a) attainment value, (b) intrinsic/interest value, (c) utility value, and (d) cost. Attainment value refers to the perceived importance of doing well on the task. Intrinsic value refers to the inherit enjoyment that the individual perceives while being engaged in the task. Utility value refers to the perceived usefulness of the task in relation to the individual's current and future goals and agenda. Although cost was discussed separately as the fourth component of task value (Eccles & Wigfield, 2002), it was conceptualized as the negative aspects of engaging in a task. Subsequently it was not conceptualized or directly measured through EVQ.

Wigfield (1994) proposed that young children's initial expectancy beliefs and task values are likely to be relatively independent of each other. He argued that children began to attach more value to the activities that they performed well over time in an achievement-related domain. Hence, Wigfield (1994) believed that the expectancy beliefs and task values should be positively correlated. Recently, Fredricks and Eccles (2002) reported that the correlation coefficients between students' expectancy-beliefs and task values ranged from .55 to .67 in mathematics. Physical education is presumed to be an achievement-related discipline similar as mathematics in schools (Chen & Ennis, 2009). It is therefore hypothesized that student expectancy beliefs and task values are positively correlated in this study.

Expectancy-Value Motivation in Elementary and Middle-School Physical Education

Elementary and middle-school students are the most studied groups that have been involved in expectancy-value motivation studies (Eccles, 1983; Eccles & Wigfield, 1995; Wigfield et al., 2009; Xiang et al., 2004; Xiang et al., 2006; Zhu & Chen, 2010). Middle-school students are viewed as a developmentally distinctive group from their elementary counterparts because of the dramatic physical and psychological changes that occur during adolescence (Eccles, Lord, Roeser, Barber, & Jozefowicz, 1997). Specifically at elementary level, Xiang et al. (2004) reported that student expectancy-beliefs were the major contributor to students' performance on the one-mile-run test explaining 22% of its variance. Among all the motivation sources, students' intrinsic value was the major contributor to their intention for future running participation, explaining 43% of its variance. At middle-school level, studies showed that expectancy-value motivation explained much smaller variances in student skill and knowledge performances (< 10%; Zhu, Chen, & Sun, 2008) and that it accounted for minimal variance for student performance improvement (Zhu & Chen, 2010). Subsequently, it is argued that expectancy-value motivation might predict engagement and performance, but not necessarily learning achievement in physical education (Zhu & Chen, 2010).

A number of studies also examined the motivational changes as well as compared motivational differences among students in elementary and middle-school years (Eccles, 1983; Eccles & Wigfield, 1995; Fredricks & Eccles, 2002). For example, Jacobs, Lanza, Osgood, Eccles, and Wigfield (2002) reported that students' expectancy-value motivation generally declined as they moved from 3rd through 12th grade for reading and mathematics. Xiang et al. (2006) reported that student task values in physical education declined as they advanced from 4th to 5th grade. As researchers gain interest in understanding student expectancy-value motivation differences/changes in elementary and middle-school years, studying measurement invariance of EVQ in physical education between these two groups of students becomes necessary and important for valid comparisons.

Measurement Invariance

Measurement invariance represents the equivalency of the factorial measurement and its underlying latent structure across different groups (Byrne, 2006; Little, 1997). Factorial invariance involves the psychometric properties of the measurement scale, including configural invariance, metric invariance, scalar invariance, and residual invariance. Latent structural invariance concerns latent means, variances, and covariances. Configural invariance indicates the extent to which the same subsets of items are associated with the same constructs. Metric invariance indicates item loadings are the same for both groups (weak invariance; Meredith, 1993). Scalar invariance indicates that the vectors of item intercepts are equivalent (strong factorial invariance; Meredith, 1993). Residual invariance tests whether the scale items measure the constructs with same degree of measurement error (strict invariance; Meredith, 1993). According to Byrne, Shavelson, and Muthén (1989), factorial invariance, or partial invariance (i.e., strong factorial invariance), is considered a prerequisite for latent structural invariance because strict invariance is difficult to achieve.

Based on the analysis of means and covariance structures (MACS), testing for measurement invariance encompasses a series of hierarchical steps (Byrne, 2006; Jöreskog, 1971; Little, 1997) to examine: (a) baseline models for each group independently, (b) configural invariance, (c) metric invariance, (d) scalar invariance, (e) residual invariance, and then (f) structural invariance such as latent means, variance, and covariance. The estimation of the hypothesized baseline models for both groups involves no between-group constraints; therefore, the data should be analyzed separately for each group in this step. Once the baseline models are established, the data from different groups must be analyzed simultaneously to test the equivalence of the constrained parameters (Jöreskog & Sörbom, 1996). The configural invariance refers to the requirements that the instrument needs to have the same number of factors and factor-load patterns across groups with no constraint imposed on the parameters. This step requires that the established baseline models should be estimated in the multi-group model. Subsequently, the test of metric invariance, scalar invariance, and residual invariance can be tested with corresponding parameters being constrained equal between the groups. These steps need to be completed in an increasingly stringent manner.

Although the previous studies (Xiang et al., 2003; Zhu & Chen, 2010) had provided evidence on the validity and reliability of EVQ for both elementary and middle-school students, respectively, it is not clear whether the questionnaire possesses measurement invariance among these groups. By following the guidelines from Byrne and Stewart (2006), Little (1997), and Vandenberg and Lance (2000), in this study the authors aimed to examine the measurement invariance of EVQ in the domain of physical education between elementary and middle-school students. Specifically, the researchers attempted to investigate EVQ's measurement baseline models for each group, factorial invariance, and structural invariance. By investigating the measurement invariance of EVQ, the researchers believed the findings would have relevant implications for cross-sectional comparison on student expectancy-value motivation in physical education.

METHODS

This study employed a cross-sectional design to examine measurement invariance of EVQ. Students from 13 elementary and 13 middle schools were sampled to participate in the study. The sampling plan, parent consent, and student assent forms were approved by the university's Institution Review Board (IRB) and the district's research office. Parental/guardian informed consent and student assent were obtained in all 26 participating schools prior to the data collection. Students and parents were both informed that participation in the study was completely voluntary and students may withdraw at any time.

Participants and Research Context

The participants were 811 students (3rd–5th grades) from 13 elementary and 903 (6th–8th grades) from 13 middle schools. The schools represented a stratified sample based on student academic achievement at school level and the percentage of federal Free and Reduced Meal System (FARMS %) from a large metropolitan school district (National Center for Education Statistics [NCES], 2003). The school district served approximately 137,800 K–12 students, representing a diverse population. As displayed in Table 1, the participants represent a gender-balanced, ethnically diverse sample with the average age of 9.35 ± 1.03 for elementary, and 12.27 ± .93 for middle-school students. The sample was representative of the school district population. At the time of data collection, the physical education curriculum focused on student learning in fundamental psychomotor skills, fitness, and fitness knowledge for both elementary and middle schools. Although specific instructional tasks may differ between elementary and middle-school students, their physical education curricula shared the same structure with similar content and learning goals.

TABLE 1. Descriptive Statistics of the Participants in Gender, Ethnicity, and Age

	Gender (%/N)		Ethnicity (%/N)					Age	
Group	**Male**	**Female**	**Asian**	**Black**	**Latino**	**White**	**Other**	**Mean**	**SD**
ES	49.20/399	50.80/412	12.50/101	19.10/155	19.60/159	39.20/318	9.60/78	9.35	1.03
MS	48.60/439	51.40/464	12.80/116	18.40/166	18.50/167	40.60/367	9.70/88	12.27	0.93

Note: ES = elementary student ($n_1 = 811$), MS = middle-school student ($n_2 = 903$).

Variable and Measures

Student expectancy beliefs and task values about physical education were measured using the EVQ (Eccles & Wigfield, 1995; Xiang et al., 2003), which includes 12 items forming four subscales. The EVQ is a five-point Likert scale with five items measuring expectancy beliefs and six items measuring the attainment (importance), intrinsic (interest), and utility (usefulness) values. In completing the questionnaire, students were asked to respond to the items by indicating their preference on the five-point scale attached to the item. For example, in responding to the item "How important do you think PE is for you?" the student can choose a number between 1 and 5 with 5 indicating *"very important"* and 1 indicating *"not important."* The descriptors *"very important"* and *"not important"* are printed explicitly on the EVQ to avoid confusion (Appendix A).

The EVQ has been reported with sound construct validity and internal reliability for both elementary and middle-school students in physical education. For example, Xiang et al. (2004) reported that Cronbach alphas ranged from .63 to .87 for expectancy belief and task value constructs for elementary students. For middle-school students, Zhu and Chen (2010) reported that Cronbach alphas ranged from .66 to .89 for expectancy belief and task value constructs. Theoretically, expectancy beliefs represent a broad unidimensional construct, and task values represent a second-order structure, encompassing attainment value, intrinsic value, and utility value. Figure 1 depicted the hypothesized measurement model for EVQ.

Data Collection

The trained data collectors and the researchers administered EVQ in the gymnasia during physical education class. The data collectors followed the exact step-by-step directions printed on the questionnaire packet. These directions consisted of having students complete the questions independently, reading the questions to students aloud, making sure students understand the questions, and enforcing the use of pencil and correct bubbling on the questionnaire. Students normally needed approximately 10 to15 min to complete EVQ. Once the students completed EVQ, the data collectors collected and put them into a sealed packet for data entry.

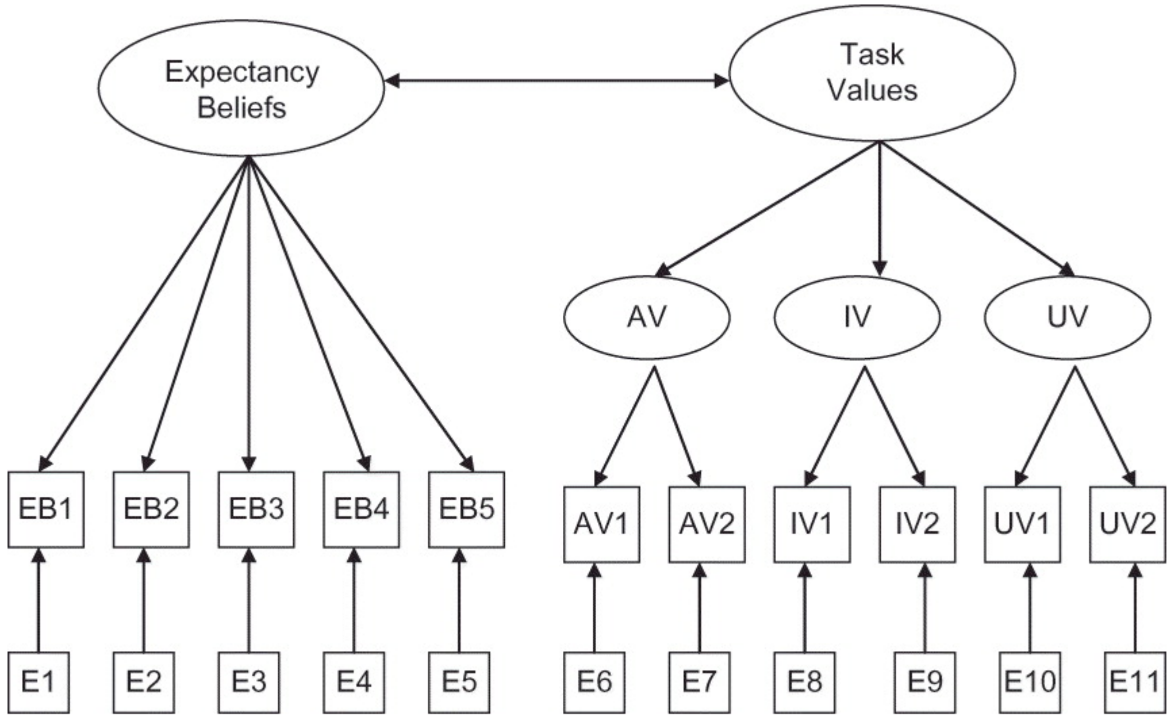

FIGURE 1. The hypothesized measurement model of Expectancy-Value Questionnaire.

Data Analyses

Three primary data analysis procedures were employed in the study. First, preliminary descriptive data analyses were conducted to examine the normality and central tendency of the data. Second, based on covariance structures, confirmatory factor analyses (CFA) were conducted to establish baseline measurement model of EVQ for each group, respectively. Eventually, based on MACS, the multi-group structural analytic methods were employed to investigate the measurement invariance of EVQ with increasingly stringent steps to release or constrain the equivalence of parameters. The data analysis was conducted using the *EQS* program (Bentler, 2005); Hu and Bentler's (1999) cutoff criteria for goodness-of-fit indices were applied. These criteria include absolute fit indices that evaluate how close the observed variance–covariance matrix is to the estimated matrix (e.g., chi-square, and Standardized Root Mean-square Residual [SRMR] < = .09), parsimony correction index that incorporates a penalty function for poor model parsimony (e.g., Root Mean Square Error of Approximation [RMSEA] < = .06), and comparative fit indices that evaluate the fit of the hypothesized model to null model (e.g., Comparative Fit Index [CFI] > = .95). The researchers used $|\Delta CFI|$ for model comparison where robust estimation approach was used (Cheung & Rensvold, 2002).

RESULTS

Preliminary Analysis

The construct means, standard deviations, and Cronbach α were reported in Table 2. On average, elementary students reported relatively high expectancy beliefs and task values (> 4.10). Middle-school students reported similar expectancy belief, yet significantly lower task values (> 3.60, < 4.00; $p < .05$). EVQ constructs showed similar Cronbach α values for both groups, ranging from .67 to .89 for elementary and from .65 to .89 for middle school. The Mardia's (1970) coefficients of multivariate kurtosis were 73.85 for elementary, and 39.77 for middle-school students, greater than 30, indicating that the multivariate normality for both datasets was violated. Consequently, the researchers used robust estimation instead of maximum likelihood to calculate the goodness-of-fit indices considering that the sample sizes were large. The scaled Satorra and Bentler (1994, 2001) chi-square statistics ($SB\ \chi^2$) were reported.

TABLE 2. Descriptive Statistics for Expectancy-Value Constructs

	Elementary School			**Middle School**		
Construct	**Mean**	**SD**	**α**	**Mean**	**SD**	**α**
Expectancy belief	4.10	.67	.81	4.08	.70	.83
Attainment value	4.18*	.81	.67	3.61*	.98	.65
Intrinsic value	4.29*	.94	.89	3.90*	1.06	.89
Utility value	4.17*	.87	.76	3.78*	1.02	.82

Note: * $p < .05$; SD = standard deviation; α = Cronbach α.

The Baseline Models

As displayed in Table 3, results from CFAs showed adequate data-model fit for both elementary and middle-school groups. However, the Lagrange multiplier test suggested that adding a covariance path between the measurement errors of item one [E1] and item two [E2] could result in significant chi-square change. Since these two items ask students similar questions (e.g., “How good are you in PE?” and “How good are you in PE in comparison with other students?”), the path was added, and both CFA models fit indices were significantly improved ($\Delta\chi^2 > 17$, $p < .05$, $|\Delta CFI| > .01$). Hence, the CFA models with the added error covariance were preserved as baseline models for measurement invariance investigation. As shown in Table 4, the item/factor loadings in the CFA models ranged from .57 to .97 for elementary students, and from .57 to .98 for middle-school students. The CFA models yielded a composite scale reliability coefficient ρ = .897 for elementary students, and ρ = .907 for middle-school students, respectively.

TABLE 3. Baseline CFA Models for Expectancy-Value Questionnaire

		Goodness-of-Fit Indices				
Group	**CFA Model**	**χ^2**	**df**	**CFI**	**RMSEA (90% CI)**	**SRMR**
ES	no EC	123.545	40	.958	.051 (.041, .061)	.043
	EC	105.167	39	.967	.046 (.035, .046)	.039
MS	no EC	133.629	40	.971	.051 (.041, .060)	.034
	EC	76.884	39	.988	.033 (.022, .043)	.026

Note: ES = elementary students (n_1 = 811), MS = middle-school students (n_2 = 903);CFA = confirmatory factor analysis; CFI = comparative fit index; RMSEA = root mean squared error of approximation; 90% CI = 90% confidence interval; SRMR = standardized root mean-square residual; EC = error covariance between E1 and E2 in *Figure* 1 (E1 and E2 are the error terms for item 1 and 2, respectively).

TABLE 4. Standardized Item and Factor Loadings

	Elementary School		Middle School	
Item/Factor	Estimate	SE	Estimate	SE
EB1	.643	.028	.723	.026
EB2	.576	.028	.647	.027
EB3	.689	.029	.576	.029
EB4	.782	.024	.816	.018
EB5	.669	.025	.723	.022
AV1	.652	.034	.750	.030
AV2	.570	.058	.568	.061
IV1	.850	.026	.870	.030
IV2	.934	.025	.911	.026
UV1	.769	.028	.826	.027
UV2	.804	.030	.845	.029
AV	.970	.028	.980	.020
IV	.772	.030	.790	.034
UV	.930	.024	.940	.025

Note: SE = standard error; EB = expectancy beliefs; AV = attainment value; IV = intrinsic value; UV = utility value.

Measurement Invariance

Based on the preserved CFA models for both groups, multi-group CFA examining configural invariance displayed a good fit model (model A, Table 5). Following Byrne and Stewart's (2006) guideline, all first-order loadings for EVQ were constrained to be equivalent between the two groups (model B), and the results showed a good fit ($SB\ \chi^2$ = 213.731, df = 85, CFI = .975, SRMR = .043, RMSEA = .042 [90% CI = .035, .049]). Subsequently, model C included second-order loadings constraining in addition to the constraints in model B. The goodness-of-fit indices showed a good fit for model C. Comparing with model A, model B and C showed no significant difference considering $|\Delta CFI| < .01$ (Cheung & Rensvold, 2002). Hence, the constraints of equal path in model C were preserved for further testing (i.e., metric invariance).

TABLE 5. Testing Measurement Invariance of Expectancy-Value Questionnaire

Model	SB χ^2	df	CFI	SRMR	RMSEA (90% CI)	Model Comparison	ΔCFI
A. Configural, no constraints	194.787	78	.977	.033	.042 (.034, .049)	–	–
B. Metric, first-order loadings	213.731	85	.975	.043	.042 (.035, .049)	B vs. A	−.002
C. Metric, all loadings	220.430	87	.974	.045	.042 (.035, .049)	C vs. B	−.001
D. Scalar, first-order intercepts	571.456	92	.922	.200	.078 (.072, .084)	D vs. C	−.052
E. Scalar, all intercepts	860.618	100	.878	.272	.094 (.088, .100)	E vs. D	−.044

Note: $p < .01$ for all standalone models; CFI = Comparative Fit Index; RMSEA = Root Mean Squared Error of Approximation; 90% CI = 90% confidence interval; SRMR = Standardized Root Mean-Square Residual.

Based on MACS, in model D, equal first-order item intercept constraints were imposed upon model C. The goodness-of-indices showed a poor model fit with substantially inflated $SRMR$, $SB\ \chi^2$, and RMSEA values (Table 5). The CFI significantly deflated comparing with model C ($|\Delta CFI|$ = .052). With all intercepts constrained equivalently between the groups, model E displayed even worse goodness-of-fit indices than model D. The researchers stopped invariance testing procedures without further imposing constraints such as residual invariance. Intercept invariance (i.e., scalar invariance) of EVQ was not supported by the results. Thus the researchers examined the Lagrange multiplier test results to identify the sources of nonequivalence. The univariate chi-square tests suggested that the equal constraint on item 7 and item 3, if released, could result in significant chi-square value changes ($\Delta\chi^2 = 28.11$, $p < .01$; $\Delta\chi^2 = 10.09$, $p < .01$). This result suggested that there was a significant difference in item (3, 7) intercepts between elementary and middle-school students. Since invariance testing involved a series of hierarchical models and each model based upon well-fit preceding models, the researchers halted further testing on structural invariance such as latent means (Byrne et al., 1989; Meredith, 1993).

DISCUSSION

The purpose of the authors in this study was to examine the measurement invariance of EVQ in the domain of physical education between elementary and middle-school students. The results were that EVQ possesses configural and metric invariance, but no scalar invariance, which stopped further testing. These findings pointed to practical implications and directions for future theoretical investigation.

Measurement Invariance

The results of this study suggested that the configural structure and the path loadings of the EVQ were preserved invariantly between elementary and middle-school students. According to Meredith (1993), the configural and metric invariance is a weak form of factorial invariance. Byrne et al. (1989) suggested that researchers should continue to further structural testing when partial invariance such as scalar invariance is achieved because strict invariance is extremely difficult to achieve. EVQ failed to obtain the strong factorial invariance with the results showing that two items possessed intercept difference between elementary and middle-school students, thus preventing the researchers from further structural testing. Latent means were not compared in this study. Nevertheless, the construct means showed significant difference in task values, but no significant difference in expectancy beliefs between elementary and middle-school students. This result appeared to support the findings that student task values generally decline as they grow older (Fredricks & Eccles, 2002; Xiang et al., 2006). Due to the noninvariant intercepts found in this study, it is difficult to fathom the extent to which the measurement of task values contributed to the group differences.

Researchers have paid a close attention to the variation of chi-square value and CFI in literature review (Vandenberg & Lance, 2000) and simulation studies (Cheung & Rensvold, 2002). In this study, chi-square and CFI were also included to evaluate the tenability of different multi-group model specifications. Yet, the substantial inflation in SRMR in testing the multi-group models (model D, E) resulted in a poor data-model fit indices, along with the other indices which were changed relatively less dramatically (e.g., CFI). It may be worthwhile to examine the variation of SRMR in future simulation studies when testing goodness-of-fit index changes in measurement invariance procedures.

Measuring Expectancy Beliefs and Task Values

The goodness-of-fit indices from the CFA model testing and the item/factor loadings (Table 4) added new evidence for the construct validity of EVQ in elementary and middle-school students. In addition, the CFA models yielded a composite scale reliability coefficient ρ = .897 for elementary students, and ρ = .907 for middle-school students, respectively. This finding complemented the Cronbach alphas (α) on the internal reliability of EVQ, which seemed to be consistent with previous studies (Xiang et al., 2003; Zhu & Chen, 2010). Taken together, the current findings were that EVQ possessed sufficient construct validity and internal reliability for measuring elementary and middle-school students' expectancy beliefs and task values in physical education.

In testing the measurement invariance of an instrument within the framework of MACS, researchers concentrate on the extent to which the factor loadings and intercepts are equivalent across groups (Byrne, 2006). When noninvariant items are identified, researchers interpret the findings using MACS analogous to item response theory (IRT, Chan, 2000; Cooke, Kosson, & Michie, 2001; Raju, Laffitte, & Byrne, 2002). IRT modeling focuses on differential item functioning such as the level of item difficulty and item discrimination which both describe the link between a test item and its underlying latent factor (Widaman & Reise, 1997). Specifically, it is deemed within MACS that the item intercept corresponds to the item difficulty parameter, and the item factor loading is analogous to the item discrimination parameter (Chan, 2000; Widaman & Reise, 1997). "The higher an item difficulty level (i.e., intercept value), the more attractive the item is in a sense that its average response level reflects its stronger endorsement (Byrne & Stewart, 2006; p. 311)." The higher a factor loading value is (i.e., item discrimination level), the less ambiguous an item is perceived to be (Chan, 2000).

Since EVQ possessed metric invariance, meaning the loadings were equivalent across the elementary and middle-school groups, the theoretical structure of expectancy beliefs and task values in physical education thus can be preserved in both groups of students. However, EVQ failed to establish its scalar invariance. Specifically, for item 7 ("Compare to math, reading, and science, how important is it for you to learn PE content?") and item 3 ("Some kids are better in one subject than in another. For example, you might be better in math than in reading. Compared to most of your other school subjects, how are you doing in PE?"), significant noninvariant intercepts were identified. Interestingly, these two items were constructed following a similar structure that posed comparison questions. Because these items were both perceived with similar ambiguity (i.e., similar loadings), elementary students reported higher intercepts for both items, representing higher endorsement than middle-school students.

The noninvariant item intercepts for item 3 and item 7 might result from two sources. First, as the students progress from elementary to middle schools, their attainment value toward sport and other academic domains decreases, as demonstrated by item 7 (Jacobs et al., 2002). Second, research has demonstrated that student expectancy beliefs for physical education tended to remain relatively stable (Xiang et al., 2006; Zhu & Chen, 2010). The noninvariant intercept for item 3 may be due to the differences in how elementary and middle-school students understand and respond to the item.

In summary, the findings suggest that the EVQ preserves its measurement structure and factorial item loadings equally well across elementary and middle-school students. In other words, for both constructs, expectancy beliefs and task values, the data from this study provided additional evidence for their construct validity in both groups of students. Hence, the questionnaire may be used to measure student expectancy-value motivation in physical education for both groups of students. Yet the noninvariant item intercepts posted questions on how the measured construct difference should be interpreted in the context.

Implications and Future Research

The findings of the authors in this study have two main practical implications. First, the results of this study added further evidence on EVQ's construct validity and internal reliability in elementary and middle-school physical education contexts. Hence, the questionnaire can be used to measure middle and elementary student expectancy beliefs and task values toward physical education. Second, when EVQ is used to measure student expectancy beliefs and task values, equal items intercepts for expectancy beliefs (item 3) and attainment value (item 7) should not be assumed to make cross-sectional comparisons. While sound comparisons might be made in student intrinsic and utility values, the differences founded through cross-sectional comparisons in expectancy beliefs and attain value may potentially result from the noninvariant item intercepts as elementary students may understand the items differently from middle-school students.

For future research, since item 3 and item 7 follow a similar wording structure, more studies are needed first to examine the sources of noninvariant item intercepts using qualitative approaches such as cognitive interview technique (Dietrich & Ehrlenspiel, 2010; Willis & Schechter, 1997). As researchers gain deeper understandings on how elementary and middle-school students understand and respond to these items, researchers will be able to revise and retest these items in these populations. Finally, this study only reported a cross-sectional examination of measurement invariance of EVQ in physical education; longitudinal studies are needed to further investigate if measurement invariance still preserves as developmental changes occur over the years.

ACKNOWLEDGEMENT

The authors wish to thank the anonymous reviewers and editors for their constructive comments.

REFERENCES

Bandura, A. 1986. *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bentler, P. M. 2005. *EQS 6 Structural equations program manual*, Encino, CA: Multivariate Software, Inc.

Byrne, B. M. 2006. *Structural equation with EQS: Basic concepts, applications, and programming*, 2nd, Mahwah, NJ: Lawrence Erlbaum Associations, Inc.

Byrne, B. M., Shavelson, R. J. and Muthén, B. 1989. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105: 456–466.

Byrne, B. M. and 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: 287–321.

Chan, D. 2000. Detection of differential item functioning on the Kirton Adaptation-Innovation Inventory using multiple-group mean and covariance structure analysis. *Multivariate Behavioral Research*, 35: 169–199.

Chen, A. and Ennis, C. D. 2004. Searching for optimal motivators: Goals, interests, and learning in physical education. *Journal of Educational Research*, 97: 329–338.

Chen, A. and Ennis, C. D. 2009. "Motivation and achievement in physical education". In *Handbook of motivation at school*, Edited by: Wentzel, K. R. and Wigfield, A. 553–574. New York, NY: Routledge.

Cheung, G. W. and Rensvold, R. B. 2002. Evaluating Goodness-of-Fit Indexes for testing measurement invariance. *Structural Equation Modeling*, 9: 233–255.

Cooke, D. J., Kosson, D. S. and Michie, C. 2001. Psychopathy and ethnicity: Structural, item, and test generalizability of the Psychopathy Checklist-Revised (PCL-R) in Caucasian and African American participants. *Psychological Assessment*, 13: 531–542.

Dietrich, H. and Ehrlenspiel, F. 2010. Cognitive interviewing: A qualitative tool for improving questionnaires in sport science. *Measurement in Physical Education and Exercise Science*, 14: 51–60.

Eccles, J. S. 1983. "Expectations, values and academic behaviors". In *Achievement and achievement motivations*, Edited by: Spence, J. T. 75–146. San Francisco, CA: W. H. Freeman.

Eccles, J. S., Lord, S. E., Roeser, R. W., Barber, B. L. and Jozefowicz, D. M. 1997. "The association of school transitions in early adolescence with developmental trajectories through high school". In *Health risks and developmental transitions during adolescence*, Edited by: Schulenberg, J., Maggs, J. L. and Hurrelmann, K. 283–320. Cambridge, UK: Cambridge University Press.

Eccles, J. S. and Wigfield, A. 1995. In the mind of the achiever: The structure of adolescents' academic achievement related-beliefs and self-perceptions. *Personality and Social Psychology Bulletin*, 21: 215–225.

Eccles, J. S. and Wigfield, A. 2002. Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53: 109–132.

Fredricks, J. A. and Eccles, J. S. 2002. Children's competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology*, 38: 519–533.

Gao, Z., Lee, A. M. and Harrison, L. 2008. Understanding students' motivation in sport and physical education: From the expectancy-value model and self-efficacy theory perspectives. *Quest*, 60: 236–254.

Hu, L. T. and Bentler, P. M. 1999. Cutoff criteria for fit indexes in covariate structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6: 1–55.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S. and Wigfield, A. 2002. Changes in children' self-competence and values: Gender and domain differences across grades one through twelve. *Child Development*, 73: 509–527.

Jöreskog, K. G. 1971. Simultaneous factor analysis in several populations. *Psychometrika*, 36: 409–426.

Jöreskog, K. G. and Sörbom, D. 1996. *LISREL 8: User's reference guide*, Chicago, IL: Scientific Software International.

Little, T. D. 1997. Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32: 53–76.

Mardia, K. V. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57: 519–530.

Meredith, W. 1993. Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58: 525–543.

National Center for Education Statistics (NCES). 2003. *Characteristics of the 100 largest public elementary and secondary school districts in the United States: 2001–02*, U.S. Department of Education: Institute of Education Sciences.

Raju, N. S., Laffitte, L. R. and Byrne, B. M. 2002. Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, 87: 517–529.

Satorra, A. and Bentler, P. M. 1994. "Corrections to test statistic and standard errors in covariance structure analysis". In *Analysis of latent variables in development research*, Edited by: Von Eye, A.and Clogg, C.C. 399–419. Newbury Park, CA: Sage.

Satorra, A. and Bentler, P. M. 2001. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66: 507–514.

Vandenberg, R. J. and Lance, C. E. 2000. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3: 4–70.

Widaman, K. F. and Reise, S. P. 1997. “Exploring the measurement invariance of psychological instruments: Applications in the substance use domain”. In *The science of prevention*, Edited by: Bryant, K. J., Windle, M. and West, S. G. 281–324. Washington, DC: American Psychological Association.

Wigfield, A. 1994. Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, 6: 49–78.

Wigfield, A. and Eccles, J. S. 2000. Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25: 68–81.

Wigfield, A. and Guthrie, J. T. 1997. Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology*, 89: 420–432.

Wigfield, A., Tonks, S. and Klauda, S. L. 2009. “Expectancy-value theory”. In *Handbook of motivation at school*, Edited by: Wentzel, K. R. and Wigfield, A. 55–75. New York, NY: Routledge.

Willis, G. B. and Schechter, S. 1997. Evaluation of cognitive interviewing techniques: Do the results generalize to the field?. *Bulletin de Methodologie Sociologique*, 55: 40–66.

Xiang, P., McBride, R., Guan, J. M. and Solmon, M. 2003. Children's motivation in elementary physical education: An expectancy-value model of achievement choice. *Research Quarterly for Sport and Exercise*, 74: 25–35.

Xiang, P., McBride, R. and Bruene, A. 2004. Fourth graders' motivation in an elementary physical education running program. *The Elementary School Journal*, 104: 253–266.

Xiang, P., McBride, R. and Bruene, A. 2006. Fourth graders' motivational changes in an elementary physical education running program. *Research Quarterly for Exercise and Sport*, 77: 195–207.

Zhu, X. and Chen, A. 2010. Adolescent expectancy-value motivation and learning: A disconnected case in physical education. *Learning and Individual Differences*, 20: 512–516.

Zhu, X., Chen, A. and Sun, H. 2008. Expectancy value, knowledge, and skill in middle school physical education [Abstract]. *Research Quarterly for Exercise and Sport*, 79(Suppl. 1): A71

APPENDIX A: Expectancy-Value Questionnaire

1. How good are you in physical education?
   Very good 5 4 3 2 1 Not good
2. If you give 5 to the best student in PE and 1 to the worst, what you give to yourself?
   Best 5 4 3 2 1 Worst
3. Some kids are better in one subject than in another. For example, you might be better in math than in reading. Compared to most of your other school subjects, how are you doing in PE?
   A lot better 5 4 3 2 1 A lot worse
4. How well do you think you are in PE?
   Very well 5 4 3 2 1 Very poorly
5. How well are you keeping yourself physically active in PE?
   Very well 5 4 3 2 1 Very poorly
6. How important do you think PE is for you?
   Not very important 1 2 3 4 5 Very important
7. Compare to math, reading, and science, how important is it for you to learn PE content?
   Not very important 1 2 3 4 5 Very important
8. In general, how fun do you think your PE classes are?
   Very boring 1 2 3 4 5 Very fun
9. How much do you like your PE classes?
   Don't like it at all 1 2 3 4 5 Like it very much
10. Some things that you learn in school help you do things better outside of school. We call this being useful. For example, learning about plants at school might help you grow a garden at home. How useful do you think the contents you learned in PE are?
    Not useful at all 1 2 3 4 5 Very useful
11. Compared to your other school subjects, how useful are the skills learned in PE?
    Not useful at all 1 2 3 4 5 Very useful