The Expectancy-Value Questionnaire in Physical Education: A Validation Study Among Norwegian Adolescents

Ove Østerlie[a]*, Audhild Løhre[a,b] , Gørill Haugan[b,c]

[a]*Department of Teacher Education, Faculty of Social and Educational Sciences, NTNU Norwegian University of Science and Technology, Trondheim, Norway;* [b]*NTNU Center for Health Promotion Research,* [c]*Department of Public Health and Nursing, Faculty of Medicine and Health, NTNU Norwegian University of Science and Technology, Trondheim, Norway*

*Corresponding author: Email: ove.osterlie@ntnu.no

Ove Østerlie, Assistant Professor
NTNU
ILU Lysholm
Postboks 8900
NO-7491 Trondheim, Norway
e-mail: ove.osterlie@ntnu.no
Phone: +47 73 41 26 18
Cell: +47 920 58 388
Østerlie is an educator and researcher in physical education. Main research areas are didactics and teaching methods in physical education with speciality in flipped learning and physical education as an inclusion arena.

Audhild Løhre, PhD, Cand Psychol, Associate Professor
NTNU
ILU Lysholm
Postboks 8900
NO-7491 Trondheim, Norway
e-mail: audhild.lohre@ntnu.no; audlo@online.no
Phone: +47 73 41 27 92
Cell: +47 918 11 362
Løhre has interdisciplinary research interests including wellbeing and health among children, adolescents and young people, the school setting, teacher education, public health and health promotion.

Gørill Haugan, PhD, RN, Professor
NTNU
Postbox 8905
7491Trondheim
Norway
e-mail: gorill.haugan@ntnu.no
Phone: +47 73 41 21 58
Cell: +47 975 03 375
Haugan is widely published in various fields such as validation and measurement, health promotion, quality of life, spirituality and wellbeing, and nurse-patient interaction among different populations (adolescents, postnatal women, nursing home residents).

The Expectancy-Value Questionnaire in Physical Education: A Validation Study Among Norwegian Adolescents

Abstract

One of the main aims of the school subject physical education (PE) is to promote a lifelong healthy lifestyle. The expectancy-value theory represents an essential theoretical perspective to examine and understand adolescents' learning and motivation in PE. Based on this theory, the Expectancy-Value Questionnaire (EVQ) measures students' expectancy-related beliefs and perceived task values related to a subject like PE. The aim of the present study was to examine the dimensionality, reliability and construct validity of the Norwegian version of the EVQ among adolescents in PE. In total, 338 students from six schools completed the EVQ in their PE classes during the spring of 2016. Explorative and confirmatory factor analyses were conducted, suggesting the four-dimensional construct of the EVQ to be superior the two-factor-model. The EVQ measurement model of adolescents' expectancy-related believes and subjective task values in PE demonstrated satisfying reliability and construct validity.

Keywords: physical education, expectancy-belief, task value, EVQ, factor analysis, dimensionality, reliability, construct validity

Introduction

One of the main aims of the school subject physical education (PE) is to promote a lifelong healthy lifestyle. This object is more or less similar all over Europe (European Commission/EACEA/Eurydice, 2013). Leading organizations investigating in public health recognize that inactivity is increasing among children and that the young population needs to be more physical active (Centers for Disease Control and Prevention [CDC], 2014; World Health Organization [WHO], 2010) as health in later life is influenced by adapted lifestyle across the life-course (WHO, 2016). School and PE are significant contributors for adolescents to acquire knowledge about health and developing abilities to take care of one's health (Bailey et al., 2009; CDC, 2014; European Commission/EACEA/Eurydice, 2013). Motivation and achievement are inherently connected (Nicholls, 1979) and adolescents' motivation for active participation in learning activities is critical for the learning outcome. It is shown that adolescents' level of physical activity in leisure time is positively correlated with PE motivational levels (Cox, Smith, & Williams, 2008). Therefore, an understanding of student motivational dynamics might perform as a vital resource in facilitating learning in PE. Adolescents who believe in their own ability of mastering various PE tasks will, in general, expose positive expectations for attainment and consequently a higher motivation and success rate (Nicholls, Cheung, Lauer, & Patashnick, 1989). Accordingly, adolescents' expectations of success or failure will influence on their motivation and participating in the PE subject at school.

Correspondingly, the expectancy-value theory is considered to be an important theoretical lens to examine and understand adolescents' learning and motivation in PE (Zhu, Sun, Chen, & Ennis, 2012). Adolescents' motivation in PE is found to be associated with their performance, engagement and intentions to partake (Xiang, McBride, & Bruene, 2004, 2006), but not necessarily with learning achievement in PE (Zhu & Chen, 2010). Informing students about the usefulness of a PE activity and how it could help them to achieve their future goals, is shown to increase persistence and performance in PE (Simons, Dewitte, & Lens, 2003).

Expectancy-Value Theory

The motivation for success is a product of an individual's perceived probability of success and the incentive value of that success (Atkinson, 1957). Likewise, the motivation to avoid failure can result from perceived probability of failure and the negative incentive value of failure. In educational research, where the expectancy-value theory is most widely applied and used, the theory is proposed, tested and found to be an important predictor of student academic and achievement performance, persistence and behavior choices (Eccles, 1983; Eccles & Wigfield, 2002). Important outcomes like engagement, continuing interest, and academic achievement are predicted by expectancies and values both theoretically (Eccles, 1983) and empirically (Nagengast et al., 2011; Trautwein et al., 2012). Relationships and the influence of significant others are important components in the expectancy-value theory (Wigfield & Tonks, 2002) . Expectancies and values are seen to be partly relationally determined (Martin & Dowson, 2009).

Expectancy-Related Beliefs (EB) represent an individual's specific belief regarding success in certain tasks, both in the short-term and long-term future (Eccles & Wigfield, 2002). It is shown that adolescents who believe in their ability to cope with schoolwork typically have positive expectations for success and, hence, high motivation and achievement (Nicholls et al., 1989). The expectancy-value theory is related to the theory of self-efficacy by Bandura (1977). However, while the self-efficacy theory ties one's beliefs about mastering to a specific task (Bong & Skaalvik, 2003) the expectancy-value theory includes beliefs about one's competence in a given domain, representing a broader area than a specific task (Gao, Lee, & Harrison, 2008; Wigfield, Tonks, & Klauda, 2009). This broader approach works well with the PE subject, acquiring a wide competence involving both students' theoretical knowledge, physical and social competence.

Subjective Task Values (STV) can be thought of as the motivation that allows an individual to respond to the question 'Do I want to do this activity and why?" (Wigfield & Cambria, 2010). These values are defined by Eccles and colleagues (1992) as task values with respect to the quality of the different tasks and how these qualities influence the individual's desire to do the task. The motivational and subjective aspects of these values are stressed; different individuals assign different values to the same activity (Wigfield et al., 2009). Expectancy-value theory breaks SVT into four subcategories: (1) Attainment Value (AV) (importance for identity or self by doing well on a given task), (2) Intrinsic Value (IV) (enjoyment or interest that results in deeper engagement and persistence), (3) Utility Value (UV) (usefulness or relevance for now and/or future) and (4) cost (loss of time, overly-high effort demands, loss of valued alternatives, or negative psychological experiences, such as stress) (Eccles, 1983) . However, cost has until now not been a prioritized topic in empirical research (Wigfield et al., 2009), with some exceptions (Battle & Wigfield, 2003; A. Chen & Liu, 2009; A. Chen, Martin, Ennis, & Sun, 2008; Zhu & Chen, 2013).

The dimensionality of STV has been argued in different ways; it has both a broad and a more specific definition (Wigfield, 1994). Some literature refer to the STV as one dimension, both theoretically (Atkinson, 1957; Feather, 1982) and empirically (Xiang, McBride, & Guan, 2004; Xiang, McBride, Guan, & Solmon, 2003), while others claim STV to include several dimensions (S. Chen, Chen, Sun, & Zhu, 2013; Eccles & Wigfield, 1995; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002; Wigfield et al., 1997). The dimensionality seems unclear especially among students in early and middle elementary school (Wigfield & Eccles, 1992; Xiang et al., 2003).

Expectancy-related beliefs and subjective task values have demonstrated a positive correlation in physical education (Gao, 2009; Xiang et al., 2003) and researchers have suggested that they develop simultaneously (A. Chen et al., 2008). *Gender* seems to affect expectancy-related beliefs and task values differently. In general, boys tend to hold higher expectancy-related beliefs compared to girls whilst the studies report no gender differences in subjective task values (Gao, 2009; Xiang, McBride, & Bruene, 2004; Xiang et al., 2006; Xiang et al., 2003). *Age* also influence the general expectancy-value motivation, with a decline across age through elementary, middle- and secondary school both in academic subjects and in PE for boys and girls (Jacobs et al., 2002; Xiang et al., 2006; Xiang, McBride, & Guan, 2004). However, a recent study has shown differences with respect to beliefs and task values. As students move from primary school to middle school and further to high school they tend to devalue PE, but their expectancy-related believes do not vary across age (Zhu et al., 2012).

Performance in PE seems to correlate with expectancy-related beliefs (Xiang, McBride, & Guan, 2004), while persistence and effort are predicted by subjective task values (Xiang, McBride, & Bruene, 2004). Children tend to value what they are good at rather than something they do not expect to succeed in (Eccles & Wigfield, 1995). Students with high expectancy-related beliefs are more likely to demonstrate motivated learning behavior and better performance. However, the fact that the learning content mediates the motivation to learn must not be overlooked (A. Chen et al., 2008). Mainly, PE is a practical subject in which the desired competence represents a mixture of practical and social skills, as well as theoretical knowledge. In Norway PE performance is assessed by means of a mark each semester. Previous research has demonstrated a positive correlation between PE competence, expectancy-related beliefs and subjective task values (A. Chen et al., 2008). Accordingly, we expected these variables to correlate positively.

In order to assess student motivation in PE a valid and reliable measurement model based on well-established theory is needed. The EVQ was developed by Eccles and Wigfield (1995) and further adapted to fit a PE setting by Xiang and colleges (2003). Zhu and colleges (Zhu & Chen, 2010; Zhu et al., 2012) made some further adjustments; this final version[1] is used in the present study. The EVQ is based on the Expectancy-Value theory (Eccles, 1983; Eccles & Wigfield, 1995), and reflects three out of the four dimensions mentioned above; (1) AV, (2) IV and (3) UV, along with the dimension EB. The EVQ measurement model comprises 11 items tapping into four dimensions. However, the dimensionality of the EVQ has shown to be unclear, studies have demonstrated both a two-factor-model (EB and STV) (e.g. Xiang, McBride, & Guan, 2004) and a four-factor model (EB, AV, IV, and UV) (e.g. Eccles & Wigfield, 1995). Still, most studies have published a four-factor-construct of the EVQ, in which subjective task values is represented by three separate constructs; AV, IV, and UV. The present study evaluates the measurement properties of the Norwegian version of the EVQ among adolescents in six Norwegian public schools.

Aims

The aim of the present study was to examine the psychometric properties of the Norwegian version of the EVQ among adolescents in secondary school. The research question was twofold: (1) How well does the two-factor model and the four-factor model of the EVQ fit to the observed data? (2) Does the EVQ questionnaire reveal good reliability and construct validity among Norwegian adolescents?

In accordance to the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Goodwin & Leech, 2003) the

---

[1] From "Measurement invariance of expectancy-value questionnaire in physical education," by Xihe Zhu, Haichun Sun, Ang Chen and Catherine Ennis, 2012, Measurement in Physical Education and Exercise Science, 16, appendix A. Copyright [2012] by the authors. Reprinted and translated with permission by Xihe Zhu (personal communication, March 12, 2018).

research question addressed evidence related to the dimensionality, reliability and construct validity, all of which considered to be interrelated measurement properties.

Dimensionality examines the extent to which the internal components of a test match the defined constructs, and is concerned with the homogeneity of the items (Netemeyer, Bearden, & Sharma, 2003). Furthermore, a scale's internal structure is relevant to its reliability, reflecting internal consistency by revealing which items are consistent with which items.

Reliability may be viewed as an instrument's consistence and relative lack of error. One type of reliability is internal consistence representing the interrelatedness among items or sets of items in a scale. Cronbach's alpha (α) and composite reliability (ρ) are reliability coefficients assessing internal consistency (Netemeyer et al., 2003) used in this study.

Construct validity refers to how well a measure actually measures the construct it is intended to measure, and is based among others things on the construct's relationships to other variables (Netemeyer et al., 2003).

In order to assess discriminant validity, the associations between expectancy-related believes, subjective task values, performance in PE, gender and age were assessed. The following hypotheses were tested: EB and STV positively correlate with each other and with performance in PE (H_1). Boys have higher EB than girls (H_2), and STV decline with age (H_3)

Methods

Participants

In total, 364 students (Year 8, 9, 10 in lower secondary school and Year 1 in upper secondary school) from six schools in three different regions in Norway were invited to participate, 338 (92.9%) of whom participated in this study. Twenty schools belonging to the three actual regions were requested to participate by an email to their respective managers. The schools involved in the study represented both rural and central communities with a normal distribution of immigration and social statuses. Gender and age distribution in the present sample were 154 girls mean age 15.25 years (SD = 1.30), and 185 boys mean age 15.01 years (SD = 1.12). The students' mark awarded for classwork the second semester of 2016 (girls: 4.45; boys: 4.48) reflected the national average for 10th grade (girls: 4.5; boys: 4.6) for the actual semester (Statistics Norway, 2016).

Measures

The adolescents' expectancy-related beliefs and subjective task values about PE were measured using a translated version of the EVQ (Zhu et al., 2012). This scale is measured by a five-point Likert scale including 11 items where item 1-5 measure expectancy-related beliefs and item 6-11 measure subjective task values. The subjective task values consists of three dimensions measuring the value of attending, intrinsic and utility value (Zhu et al., 2012) by two items each. A sample item is "How well do you think you are in PE?" where 5 indicated "*very well*" and 1 indicated "*very poorly*". In addition to the Likert scale the text "*very well*" and "*very poorly*" are printed to avoid confusion. The English version of EVQ has been reported with sufficient construct validity and internal reliability in elementary and middle school PE contexts (Xiang, McBride, & Bruene, 2004; Zhu & Chen, 2010; Zhu et al., 2012). Appendix A and Appendix B present the English (Zhu et al., 2012) and the Norwegian version of the EVQ, respectively.

Performance in PE was measured by the mark in PE for the semester during which the EVQ data were collected. Gender and date of birth was self-reported on the EVQ.

Translation of the EVQ from English to Norwegian

For use in this study the EVQ was translated from English to Norwegian using the cross-cultural adaption process (Gjersing, Caplehorn, & Clausen, 2010). Two professors in English, one native Norwegian and one English, in cooperation with the researcher (forming the "expert panel") made the translation. The process followed the suggested steps for cross-cultural adaption except one. The present study did not include a pretesting of the EVQ Norwegian version among adolescents. The expert panel discussed and evaluated the Norwegian version resulting in some vocabulary adjustments to fit a PE setting.

Data Collection

In order to assess a possible shift in the students' motivation for participation in PE after attended a 4-week intervention, data were collected at two points of time (T1 = baseline, T2 = after attending the 4-week PE course) during spring 2016. However, such changes in motivation were not focused in the present study assessing the psychometric properties of the EVQ Norwegian version. Still, including data from both T1 and T2 provided stronger evidence on the EVQ properties. The participating schools and PE teachers received detailed information from the researchers in accordance with a written instruction on how to conduct the data collection. The students filled in the EVQ anonymously in paper format at the start of a PE class. There was no time limit. All students had the opportunity to mark their answers without being observed and to ask questions if something was unclear. To minimize the adolescent's tendency to give socially desirable responses, they were asked to answer as truthfully as possible, along with an assurance that the teacher would not be able to read their responses and that their marks would not be affected by how they responded. A written consent in accordance with the procedure acknowledged by the Norwegian Center for Research Data (NSD, Project #47604) was given by their parents. The students normally needed approximately 10 minutes to complete the EVQ. The data collectors immediately put all EVQ's in a sealed package. This material was then sent to the researchers for data entry.

Statistical analysis

The data were analysed by descriptive statistics and explorative factor analysis (EFA) using IBM SPSS version 24, and CFA by means of Stata 14.1 (StataCorp, 2015). When evaluating a measurement scale, researchers face two important questions: (1) the underlying dimensionality of data, and (2) the adequacy of individual items. In these instances, EFA and CFA can provide complementary perspectives on data, giving different pieces of information (Hurley et al., 1997; Netemeyer et al., 2003). The implicit assumption underlying the use of EFA in the present study is the insecurity with respect to the dimensionality of the EVQ, which has not previously been tested in Norway among adolescents. Therefore, this study intended to gain insight into a potential factor structure of the EVQ, and provide a broad perspective on the observed data using EFA followed by the confirmation procedure by means of CFA.

Confirmatory factor analysis is a sub-model in SEM that deals specifically with measurement models (Brown, 2006), accounting for random measurement error. Thus, the psychometric properties of the scales used are more accurately derived. A high loading of an item indicates that there is much in common between the factor and the respective item (Sharma, 1996). Loadings below 0.32 are considered poor, ≥ 0.45 fair, ≥ 0.55 good, ≥ 0.63 very good, and above 0.71 excellent (Sharma, 1996).

A substantial body of research has indicated that Cronbach's alpha cannot be generally relied on as an estimator of reliability (Raykov, 2001). Thus, inter-item consistency was assessed by Cronbach's alpha (α) as well as the composite reliability (ρc) (Hair, Black, Babin, & Anderson, 2010) coefficients; and values (ρc) ≥ 0.7 are considered to be good (Bagozzi & Yi, 1988; Hair, Black, Babin, & Anderson, 2010; Mehmetoglu & Jakobsen, 2017). The present study assessed model fit adequacy by χ^2-statistics and various fit indices. In line with the "rules of thumb" given as conventional cut-off criteria (Mehmetoglu & Jakobsen, 2017) the following fit indices were used; the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMS) with values below 0.05 indicating good fit, whereas values smaller than 0.10 is interpreted as acceptable (Mehmetoglu & Jakobsen, 2017). Further, the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) with acceptable fit set at 0.90 (Mehmetoglu & Jakobsen, 2017) were used. The frequency distribution of the measurements was examined to assess deviation from normality; both skewness and kurtosis were significant, thus, the Robust Maximum Likelihood (RML) estimate procedure was applied. When analysing continuous but non-normal endogenous variables, the Satorra-Bentler corrected χ^2 should be reported (Kline, 2011; Satorra & Bentler, 1994).

Results

Descriptive Analysis

Table 1 presents the means (M), standard deviations (SD), Cronbach's α, and Pearson's correlation matrix for the EVQ scale estimated at both assessments (T1 and T2). Significant correlations in the predicted direction for EVQ towards performance, gender, and age were disclosed (Table 1).

Cronbach's alpha coefficients between .74 and .92 (Table 1) and composite reliability coefficients between .72 and 0.92 (Table 2) indicated good inter-item consistency.

Table 1. Expectancy-Value Questionnaire (EVQ). EVQ T1(T2) in relation to selected measures: Means (*M*), standard deviation (*SD*), Cronbach's alpha and correlation coefficients for EVQ and its dimensions.

EVQ	*M*	*SD*	*Cronbach's alpha*	*Correlation coefficients Pearson's* r^2 EB	STV	AV	IV	UV
Model-1 two-factor solution								
Expectation beliefs (EB)	19.54 (19.94)	4.06 (3.98)	.92 (.92)	-				
Subjective Task Values (STV)	22.13 (22.64)	5.23 (4.95)	.90 (.90)	.62** (.57**)	-			
Model-2 four-factor solution								
Expectation beliefs (EB)	19.54 (19.94)	4.06 (3.98)	.92 (.92)	-				
Attainment value (AV)	7.43 (7.72)	2.04 (1.95)	.74 (.79)	.55** (.53**)	.88** (.89**)	-		
Intrinsic value (IV)	7.48 (7.60)	2.04 (1.75)	.90 (.87)	.62** (.56**)	.89** (.85**)	.68** (.63**)	-	
Utility value (UV)	7.18 (7.35)	1.97 (1.89)	.85 (.83)	.43** (.40**)	.88** (.89**)	.65** (.70**)	.67** (.62**)	-
Performance	4.46	.70		.42** (.42**)	.28** (.29**)	.22** (.22*)	.39** (.23*)	.16 (.21*)
Gender	.55	.498		.19** (.20**)	.08 (.02)	.05 (-.05)	.11 (.09)	.01 (.02)
Age	15.12	1.21		-.03 (-.03)	-.25** (-.27**)	-.30** (-.30**)	-.18** (-.14*)	-.21** (-.22**)

Note. STV = mean score of AV+IV+UV; Performance = semester mark, spring 2016.
*$p < .05$, two-tailed.
**$p < .01$, two-tailed.

Table 2. Goodness-of-fit measures for measurement models of the Expectancy Value Questionnaire (EVQ). Confirmatory Factor Analysis for Model-1 and Model-2, at two points of assessment T1 and T2. Reliability estimates for T2 are listed in parentheses.

Fit measure	T1 N = 283 Model-1 2-Factors	T1 N = 283 Model-2 4-Factors	T2 N = 272 Model-1 2-Factors	T2 N = 272 Model-2 4-Factors
χ^2 *Satorra Bentler*	265.799	129.605	219.986	103.850
p-value	0.00001	0.00001	0.00001	0.00001
$\frac{x^2}{df}$ *Satorra Bentler*	6.18 (Df[3] = 43)	3.41 (Df = 38)	5.12 (Df = 43)	2.73 (Df = 38)
RMSEA	0.135 (CI: 0.120-0.151)	0.092 (CI: 0.075-0.110)	0.123 (CI: 0.107-0.139)	0.080 (CI: 0.062-0.098)
p-value (close fit test)	0.00001	0.0001	0.00001	0.0001
SRMR	0.067	0.052	0.064	0.048
TLI	0.88	0.94	0.898	0.96
CFI	0.90	0.96	0.92	0.97
$\rho c = \frac{(\sum\lambda)^2}{\left[(\sum\lambda)^2 + \sum(\theta)\right]}$	T1 (T2): Expectation: 0.92 (0.92) Task-Value: 0.89 (0.90)		T1 (T2): Expectation: 0.92 (0.92) Task-AV: 0.72 (0.79) Task-IV: 0.91 (0.89) Task-UV: 0.85 (0.84)	

Note: EVQ = Expectancy-Value Questionnaire measurement model; RMSEA = Root Mean Square Error of Approximation; SRMS = Standardized Root Mean Square Residual; TLI = Tucker-Lewis Index; CFI = comparative Fit Index; Df = degrees of freedom; ρc = composite reliability.

Dimensionality of the EVQ

EFA

The EVQ was assessed in the same sample on two occasions, approximately four weeks apart. Since previous studies have shown that the EVQ dimensionality is unclear, the EVQ items were subjected to EFA. The Kaiser-Meyer-Olkin measure (Tabachnick & Fidell, 2007) of sampling adequacy exceeded the recommended value of .60 (T1 = .895, T2 = .901) and Bartlett's test of Sphericity reached statistical significance ($p < .0001$), supporting the factorability of the correlation matrix for both assessments. The EVQ-factors were expected to be correlated (H_1). Thus, principal component analysis with an oblique promax rotation and Kaiser Normalization was used. Table 3 lists the loadings (cross-loadings), factors and variance explained for both models at both assessments (T1 and T2). Exploratory factor analysis revealed two factors with eigenvalue 1.0 and above. This two-factor-solution

(Model-1) disclosed factor loadings between .42 and .91 and four cross-loadings, explaining 71% of the total variance. However, several studies have shown that the EVQ comprises four factors; therefore, the data were run once more (both T1 and T2) setting the number of factors to four. This four-factor-solution (Model-2) displayed four cross-loadings, factor loadings between .33 and .89, and explained 81% of the variance. Hence, the dimensionality seemed uncertain, and we turned to CFA.

CFA

The two-factor construct: Model-1. The two-factor construct's fit to the observed data (Model-1, T1 and T2) was tested by means of CFA, showing significant t-values for all estimates ($p < .05$), factor loadings (λ) ranging between .93 and .66, with squared multiple correlations (R^2) varying from .87 to .43. The two-factor solution gave a Satorra-Bentler scaled $\chi^2(43) = 265.799$; $p < .00001$ and RMSEA = .135 $p < .00001$, SRMR = .067, CFI = .90 and TLI = .88. Accordingly, Model 1 did not reveal a god fit to our data (Table 2).

The four-factor construct: Model-2. Next, the previous published four-factor solution (framed Model-2) was tested, showing a significantly increased fit at T1 ($\chi^2(38) = 129.605$; $p < .0001$, RMSEA = .092, CI between .075 and .110, close fit $p = .04$, SRMR = .054, CFI = .96 and TLI = .94), and even a better fit at T2 ($\chi^2(38) = 103.85$; $p < 0.0001$, RMSEA = .080, CI between .075 and .110, close fit $p = .0001$, SRMR = .048, CFI = .97 and TLI = .96). Hence, in both assessments (T1 and T2) Model-2 comprising of four factors showed a significantly better fit than Model-1 (Table 2) and presented significant t-values for all estimates ($p < .05$). For the model to be significantly better, the change in χ^2-value should exceed the critical value belonging to the difference in degrees of freedom at the 5% level. The analysis confirmed a significant difference in χ^2 for Model 1 versus Model 2 for both assessments (T1: χ^2diff = 136.19, T2: χ^2diff = 116,136, critical value = 15.09 at 5 df, $p = .01$). Thus, the null-hypothesis of equal fit for Model 1 and Model 2 did not find support. Table 2 lists the estimates for Model-1 and Model-2 at both assessments (T1 and T2).

Table 3. Exploratory Factor Analysis of the Expectancy-Value Questionnaire – Rotated Component Matrix, T1 and T2. Estimates for T2 are listed in parentheses.

	T1 Model-1 (2 factors)		T1 Model-2 (4 factors)			
EVQ 11-items	EVQ-1[1]	EVQ-2[2]	EVQ-1[1]	EVQ-3[3]	EVQ-4[4]	EVQ-5[5]
EVQ1: How good are you in physical education?	.900 (-)	899)	.889 (.869)	-	-	-
EVQ2: If you give 5 to the best student in PE and 1 to the worst, what you give to yourself?	.872 (-)	(.867)	.864 (.849)	-	-	-
EVQ3: 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?	.741 (-)	(.693)	.735 (.541)	-	-	.341 (.714)
EVQ4: How well do you think you are in PE?	.858 (-)	(.905)	.853 (.878)	-	-	-
EVQ5: How well are you keeping yourself physically active in PE?	.700 (-)	.426 (.755)	.635 (.774)	.496	-	-
EVQ6: How important do you think PE is for you?	.421 (.770)	.690 (-)	.333	.597 (.442)	- (.563)	.428 (.410)
EVQ7: Compare to math, reading, and science, how important is it for you to learn PE content?	- (.742)	.682 (-)	-	- (.529)	-	.878 (.623)
EVQ8: In general, how fun do you think your PE classes are?	.351 (.732)	.772 (.361)		.805	.362 (.826)	-
EVQ9: How much do you like your PE classes?	.454 (.729)	.721 (.393)	.344	.798	- (.809)	-
EVQ10: 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?	- (.834)	.856 (-)	-	.412 (.867)	.812	-
EVQ11: Compared to your other school subjects, how useful are the skills learned in PE?	- (.836)	.826 (-)	.717	-(.819)	.871	-
Cumulative % of total variance explained	56.61 (56.37)	14.53 (15.21)	56.61 (56.37)	14.53 (15.21)	5.85 (6.57)	5.17 (4.68)
Cronbach's Alpha	.92 (.92)	.90 (.90)	.92 (.92)	.74 (.79)	.90 (.87)	.85 (.83)

Note: [1]EVQ-1 = Expectation Beliefs (EB). [2]EVQ-2 = Subjective Tasks Values (STV). [3]EVQ-3 = Attainment value (AV). [4]EVQ-4 = Intrinsic value (IV). [5]EVQ-5 = Utility value (UV). Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Values < 0.32 are suppressed. Model-1: Two components extracted based on Eigenvalue > 1. Total variance explained: 71.14% (71.57%). Rotation converged in 3 (3) iterations. Model-2: Four components extracted. Total variance explained: 82.16% (82.82%). Rotation converged in 6 (6) iterations.

The best fitting model was Model 2 assessed at T2, displaying factor loadings (λ) ranging between .93 and .66, and squared multiple correlations (R^2) varying from .87 to .43. Figure 1 displays Model 2 at T2.

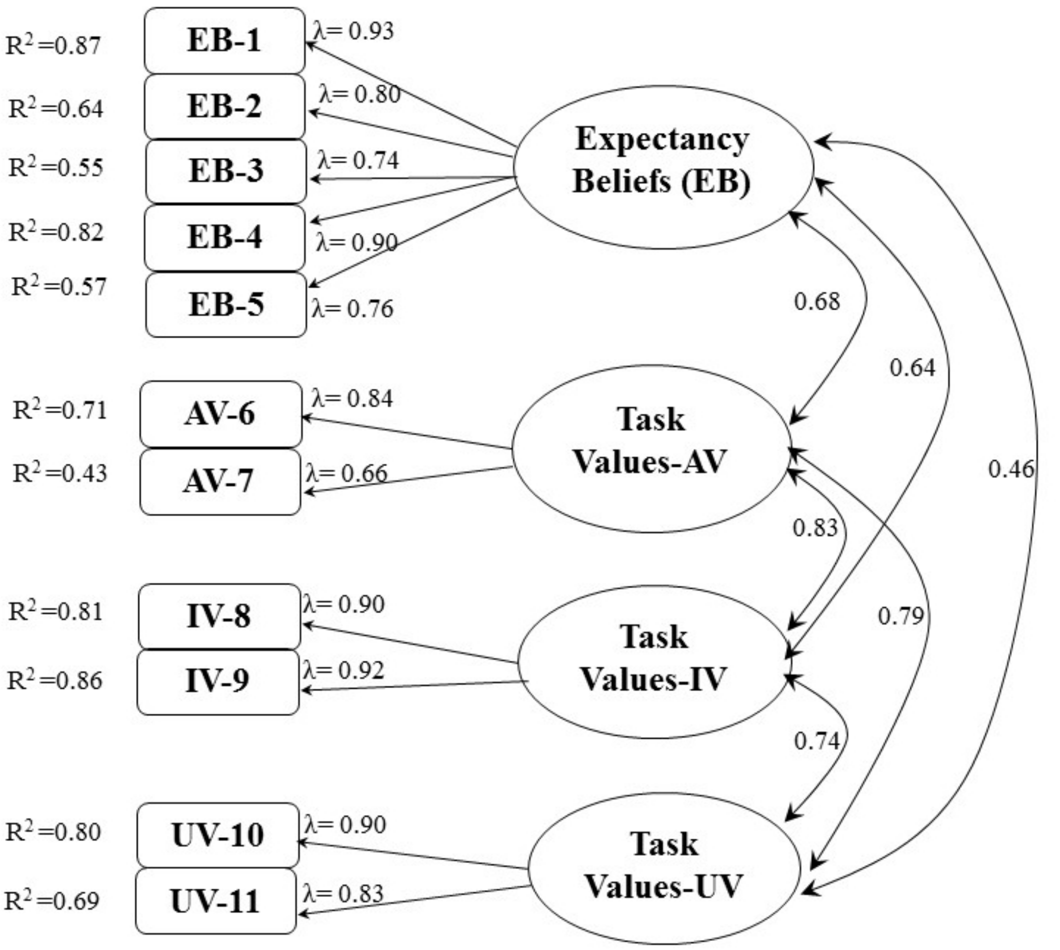
Discussion

The research question of this study addressed evidence related to dimensionality, reliability, and construct validity of the EVQ questionnaire. The aim was to assess the psychometric properties of the EVQ measure among adolescents in Norway. The research question was two-fold: (1) How well does the two-factor model and the four-factor model of the EVQ fit to the observed data? (2) Does the EVQ questionnaire reveal good reliability and construct validity among Norwegian adolescents?

Dimensionality

In accordance with previous studies (Zhu et al., 2012), the present results indicated that the four-factor model of the EVQ is psychometrically superior to the two-factor model. However, the EFA suggested two factors, whereas CFA discovered that the two-constructs-solution tailored inferior to the observed data. The first factor "Expectancy-related beliefs," including five items, performed as a separate and unique construct. This point seems indisputable. However, the three other factors (Tasks AV, IV and UV) seemed somewhat unclear, revealing highly significant factor correlations indicating that the dimensionality might be questioned; does the EVQ contain four unique constructs?

Due to the developmental characteristics in young adolescents, it is argued that few items and short instruments better generate accurate accounts (Feltz & Chase, 1998). Thus, the shorter the better. However, these three factors (Task AV, Task IV, and Task UV) comprised of only two indicator items, implying these factors to be weak constructs (Hair et al., 2010; Kline, 2011; Mehmetoglu & Jakobsen, 2017). Still, a scrutinizing of the theoretical content of these three constructs supported the plausibility of the suggested four-factor model. The AV-factor focuses on the *importance* of physical education (EVQ6 "how important do you think physical education is for you" and EVQ7 "compared to math, reading, and science, how important is it for you to learn physical education content?"). Since both of these two items direct the importance of physical education, it seems logical that they belong to each other and represent a factor. The next construct, the IV factor, consists of two items both centering on *enjoyment* in relation to physical education (how fun is PE and how much do adolescents like PE). When young people find something to be fun, it is likely that they enjoy it. Therefore, it is also reasonable that these two items constitute a factor. Looking at the fourth construct, the UV factor, both items concentrate on the *usefulness* of physical education. Therefore, based on theoretical assumptions, the four-construct solution appears to be proper.



Model-2: χ^2 =129.605, df=38, p-value <0.0001, RMSEA=0.092, SRMR=0.052, CFI=0.96, TLI=0.94

Figure 1. CFA Model-2 of the EVQ measurement, T2.

Reliability

Reliability is supported by items in each factor with highly significant standardized factor loadings–preferably greater than 0.7 (Brown, 2006; Hair et al., 2010; Kline, 2011). This was the fact for 10 out of the 11 indicators; one indicator (AV 7) revealed a loading at 0.66. Accordingly, all standardized factor loadings showed good to very good values ranging between 0.93 and 0.66. The square of a standardized loading represents how much variation in an item is explained by the latent factor and is termed the variance extracted of the item (Hair et al., 2010). As loadings fall below 0.7, they can still be considered significant, but more of the variance in the measure is error variance than explained variance. In the present study, this was the fact for only one item. As a result, Cronbach's alpha and composite reliability also revealed good values (Table 2), indicating good internal consistency: values greater than 0.7 are good (Acock, 2013; Hair et al., 2010; Mehmetoglu & Jakobsen, 2017). Hence, in this study the reliability was very well supported.

Construct Validity

Construct validation is a lengthy and ongoing process of learning more about the constructs in focus, making new predictions and then testing them. Each study that supports the theoretical construct serves to strengthen the theory (Netemeyer et al., 2003). Construct validity for the EVQ refers to the assumption that this questionnaire validly measures expectancy-related beliefs and subjective task values among adolescents. The observed data supported that expectancy-related believes and subjective task values positively correlated with each other and with performance in PE (H_1). Boys had higher expectancy-related believes than girls (H_2) and subjective task values declined with age (H_3). Hence, convergent and discriminant validity were supported by significant correlations in the predicted direction with the selected constructs of performance, gender and age. Furthermore, despite three factors in the four-factor model included only two indicators, Model 2 revealed good composite reliability coefficients (ρ) for all the four latent variables, ranging between .92 and .79 (Table 2). Moreover, the factor correlations ranged between .83 and .46. Nevertheless, some correlations showed quite high estimates (Figure 1) indicating some problems with discriminant validity. These were: (1) $r = .83$ for the relation between Task- AV and Task-IV, (2) $r = .79$ for Task AV with Task UV, and (3) $r = .74$ for Task IV with Task UV. Based on these high factor-correlations, construct validity in terms of discriminant validity can be questioned. Revealing a factor correlation of $r = 0.83$, how much of the factors' variance is unique? At the same time, the CFA clearly signified that the two-factor model corresponded poorly to the observed data, while the four-factor model exposed a good fit.

Besides, three of the four factors did only consist of two indicator items, implying these factors to be weak. From this, we suggest that including one more item in the three factors comprising only two indicators might strengthen the construct validity of the EVQ.

Taken together, the evidence supports satisfying construct validity of the EVQ construct among adolescents in Norway. Content validity is an obligation both for reliability and construct validity (Mokkink et al., 2010; Potter & Levine-Donnerstein, 1999), and is assessed by judging the relevance and the comprehensiveness of the items, both with regard to relevance for the construct to be measured as well as for the study population. In the present adolescent population, the 11 items appeared to be relevant, signified by the high factor loadings and the high R^2-values.

Strengths and limitations

The participation of 338 adolescents (response rate 93%) from six schools involving three counties in Norway signifies a strength of this study. The present sample represents a diversity of locations in urban and rural areas, reflecting the general adolescent population in Norway. Next, the students' semester marks corresponded with the national average mark for the actual semester, indicating that the present sample do not differ from the general Norwegian adolescent population at the actual ages. The PE teacher administered the data collection at the start of a PE lesson, ensuring anonymity and enough time for the students to fill in the questionnaire. This procedure, using a well known teacher in well known surroundings contributed to students feeling comfortable in the assessment situation, supporting reliable data, represents a strength of this study. The fact that the same sample assessed the EVQ items twice, allowing analysis of two datasets from the same sample, signifies another strength of the present study as the psychometric properties of the EVQ did not vary. Measures of validity and factor dimensionality were substantially consistent from T1 to T2, demonstrating the four-factor model being significantly superior to the two-factor model at both points in time. These results suggest that the Norwegian version of the EVQ is a valid and reliable measure among adolescents.

Nevertheless, some limitations should be taken into consideration. This study of the Norwegian version of the EVQ included adolescents aged 13-17 years old. Thus, the present results cannot be generalized to younger children nor to older adolescents. Also, a pretesting of the Norwegian version among Norwegian adolescents was not undertaken, which represents a limitation.

Conclusion

This study evaluated the psychometric properties of the Norwegian version of the EVQ among adolescents in secondary school, by assessing the dimensionality, reliability and construct validity. The EVQ demonstrated satisfying reliability and construct validity, while the dimensionality seemed somewhat blurred or indistinct. However, the four-factor model seems superior to the two-factor model. Hence, the Norwegian version appeared to be appropriate and can be used to measure adolescents' expectancy-related believes and subjective task values in PE. However, some further development might be useful to strengthen the construct validity of the EVQ. The three factors measuring the task values (AV, IV and UV) comprising of two indicators each, indicate rather weak and thus uncertain constructs. Therefore, a further development of the EVQ should include at least one more item tapping into these factors, which possibly, might strengthen the EVQ construct validity.

Disclosure Statement

No potential conflict of interest was reported by the authors.

References

Acock, A. C. (2013). *Discovering structural equation modeling using stata. (rev. ed.)*. College Station, TX: STATA Press.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review, 64*(6p1), 359

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the academy of marketing science, 16*, 74-94. https://doi.org/10.1007/BF02723327

Bailey, R., Armour, K., Kirk, D., Jess, M., Pickup, I., Sandford, R., & BERA Physical Education and Sport Pedagogy Special Interest Group. (2009). The educational benefits claimed for physical education and school sport: an academic review. *Research papers in education, 24*, 1-27. https://doi.org/10.1080/02671520701809817

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215

Battle, A., & Wigfield, A. (2003). College women's value orientations toward family, career, and graduate school. *Journal of Vocational Behavior, 62*, 56-75. https://doi.org/10.1016/S0001-8791(02)00037-4

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1-40

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Centers for Disease Control and Prevention [CDC]. (2014). *State indicator report on Physical Activity, 2014*. Atlanta: Division of Nutrition, Physical Activity, and Obesity Retrieved from http://www.cdc.gov/physicalactivity/downloads/pa_state_indicator_report_2014.pdf.

Chen, A., & Liu, X. (2009). Task values, cost, and choice decisions in college physical education. *Journal of Teaching in Physical Education, 28*(2), 192-213

Chen, A., Martin, R., Ennis, C. D., & Sun, H. (2008). Content specificity of expectancy beliefs and task values in elementary physical education. *Research Quarterly for Exercise and Sport, 79*, 195-208. https://doi.org/10.1080/02701367.2008.10599483

Chen, S., Chen, A., Sun, H., & Zhu, X. (2013). Physical activity and fitness knowledge learning in physical education: Seeking a common ground. *European Physical Education Review, 19*, 256-270. https://doi.org/10.1177/1356336X13486058

Cox, A. E., Smith, A. L., & Williams, L. (2008). Change in physical education motivation and physical activity behavior during middle school. *Journal of Adolescent Health, 43*, 506-513. https://doi.org/10.1016/j.jadohealth.2008.04.020

Eccles, J. S. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives : Psychological and sociological approaches* (pp. 70-146). San Francisco, CA: Freeman.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescent achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*(3), 215-225

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology, 53*, 109-132

European Commission/EACEA/Eurydice. (2013). *Physical Education and Sport at School in Europe*. Retrieved from Luxembourg: http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/150en.pdf

Feather, N. T. (1982). Human values and the prediction of action: An expectancy-valence analysis. In N. T. Feather (Ed.), *Expectations and actions: Expectancy-value models in psychology* (pp. 263-289). Hillsdale NJ: Lawrence Erlbaum.

Feltz, D. L., & Chase, M. A. (1998). The measurement of self-efficacy and confidence in sport. In J. L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 65-80). Morgantown, WV: Fitness Information Technology, Inc.

Gao, Z. (2009). Students' motivation, engagement, satisfaction, and cardiorespiratory fitness in physical education. *Journal of Applied Sport Psychology, 21*, S102-S115. https://doi.org/10.1080/10413200802582789

Gao, Z., Lee, A. M., & Harrison, L., Jr. (2008). Understanding students' motivation in sport and physical education: From the expectancy-value model and self-efficacy theory perspectives. *Quest, 60*, 236-254. https://doi.org/10.1080/00336297.2008.10483579

Gjersing, L., Caplehorn, J. R. M., & Clausen, T. (2010). Cross-cultural adaptation of research instruments: Language, setting, time and statistical considerations. *BMC Medical Research Methodology, 10*, 13-13. https://doi.org/10.1186/1471-2288-10-13

Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and evaluation in Counseling and Development, 36*(3), 181-191

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of organizational behavior, 18*, 667-683

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

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Martin, A. J., & Dowson, M. (2009). Interpersonal relationships, motivation, engagement, and achievement: Yields for theory, current issues, and educational practice. *Review of educational research, 79*, 327-365. https://doi.org/10.3102/0034654308325583

Mehmetoglu, M., & Jakobsen, T. G. (2017). *Applied statistics using stata: A guide for the social sciences*. London, UK: SAGE.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*, 737-745. https://doi.org/10.1016/j.jclinepi.2010.02.006

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the "×" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science, 22*, 1058-1066. https://doi.org/10.1177/0956797611415540

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures. Issues and applications*. Thousand Oaks, CA: SAGE.

Nicholls, J. G. (1979). Quality and equality in intellectual development: The role of motivation in education. *American Psychologist, 34*(11), 1071-1084

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

Potter, W. J., & Levine-Donnerstein, D. (1999). Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research, 27*, 258-284. https://doi.org/10.1080/00909889909365539

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. V. Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: SAGE.

Sharma, S. (1996). *Applied multivariate techniques*. New York: Wiley.

Simons, J., Dewitte, S., & Lens, W. (2003). "Don't do it for me. Do it for yourself!" stressing the personal relevance enhances motivation in physical education. *Journal of Sport and Exercise Psychology, 25*, 145-160. https://doi.org/10.1123/jsep.25.2.145

StataCorp. (2015). Stata statistical software: Release 14. College Station, TX: StataCorp LLC.

Statistics Norway. (2016). Marks, lower secondary school, 2016. Retrieved from https://www.ssb.no/en/utdanning/statistikker/kargrs

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Pearson/Allyn and Bacon.

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy-value theory: A latent interaction modeling study. *Journal of Educational Psychology, 104*, 763-777. https://doi.org/10.1037/a0027470

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*, 49-78. https://doi.org/10.1007/bf02209024

Wigfield, A., & Cambria, J. (2010). Students' achievement values, goal orientations, and interest: Definitions, development, and relations to achievement outcomes. *Developmental Review, 30*, 1-35. https://doi.org/10.1016/j.dr.2009.12.001

Wigfield, A., & Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. *Developmental Review, 12*(3), 265-310

Wigfield, A., Eccles, J. S., Yoon, K. S., Harold, R. D., Arbreton, A. J., Freedman-Doan, C., & Blumenfeld, P. C. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology, 89*, 451-469

Wigfield, A., & Tonks, S. (2002). Adolescents' expectancies for success and achievement task values during the middle and high school years. In F. Pajares & T. Urdan (Eds.), *Academic motivation of adolescents* (pp. 53-82). Connecticut, CT: Information Age Publishing.

Wigfield, A., Tonks, S., & Klauda, S. L. (2009). Expectancy-value theory. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 55-75). New York: Routledge.

World Health Organization [WHO]. (2010). *Global recommendations on physical activity for health*. Retrieved from http://www.cdc.gov/physicalactivity/downloads/pa_state_indicator_report_2014.pdf

World Health Organization [WHO]. (2016). *Physical activity strategy for the WHO European Region 2016–2025*. Retrieved from https://www.dgs.pt/ficheiros-de-upload-2013/pnpaf-who-eu-physical-activity-strategy-2016-2025-final-2016-pdf.aspx.

Xiang, P., McBride, R., & Bruene, A. (2004). Fourth graders' motivation in an elementary physical education running program. *Elementary School Journal, 104*, 253-266. https://doi.org/10.1086/499752

Xiang, P., McBride, R., & Bruene, A. (2006). Fourth-grade students' motivational changes in an elementary physical education running program. *Research Quarterly for Exercise and Sport, 77*, 195-207. https://doi.org/10.1080/02701367.2006.10599354

Xiang, P., McBride, R., & Guan, J. (2004). Children's motivation in elementary physical education: A longitudinal study. *Research Quarterly for Exercise and Sport, 75*, 71-80. https://doi.org/10.1080/02701367.2004.10609135

Xiang, P., McBride, R., Guan, J., & Solmon, M. (2003). Children's motivation in elementary physical education: An expectancy-value model of achievement choice. *Research Quarterly for Exercise and Sport, 74*, 25-35. https://doi.org/10.1080/02701367.2003.10609061

Zhu, X., & Chen, A. (2010). Adolescent expectancy-value motivation and learning: A disconnected case in physical education. *Learning and Individual Differences, 20*, 512-516. https://doi.org/10.1016/j.lindif.2010.04.013

Zhu, X., & Chen, A. (2013). Motivational cost aspects of physical education in middle school students. *Educational Psychology, 33*, 465-481. https://doi.org/10.1080/01443410.2013.785043

Zhu, X., Sun, H., Chen, A., & Ennis, C. (2012). Measurement invariance of expectancy-value questionnaire in physical education. *Measurement in Physical Education and Exercise Science, 16*, 41-54. https://doi.org/10.1080/1091367X.2012.639629

Appendix A:
EVQ English version

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

Appendix B:
EVQ Norwegian version

1. Hvor god er du i kroppsøving?
   Veldig god 5 4 3 2 1 Veldig dårlig
2. Dersom du gav karakteren 5 til den beste studenten i kroppsøving og 1 til den dårligste, hvilken karakter ville du gitt deg selv?
   Beste 5 4 3 2 1 Dårligste
3. Noen barn er bedre i et fag/emne enn et annet. Du kan for eksempel være bedre i matematikk enn i lesing. Sammenlignet med de fleste av de andre fagene dine på skolen, hvordan vil du beskrive dine ferdigheter i kroppsøving?
   Mye bedre 5 4 3 2 1 Mye dårligere.
4. Hvor god mener du at du er i kroppsøving?
   Veldig god 5 4 3 2 1 Veldig dårlig.
5. Hvor fysisk aktiv er du i kroppsøving?
   Veldig aktiv 5 4 3 2 1 Lite aktiv.
6. Hvor viktig synes du kroppsøving er for deg?
   Veldig lite viktig 1 2 3 4 5 Veldig viktig.
7. Sammenlignet med matematikk, lesing og naturfag, hvor viktig er det for deg å lære hva kroppsøving består av?
   Veldig lite viktig 1 2 3 4 5 Veldig viktig.
8. Hvor morsomme synes du kroppsøvingstimene er generelt?
   Veldig kjedelige 1 2 3 4 5 Veldig morsomme.
9. Hvor godt liker du kroppsøvingstimene?
   Liker dem ikke i det hele tatt 1 2 3 4 5 Liker dem veldig godt.
10. Noen av tingene du lærer på skolen hjelper deg til å gjøre andre ting på en bedre måte utenfor skolen. Vi kaller disse tingene nyttige. Det du lærer om planter på skolen kan for eksempel hjelpe deg til å dyrke planter hjemme i hagen. Hvor nyttig mener du innholdet du lærte i kroppsøvingstimene er?
    Ikke nyttig i det hele tatt 1 2 3 4 5 Veldig nyttig.
11. Sammenlignet med de andre fagene på skolen, hvor nyttig er det du lærer i kroppsøving?
    Ikke nyttig i det hele tatt 1 2 3 4 5 Veldig nyttig.