Qin et al. BMC Public Health (2024) 24:3117
https://doi.org/10.1186/s12889-024-20563-0

BMC Public Health

RESEARCH

Open Access

The Chinese version of the physical activity questionnaire for adolescents: a psychometric validity, reliability, and invariance study

Ling Qin[1,2], Walter King Yan Ho[3] and Selina Khoo[2*]

Abstract

Background The Physical Activity Questionnaire for Adolescents (PAQ-A) is one of the most commonly used questionnaires for assessing adolescents' levels of moderate to vigorous physical activity (PA). Although a Chinese version of the PAQ-A (comprising nine items) exists, it has been assessed for only internal consistency coefficients, test–retest reliability, and correlations with the GT3X + accelerometer. Therefore, this study aimed to evaluate the psychometric properties of the Chinese version of the PAQ-A among Chinese adolescents.

Methods This study included three samples of 1,101 secondary school students (12–19 years old). The first sample consisted of 518 participants (50.4% males, 49.6% females) for correlation tests and exploratory factor analysis (EFA). The second sample consisted of 227 participants (55.75% males, 44.24% females) for confirmatory factor analysis (CFA), convergent validity, and sample size invariance tests. The third sample consisted of 356 participants (54.50% male, 45.50% female) for the internal consistency reliability test, invariance test (cross sample size, education level and gender), and t-test.

Results Pearson correlation analysis of the PAQ-A indicated that all the total-item correlations exceeded 0.2, indicating good consistency across the items. Subsequent EFA of the Chinese version of the PAQ-A revealed a two-factor structure. CFA subsequently validated this structure. One of the items exhibited a standardized loading below 0.4 and was excluded. The exclusion of this item resulted in increased standardized loadings for the remaining items, ranging from 0.40 to 0.82, which indicates improved fit indices. This adjustment underscores the questionnaire's satisfactory convergent validity and robust discriminant validity. The overall Cronbach's alpha was 0.821, with the values for the first and second factors being 0.757 and 0.716, respectively. The questionnaire demonstrated stable invariance across sample sizes and education levels and additionally showed partial scalar invariance across genders. A t-test revealed a significant difference between males and females, which aligns with previous findings. These findings supported the construct validity of the questionnaire.

Conclusion This study validated the Chinese version of the PAQ-A for assessing adolescent PA in China, with a two-factor structure improved by removing one item.

Keywords Validity, Reliability, Invariance, Psychometric properties, Questionnaire validation, Physical activity, Youth, China

*Correspondence:
Selina Khoo
selina@um.edu.my
Full list of author information is available at the end of the article

Introduction

Physical activity (PA) is any bodily movement produced by skeletal muscles that requires energy expenditure [1]. The involvement of adolescents in moderate-to-vigorous PA has been associated with a wide range of health benefits, including improvements in cardiovascular fitness [2], metabolic efficiency [3], muscular fitness [4], bone health [5], and body composition [6]. In addition, engaging in moderate-to-vigorous PA has a favorable impact on adolescents' self-perception [7], cognitive capacities [8], and general quality of life [9]. Additionally, several studies have demonstrated a beneficial impact of PA on adolescents' academic achievements, including grade point averages, scores on standardized tests, and grades in specific courses [10–12]. Therefore, the World Health Organization (WHO) recommends that children and adolescents (5–17 years old) perform an average of at least 60 min of moderate-to-vigorous intensity PA daily [13].

Despite the numerous beneficial effects of PA, participation in PA typically diminishes throughout adolescence. Guthold et al. [14] analyzed 298 school-based surveys from 146 countries covering 1.6 million students aged 11 to 17 years to assess adherence to the WHO's PA recommendations [13]. Their study revealed a high prevalence of insufficient PA globally, with notable variations across regions. Specifically, the highest rates were observed in high-income Asia–Pacific countries (92.2%), while the lowest were in South Asia (75.2%). In China, a national survey was conducted to assess PA levels among children and adolescents utilizing the Global Matrix 4.0 Indicators as a framework [15]. The extensive survey included a representative sample of 133,006 school-aged children and adolescents (aged 9 to 17 years). The study highlights significantly low PA levels among Chinese youth, with only 14% meeting the WHO's guidelines in 2022 [1], which is marginally higher than the 13% in 2018. An age-related decrease in PA compliance was also detected, decreasing from 58.5% among primary secondary school students to 35.5% among upper secondary school students. Insufficient PA is associated with adverse health outcomes, including an elevated risk of overweight, obesity, and cardiometabolic disorders [16, 17]. Zhu et al.'s [18] study revealed that Chinese adolescents in upper secondary schools were significantly more likely than children in primary schools to be overweight. The decline in adherence to PA and the increase in overweight among adolescents across educational levels highlight the urgent need for research on PA among Chinese adolescents.

To address physical inactivity among Chinese adolescents, PA needs to be accurately measured for analyzing health associations, risk assessments, guideline formulation, identifying disparities and low-activity populations, and evaluating intervention impacts [19, 20]. PA can be assessed by objective and subjective methods [21]. Objective methods (e.g., heart rate monitors and accelerometers) have some challenges, such as the complex calibration and the invasiveness (e.g., chest strap heart rate monitors) of equipment [22]. Furthermore, a critical observation in equipment assessments is the variability in PA estimates among individuals attributed to disparities in equipment brands, the location of wear, and the specific PA thresholds set by different devices [23, 24]. Moreover, the cost of devices can be prohibitive for widespread use in large-scale studies or by individuals with limited resources [25]. Subjective methods of measuring PA include recall questionnaires, activity logs or diaries, and proxy reports, either self-administered or administered through interviews [26]. These methods are the predominant techniques for evaluating PA [27]. They offer a cost-effective alternative for assessing PA, capturing details on the intensity, frequency, and type of PA in individuals, making them well-suited for large-scale studies [21, 28]. Questionnaires are the most widely used form of self-reporting. There are 61 questionnaires measuring physical activity in children and adolescents [29]. The number of items on these questionnaires varies, and each questionnaire may focus on different demographics, and cover different dimensions. One of the most widely used questionnaires to assess PA in adolescents is the Physical Activity Questionnaire for Adolescents (PAQ-A) [30], which was developed to assess the general level of PA among adolescents, typically those in secondary school.

Kowalski et al. [30] evaluated the convergent validity of the PAQ-A by comparing it with other measures of PA (e.g., the Seven-Day Recall Interview, Activity Rating, Leisure Time Exercise Questionnaire, and Caltrac Motion Sensor) to determine its effectiveness in assessing the general level of PA among 85 secondary school students. The study found low to high correlations between the PAQ-A scores and the other PA measures (strong correlation with the Activity Rating [$r = 0.73$], moderate correlation with the Leisure Time Exercise Questionnaire [$r = 0.57$], weak correlation with the Caltrac motion sensor [$r = 0.33$], and moderate correlation with the Seven-Day Recall Interview [$r = 0.59$]). Martínez-Gómez et al. [31] evaluated the reliability and validity of the PAQ-A in a Spanish adolescent population using accelerometer measurements as a criterion for comparison. The study showed the reliability of the PAQ-A, with an ICC of 0.71. Internal consistency measured by Cronbach's alpha ranged from 0.65 to 0.67 among 82 adolescents and was 0.74 in a larger group of 232. The questionnaire moderately correlated with PA levels ($rho = 0.39$ for total activity and rho = 0.34 for MVPA) per accelerometer. Both studies showed that the PAQ-A questionnaire is reliable and valid for assessing PA levels in secondary school students with different backgrounds. The PAQ-A has been translated and validated in various languages, including Amharic [32], Dutch [33], Malay [34], Thai [35], Indonesian [36], and Polish [37].

The Chinese version of the PAQ-A was translated by Li et al.[38], and it was found to have acceptable reliability and validity among 125 adolescents. The study reported internal consistency coefficients of 0.82 and 0.85, alongside a test–retest reliability of 0.81. Additionally, moderate correlations with GT3X + accelerometer data were observed for overall activity (0.51) and moderate to vigorous PA (0.40), both of which were significant at $p < 0.05$. This validates the effectiveness of the Chinese PAQ-A for measuring PA levels. Subsequently, the Chinese version of the PAQ-A has been widely utilized in various studies on personal health and physical education within Chinese contexts, specifically to assess PA among secondary school students [39–44].

Although the Chinese version of the PAQ-A has been used in several Chinese studies, exploratory factor analysis (EFA) has not been conducted to test construct validity, which deviates from the standard scale testing process [45]. Moreover, this gap may lead to uncertainties in accurately capturing Chinese adolescents' PA. When validating the Chinese version of the PAQ-A, Li et al. [38] reported only the internal consistency coefficient, test–retest reliability, and correlations with the GT3X + accelerometer; other psychometric property indicators, such as composite reliability (CR), average variance extracted (AVE), and invariance, were not tested. Given that the specific characteristics of diverse populations can significantly affect the efficacy of an instrument in measuring a construct, it is important to evaluate the instrument's psychometric properties within the target population before its application in an adapted format [46]. Therefore, this study aims to evaluate the psychometric properties of the Chinese version of the PAQ-A among adolescents in China.

Methods

Participants

In 2023, three samples of secondary school students from the urban area of Chongqing, China, were sequentially recruited via random sampling. A total of 1,101 secondary school students participated in this study. According to Comrey and Lee [47], the standards for sample size in EFA are as follows: 50 is very poor, 100 is poor, 200 is fair, 300 is good, and 500 is very good. Based on these guidelines, the first sample consisted of 518 participants (male:261, female: 257) aged 12–19 years (mean: 14.61, SD: 1.43) for item analysis and EFA. Through an online sample size calculation tool [48], the following parameters were established for determining the necessary minimum sample size for our study: an effect size of 0.3, a statistical power level of 0.8, the inclusion of two latent variables, and eight observed variables, all with a significance threshold (alpha level) set at 0.05. The recommended minimum sample size was 100. According to this criterion, the second sample consisted of 227 participants (male:126, female: 101), aged 12–19 years (mean: 14.71 SD: 1.47), for CFA, convergent validity, discriminant validity, and sample size invariance tests. The third sample consisted of 356 participants (194 males, 162 females), aged 12–18 years (mean: 14.73, SD: 1.51), for internal consistency reliability, invariance tests (sample size, education level, and gender), and t-test. To ensure data quality, we applied the following inclusion criteria: (1) currently enrolled secondary school students, (2) willing and able to participate with parental consent, and (3) without disability or illness.

Measures

A demographic survey collected information on participants' gender, age, and education level. The Chinese version of the PAQ-A was used based on Li et al.'s version [38]. The Chinese version of the PAQ-A consists of eight items aimed at assessing various contexts of weekly PA among adolescents and adheres to the protocols outlined in the instrument's manual by Kowalski et al. [49]. Specifically, participants are asked to rate their PA levels in various contexts (including physical education class, after school, and on weekends) on a 5-point scale, where 1 indicates low activity and 5 indicates high activity. According to Kowalski et al.'s [49] guideline, the mean of the eight items is the final PAQ-A summary score. A score of 1 indicates a low level of PA, while a score of 5 indicates a high PA level.

Procedure

Before the study, ethics approval was obtained from the first author's University Research Ethics Committee. Before data collection, secondary school principals were informed about the study in order to invite their students to participate in the study. Informed written consent was obtained from the students' parents or legal guardians and assent was obtained from students who agreed to participate in the study. The students were informed that participation in the survey was entirely voluntary, with the option to withdraw from the study at any time without penalty. The students were told that the study's objective was to assess their PA levels over the previous seven days, emphasizing that the survey did not seek 'right' or 'wrong' answers. The questionnaires were to be completed in a quiet classroom setting supervised by a researcher in the absence of their teachers to ensure an environment conducive to honest responses. Furthermore, the survey's anonymity was highlighted, assuring students that all provided information would remain confidential. It was also explicitly stated that their responses would not be accessible to their teachers, reinforcing the privacy and confidentiality of their participation.

Data analysis

SPSS (IBM) version 27 and SPSS Amos (IBM) version 26 were used for all the data analyses. Statistical significance was set at a probability level of $p < 0.05$.

Both EFA and CFA were utilized to assess the structural validity of the questionnaire. EFA aimed to uncover underlying common factors, enhancing the convergent and discriminant validity of questionnaire items [50]. The data's appropriateness for principal component analysis was assessed through the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity [51]. The extraction of factors was based on eigenvalues surpassing 1.0. According to Meyers et al. [52], Promax rotation was applied to account for inter-factor correlations. Following the recommendation by Nunnally [53], items exhibiting primary factor loadings below 0.50 or significant cross-loadings (secondary loadings above 0.40) were excluded. Subsequently, we conducted a CFA to further examine the questionnaire's structure. Multiple fit indices were used to evaluate the adequacy of model fitting to the data, including the chi-square statistic, comparative fit index (CFI), Tucker–Lewis index (TLI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA), which incorporated a 90% confidence interval (CI). An acceptable model fit is indicated by CFI and TLI values above 0.90 and SRMR and RMSEA values at or below 0.08 [54]. This study tested the convergent validity of the questionnaire. Convergent validity indicates how well items or tests measuring the same latent trait belong to the same dimension, and there should be a high correlation between the measurements obtained from these items [54]. Convergent validity can be evaluated in three ways: (1) factor loadings, or the items' loadings on the latent dimension being significant, with loadings exceeding 0.40; (2) average variance extracted (AVE), or how much of the variance in the latent factor is due to measurement error, with more significant AVE indicating greater variance in the item explained by the latent factor and the smaller the measurement error. Hair et al. [55] suggested that an AVE greater than 0.50 is perfect; however, an AVE greater than 0.40 is acceptable if the CR is greater than 0.60 [56–59], and (3) composite reliability, or the consistency of a set of items in CFA, is a more rigorous internal consistency measure. Hair [60] recommended that the CR meet or exceed the 0.70 cutoff point. Moreover, discriminant validity was tested in this study, ensuring that the constructed measure distinctly captures the phenomena of interest and not those exhibited by other variables within the structural equation model [54]. For the consistency reliability test, Cronbach's alpha was employed as the testing measure [61].

To assess the consistency across groups, we implemented an approach to model testing that progressively introduces constraints. Initially, we established configural invariance by ensuring an identical factor structure across groups. Subsequently, we advanced to metric invariance by mandating uniform factor loadings across the groups. The process culminated in the attainment of scalar invariance through setting uniform item intercepts across groups. We did not consider strict invariance (which involves equal error variances) as metric and scalar invariance are suffice for meaningful comparative group analyses [54]. Chen [62] proposed a criterion for assessing measurement invariance, advocating for a threshold of a 0.01 variation in the CFI, accompanied by alterations in the RMSEA of up to 0.015. Additionally, a fluctuation in the CFI ranging from 0.01 to 0.02 is also indicative of measurement invariance [63, 64]. Cheung and Rensvold [64] argued that although a change in CFI between 0.01 and 0.02 signifies a moderate deterioration in model fit, it does not indicate differences exist. A change exceeding 0.02 clearly suggests significant differences, thus rejecting the invariance model.

Results

Item analysis

Item-total correlations were evaluated as a crucial step for identifying any items that exhibited weak correlations (i.e., < 0.20) with the overall questionnaire [65], necessitating further scrutiny before proceeding to assess the questionnaire's validity. Pearson correlation coefficients were calculated, and the correlation results are presented in Table 1. All total-item correlations are greater than 0.20, suggesting that the PAQ-A items are appropriately related. This finding suggested good consistency across each item and supported the retention of all items.

Construct validity

Exploratory factor analysis

To evaluate the construct validity of the PAQ-A, we initially conducted a Kaiser–Meyer–Olkin (KMO) test and Bartlett's test of sphericity to determine the appropriateness of our sample size for EFA. The KMO test yielded a measure of sampling adequacy of 0.85, indicating a high degree of common variance among items suitable for factor analysis. Furthermore, Bartlett's test of

Table 1 Pearson's product-moment correlations for PAQ-A (*N* = 518)

Variable	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Total
Item 1	-								
Item 2	.158**	-							
Item 3	.077	.138**	-						
Item 4	.317**	.239**	.169**	-					
Item 5	.324**	.268**	.216**	.621**	-				
Item 6	.339**	.302**	.198**	.471**	.531**	-			
Item 7	.318**	.400**	.347**	.468**	.485**	.533**	-		
Item 8	.274**	.423**	.322**	.410**	.407**	.473**	.645**	-	
Total	.537**	.527**	.441**	.729**	.758**	.740**	.789**	.737**	-

** Correlation is significant at the 0.01 level (2-tailed)

Table 2 Factor loadings from EFA analysis of the PAQ-A

	Factor 1	Factor 2
Item 4	.800	-.017
Item 1	.788	-.262
Item 5	.776	.045
Item 6	.652	.182
Item 3	-.353	.883
Item 8	.218	.673
Item 2	.037	.598
Item 7	.344	.596

Extraction Method: Principal Component Analysis
Rotation Method: Promax with Kaiser Normalization
Rotation converged in 3 iterations

sphericity produced a chi-square value of 1275.739 with 28 *df* ($p < 0.001$), confirming the suitability of our data for factor analysis due to the significant correlation among variables. The EFA results revealed a two-factor solution based on eigenvalues exceeding 1. Specifically, Factor 1 consisted of four items (Item 1, Item 4, Item 5, and Item 6) associated with extracurricular PA, which loaded strongly onto Factor 1, with loadings ranging from 0.65 to 0.80. This factor accounted for 44.93% of the total variance. Factor 2 comprised four items (Item 2, Item 3, Item 7, and Item 8) mainly related to school-based PA, with factor loadings between 0.59 and 0.88, explaining 12.83% of the total variance (as shown in Table 2). This two-factor structure underscores the ability of the PAQ-A to distinguish between extracurricular and school-related PA, supporting its construct validity.

Confirmatory factor analysis

The initial model fit indices of the PAQ-A showed acceptable fit indices. Specifically, $\chi2 = 30.103$, $df = 19$, CFI = 0.979, TLI = 0.970, SRMR = 0.042, RMSEA = 0.051 (90% CI: 0.000 to 0.084). However, Item 3's standard loading was 0.36, which was below the cutoff of 0.40 (see Fig. 1).

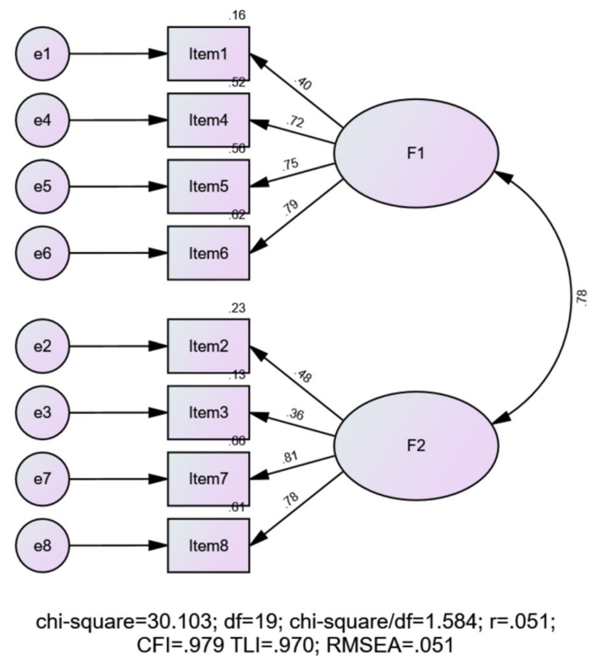

Fig. 1 Initial CFA model of the PAQ-A (F1 = Factor 1, F2 = Factor 2)

After removing Item 3, the final model demonstrated a good fit: $\chi2 = 21.263$, $df = 13$, CFI = 0.984, TLI = 0.974, SRMR = 0.035, and RMSEA = 0.053 (90% CI: 0.000 to 0.092) (see Fig. 2).

Convergent validity

The two factors of the structure demonstrated acceptable internal consistency reliability, with standard loadings ranging from 0.045 to 0.820, CR values ranging from 0.748 to 0.768, and AVE values ranging from 0.466 to 0.509. Table 3 shows the CFA results, including standardized loadings, t values, *p* values, factor loadings, squared

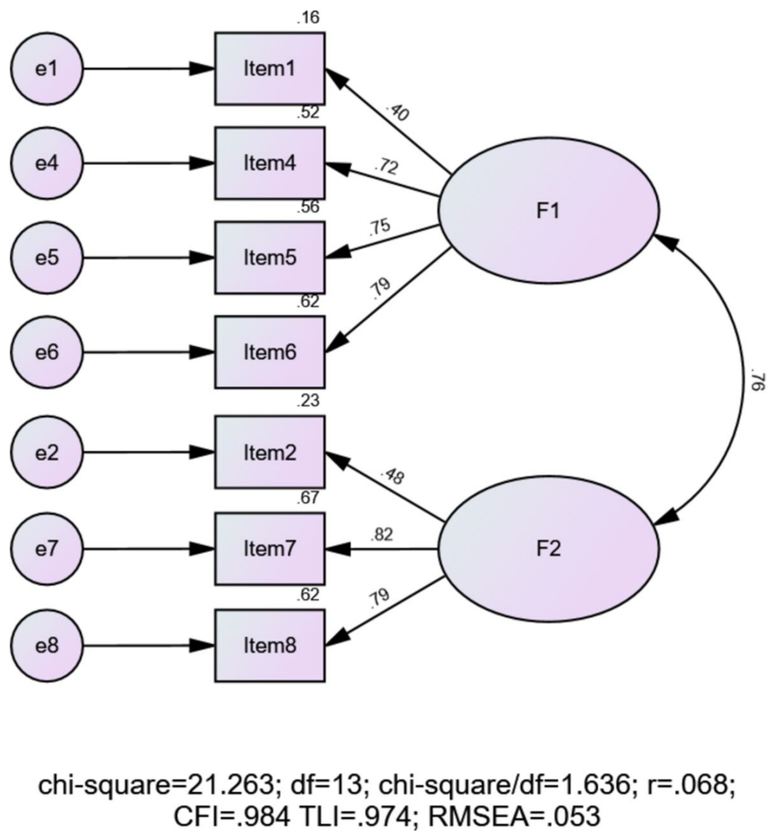

Fig. 2 Modified CFA model of the PAQ-A (F1 = Factor 1, F2 = Factor 2)

Table 3 CFA item standard errors (SE), *t*-value, *p*-value, factor loadings (std.), squared multiple correlations (SMC), composite reliability (CR) and average extracted variance (AVE) values for PAQ-A

	SE	*t*-value	*p*-value	Std	SMC	CR	AVE
Factor 1							
Item 1				.405	.164	.768	.466
Item 4	.607	.515	***	.722	.521		
Item 5	.612	.569	***	.748	.559		
Item 6	.563	.636	***	.787	.619		
Factor 2							
Item 2				.482	.232	.748	.509
Item 7	.261	.705	***	.820	.672		
Item 8	.259	.748	***	.790	.623		

*** $p < 0.001$

multiple correlations, and CR and AVE values for the PAQ-A.

Discriminant validity

Since two factors were explored in the EFA, this study tested the discriminant validity of the Chinese version of the PAQ-A questionnaire. Discriminant validity guarantees that a measurement of a construct is distinct and represents specific phenomena that are not captured by other measurements in a structural equation model [54]. According to Hair et al., the HTMT ratio should be less than 0.85 or 0.90 [55]. Table 4 shows that the HTMT ratio for the PAQ-A is less than 0.85, indicating that the two factors measured are statistically well differentiated.

Table 4 PAQ-A HTMT ratio

	Factor 1	Factor 2
Factor 1	—	
Factor 2	.792	—

Internal consistency reliability

The internal consistency reliability of the questionnaire factors was tested using Cronbach's alpha coefficient. Cronbach's alphas for the overall questionnaire, Factor 1, and Factor 2 were 0.821, 0.757, and 0.716, respectively (see Table 5).

Invariance test

This study employed a sequential model-testing strategy using multigroup CFA to assess the invariance of the PAQ-A measurement model across samples. Table 6 outlines the findings from tests assessing invariance across sample size, education level and gender.

Invariance across sample size

Initially, the analysis concentrated on examining configural invariance across sample sizes. This approach assumes a consistent model structure across groups, characterized by uniform items and an equal number of factors, with all items allowed to load on all factors. The configural invariance model related to cross-sample size exhibited a satisfactory model fit, indicated by χ2 (583) = 57.032, *df* = 26, CFI = 0.976, RMSEA = 0.045, SRMR = 0.035. This outcome supports the model's stability in factor structure across varying sample sizes. Upon verifying configural invariance across sample sizes, we examined metric invariance. We tested metric invariance by enforcing equivalence constraints on item factor loadings. The objective was to evaluate changes in model fit compared to the configural invariance model. The metric invariance model was also strongly supported, showing minimal changes in the model (ΔCFI = 0.002, ΔRMSEA = 0.005), with: χ2 (583) = 59.249, *df* = 31, CFI = 0.978, RMSEA = 0.040, SRMR = 0.030. Finally, we assessed scalar invariance by equalizing intercept parameters across groups and examining the impact on model fit relative to the metric invariance model. The analysis indicated a minimal change in model fit, ΔCFI = 0.009, ΔRMSEA = 0.004. The scalar invariance model also exhibited a good fit, with χ2 (583) = 77.925, *df* = 38, CFI = 0.969, RMSEA = 0.044, and SRMR = 0.047, further validating the measurement model's invariance regarding sample size. These results indicate that metric and scalar variances remained invariant across the sample sizes.

Table 5 Reliability of the PAQ-A

	M	SD	Cronbach's alpha
Total	2.623	.775	.821
Factor 1	2.410	.897	.757
Factor 2	2.907	.819	.716

Invariance across education level

The configural invariance model related to cross-education level exhibited a satisfactory model fit, indicated by χ2 (356) = 56.747, *df* = 26, CFI = 0.960, RMSEA = 0.058, SRMR = 0.053. This outcome supports the model's stability in factor structure across education levels. Upon verifying configural invariance across education levels, we examined metric invariance. The metric invariance model was also strongly supported, showing minimal changes in the model (ΔCFI = 0.001, ΔRMSEA = 0.005), with: χ2 (356) = 61.674, *df* = 31, CFI = 0.961, RMSEA = 0.053, SRMR = 0.059. Finally, we assessed scalar invariance by equalizing intercept parameters across groups and examining the impact on model fit relative to the metric invariance model. The results revealed minimal change in model fit, with a ΔCFI of 0.012 and a ΔRMSEA of 0.005. These results suggest a moderate deterioration in the model fit, which does not indicate the presence of differences. Therefore, strict invariance is still considered acceptable. The scalar invariance model also exhibited a good fit, with χ2 (356) = 73.861, *df* = 34, CFI = 0.959, RMSEA = 0.057, and SRMR = 0.057, further validating the measurement model's invariance regarding education level. These results indicate that metric and scalar variances remained invariant across education levels.

Invariance across gender

In general, the configural invariance model across genders demonstrated good model fit, χ2 (356) = 55.882, *df* = 26, CFI = 0.959, RMSEA = 0.057, SRMR = 0.057. Because good model fit was found in other model fit indices, we concluded that the results provide sufficient support for an equal number of factors between males and females. Next, we tested the viability of the metric invariance model. The metric invariance model was also strongly supported, showing minimal changes in model (ΔCFI = 0.004, ΔRMSEA = 0.007), with χ2 (583) = 57.931, *df* = 31, CFI = 0.963, RMSEA = 0.050, SRMR = 0.055. Finally, we examined the viability of the scalar invariance model. The results showed a decrease in model fit relative to the metric invariance model when the intercept parameters were constrained to be equal, ΔCFI = 0.031, ΔRMSEA = 0.011, and ΔCFI, slightly exceeding the standard (ΔCFI = 0.01) and rejecting the hypothesis of invariant intercepts across gender. We found that the

Table 6 Fit indices for the invariance testing of measurement model across sample size, education level and gender

Invariance Testing	Model	χ2	*df*	*p*-value	CFI	RMSEA (90%CI)	SRMR	Model Comparison	ΔCFI	ΔRMSEA
Across sample size (The first sample: $N=227$, the second sample: $N=356$)	Configural	57.037	26	.000	.976	.045 (.029 to .061)	.035			
	Metric	59.249	31	.002	.978	.040 (.024 to .055)	.040	Configural vs Metric	.002	.005
	Scalar	77.925	38	.000	.969	.044 (.029 to .056)	.047	Metri vs Scalar	.009	.004
Across education level (Secondary level = 198, upper secondary level = 157)	Configural	56.747	26	.000	.960	.058 (.037 to .079)	.053			
	Metric	61.674	31	.001	.961	.053 (.033 to .072)	.059	Configural vs Metric	.001	.005
	Scalar	73.861	34	.000	.949	.058 (.040 to .076)	.069	Metri vs Scalar	.012	.005
Across gender (Males: $N=194$, females: $N=162$)	Configural	55.882	26	.001	.959	.057 (.036 to .078)	.057			
	Metric	57.941	31	.002	.963	.050 (.029 to .069)	.055	Configural vs Metric	.004	.007
	Scalar	87.735	38	.000	.932	.061 (.044 to .078)	.061	Metri vs Scalar	.031	.011

Chinese version of the PAQ-A questionnaire has configural and metric invariance but not scalar invariance, which means that at least one item intercept differs between the two groups [66]. Based on this, we tested the difference in scores between gender on the Chinese version of the PAQ-A in the next section using a t-test.

T-test

We calculated the sample size for the t-test sing G*Power, setting it to a two-tailed test with an effect size of 0.5, an alpha error of 0.05, a power of 0.95, and an allocation ratio of 1. The result indicated a required sample size of 210 (Group 1: 105, Group 2: 105). Therefore, the third sample ($N=356$, 194 males, 162 females) in this study met the requirements. Figure 3 presents the results of the t-test for the PAQ-A. Specifically, a two-tailed independent samples t-test was conducted to compare the means of males and females, with a 95% confidence interval for the difference. The results of the analysis revealed significant gender differences across six items. For Item 1 (Spare time PA), no significant difference was found between males (M = 1.77, SD = 0.65) and females (M = 1.70, SD = 0.63), t (356) = 1.01, $p = 0.31$. However, significant gender differences were found: Item 2 (PA during physical education classes), t (356) = 3.35, $p < 0.001$, Item 4 (PA after school), t (356) = 3.52, $p < 0.001$, Item 5 (PA in the evening), t (356) = 3.30, $p = 0.001$, Item 6 (PA on the weekend), t (356) = 4.03, $p < 0.001$, Item 7 (weekly PA frequency), t (356) = 3.83, $p < 0.001$; and Item 8 (overall weekly PA), t (356) = 2.89, $p = 0.004$.

Discussion

This study aimed to assess the psychometric properties of the Chinese version of the PAQ-A among Chinese adolescents. Correlational tests were conducted on the first

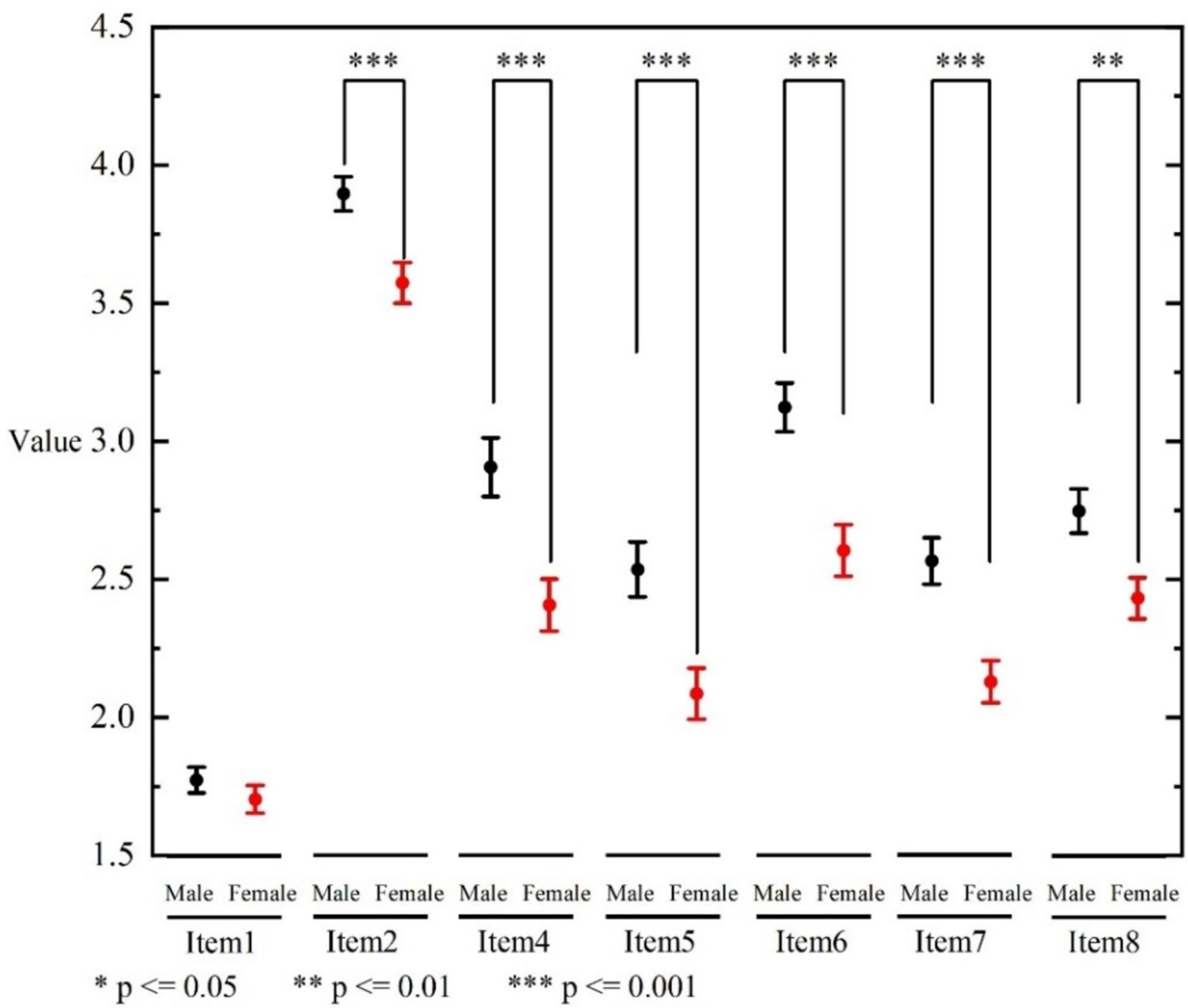

Fig. 3 Gender group t-test results

sample, which showed low to high correlations, and the total correlation of the items was greater than 0.2. Based on the EFA results, the Chinese PAQ-A showed a two-factor structure with eight items. The two-factor model was further tested using CFA. The standard factor loading for Item 3 was 0.36, which was less than the cutoff of 0.4. After Item 3 was removed, the model fit indices improved. Convergent and discriminant validity were tested, which presented acceptable results. Reliability was supported by the acceptable internal consistency of the two-factor questionnaire and the good internal consistency of the questionnaire. Furthermore, tests of configural, metric, and scalar invariance have shown that the assumptions of measurement invariance hold across sample sizes and education levels. Finally, through the t-test, except for Item 1, the other items showed statistically significant differences between genders.

The eight items of the PAQ-A demonstrated a range of correlations from weak to strong, for both inter-item and overall correlations. These results were similar to those of Bajamal's [67] study, which showed moderate to strong correlations with each other, with coefficients ranging from $r=0.23$ ($p<0.01$) to $r=0.49$ ($p<0.01$). These findings underscore the internal consistency and dimensional coherence of the PAQ-A, suggesting that the instrument can capture a unified construct of PA among adolescents. Furthermore, the alignment with other studies reinforces the reliability of the PAQ-A as a tool for assessing adolescent PA levels across diverse settings.

This study employed EFA to explore the structure of the PAQ-A. This analysis identified a two-factor structure in the Chinese PAQ-A that distinguishes between extracurricular and school-based PA. The CFA supported the two-factor structure obtained from the EFA. In the initial model, all other fit indices achieved the anticipated levels, except for Item 3 about noon-break PA, which exhibited a standardized loading below 0.4. This finding is consistent with those of some studies [36, 68]. Li et al. [38] reported that the test–retest reliability of the Chinese version of the PAQ-A increased from 0.79 to 0.81 after the deletion of Item 3. Furthermore, when conducting a correlation analysis between the overall questionnaire scores and accelerometer counts, the deletion of Item 3 resulted in an increase in correlation coefficients from 0.48 and 0.39 ($p<0.05$) to 0.51 and 0.40 ($p<0.05$), respectively. In a Spanish study conducted by Martínez-Gómez et al. [31], the PAQ-A was administered to 203 adolescents aged between 12 and 17 years, and the results similarly found no statistically significant correlation between Item 3 and pedometer readings. Similar findings regarding the low consistency of Item 3 with the remaining questionnaire have also been repeatedly observed in previous studies analyzing the psychometric properties of the PAQ-A [34, 68, 69]. These findings indicate that Item 3 exhibits low standardized loading and negatively impacts the reliability of the PAQ-A questionnaire and its correlation with adolescent accelerometer counts. In this study, excluding Item 3 from the analysis increased the fit indices for the model.

Convergent validity presented acceptable values, aligning with Fornell and Larcker's [56] findings, suggesting that an AVE exceeding 0.40 is deemed acceptable as long as the CR surpasses 0.60. However, these metrics did not meet the stricter guidelines recommended by Hair et al. [55], which require an AVE of at least 0.50. This discrepancy may be explained by significant correlations among questionnaire items or the presence of items with cross-loadings [70]. Another reason might be potential inaccuracies in the questionnaire's translation process. Despite a cultural adaptation of the PAQ-A for the Chinese context, with participants reporting no comprehension issues with any of the items during testing [38], there remains a possibility that the translation of the questionnaire failed to fully capture the original concept of weekly PA engagement. In subsequent studies, the convergent validity of the PAQ-A should be examined across cultures and areas. This study tested the questionnaire and found it to possess good discriminant validity, indicating that the Chinese PAQ-A effectively distinguishes between two factors.

In previous studies, Cronbach's alpha coefficients of the PAQ-A were tested. Bajamal et al. [67] reported a Cronbach's alpha of 0.81, while Koh et al. [34] found a coefficient of 0.84. Bervoets et al. [33] observed a Cronbach's alpha of 0.77. Moreover, Li et al. [38] found Cronbach's alpha to be 0.82 for the Chinese PAQ-A. Consistent with the findings of Li et al. [38], our research also identified a Cronbach's alpha coefficient of 0.821. This consistency across studies underscores the reliability of the PAQ-A in measuring PA levels among adolescents with a high degree of internal consistency. Furthermore, this study identified a two-factor structure through EFA, which also exhibited acceptable Cronbach's alpha coefficients (Factor $1=0.757$, Factor $2=0.716$), indicating the reliability of these factor structures.

This study presents the first examination of invariance measurement for the PAQ-A questionnaire, finding that the test for measurement invariance of the Chinese version of the PAQ-A supports the equivalent use of a two-factor structure of the PAQ-A across sample sizes and education levels, indicating the cross-cultural applicability and reliability of the PAQ-A for assessing PA levels among adolescents. However, we found that the PAQ-A does not support gender invariance in China. Specifically, while configural and metric invariance were supported, scalar invariance was not, which suggests that at least one item intercept in the questionnaire is inconsistent between males and females [66]. Consequently, we conducted a t-test to further investigate this discrepancy. Given the distinct psychological attributes of male and female groups that may uniquely influence their behavior, it is crucial to conduct this analysis [71].

Our study revealed that, in addition to Item 1, there were statistically significant differences between males and females in the other items, consistent with the more recent findings of Liu et al. [15]. They discovered that boys, constituting 54.5% of the sample, exhibited higher overall PA levels than girls at 44.9%. Furthermore, the percentage of boys participating in organized sports, at 18.3%, surpassed that of girls, at 16.5%. Regarding active play levels, boys, representing 52.0% of the sample, were more active than girls, who accounted for 41.0%. Moreover, an earlier study also demonstrated that among children and adolescents in China, boys were significantly more likely to achieve the recommended 60 min of MVPA per day than girls [18]. Future research should focus on developing interventions aimed at increasing PA among females, addressing the specific challenges they face in maintaining adequate PA levels. Furthermore, based on the Chinese version of the PAQ-A questionnaire scores in this study, Item 1 (Spare time PA) received the lowest score, while Item 2 (PA during physical education classes) received the highest score. This finding is similar to those reported by Bervoets et al. [33], Rahayu et al. [36], and Wyszyńska et al. [37]. This suggests that students' participation in PA during physical education classes is higher than during their spare time across cultures and regions. The consistency of these findings across gender and regional contexts affirms the reliability of the Chinese PAQ-A in accurately capturing differences in PA participation among students, thus providing additional evidence of its construct validity [72].

The observed results indicate that females have lower PA levels and that students participate more in PA during physical education classes than during their spare time. These findings suggest the need for targeted policy and educational practice interventions. For policymakers, this means creating more opportunities tailored to female participation in PA, such as implementing gender-specific programs and fostering inclusive environments that actively support female engagement. Additionally, the preference for structured physical education classes over unstructured spare-time activities highlights the importance of extending PA opportunities beyond the school setting. Establishing accessible community programs and recreational facilities can provide diverse options for adolescents to engage in PA during their leisure time. From an educational practice, schools should design interventions that promote PA among female students, including creating inclusive physical education programs that align with female interests and can contribute to a more supportive environment. Moreover, schools should leverage physical education classes effectively while encouraging students to remain active outside structured class times. Incorporating PA into other parts of the school day, offering extracurricular activities, and educating students on the benefits of an active lifestyle can potentially enhance overall PA levels.

Strengths and limitations

This study has several strengths. First, this study contributes significantly to the literature by initially validating the psychometric properties of the Chinese version of the PAQ-A, including item analysis, construct validity, and invariance testing. This rigorous assessment underscores the validity, reliability and invariance of the PAQ-A for future research endeavors, confirming its utility as a robust tool for assessing adolescent PA patterns. Second, this study subjected the Chinese version of the PAQ-A questionnaire to EFA for the first time and identified a two-factor structure contrary to previous studies. This finding reveals the presence of two independent dimensions involved in measuring PA among adolescents in the Chinese version of the PAQ-A. The identification of this structure not only provides a new perspective on the complexity of adolescent PA but also holds significant implications for more precise measurements and interventions using the PAQ-A questionnaire within Chinese contexts. Moreover, this result fosters the adaptation of tools for cross-cultural research, offering a richer and more nuanced theoretical foundation for studies on adolescent PA. Third, this is the first study to examine the measurement invariance of the PAQ-A across variables such as sample size, education level, and gender. The results indicated stable invariance across sample sizes and education levels and demonstrated partial scalar invariance across genders. Establishing measurement invariance is a critical prerequisite for comparing scores across groups and confirms that variations in scores are due to attitudinal differences rather than psychometric differences among study populations [62]. Finally, this study validated the Chinese PAQ-A as a useful instrument for researchers and school policymakers to determine the levels of PA among Chinese adolescents.

This study has several limitations. The findings are based on a sample of adolescents from urban areas in China, which may limit the generalizability of the results to a broader Chinese population. Future investigations should include more representative samples to better reflect Chinese society. This study did not examine test–retest reliability due to time and financial constraints, further compounded by limitations in resource allocation, which should be addressed in future research through a retest design. Finally, this study employed a cross-sectional design, so future research should consider conducting longitudinal studies to investigate the stability of the PAQ-A over time. Additionally, future studies must evaluate the relationships between PAQ-A scores and other variables (e.g., self-efficacy, well-being, and quality of physical education).

Conclusion

The revised Chinese version of the PAQ-A has good psychometric properties. The results indicate that the two-factor structure, following the removal of Item 3, exhibits good model fit, convergent validity, and discriminant validity. It also shows strong measurement invariance across sample sizes, and the results of this study revealed significant gender differences in PA, as well as differences across various contexts, such as during physical education classes, after school, in the evening, and on weekends, which aligns with previous studies. Consequently, the Chinese version of the PAQ-A can be a useful tool for measuring PA in adolescents. Future research should further validate the questionnaire's structural validity in a broader Chinese population.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-024-20563-0.

Supplementary Material 1.

Acknowledgements

We sincerely thank the young participants and their parents who contributed to this study and the school staff who facilitated and supported the data collection process.

Authors' contributions

LQ and SK significantly contributed to the conception and design of the work, drafting the manuscript, while WH played a crucial role in conceptualization, design, and critically revising the manuscript for important intellectual content. All authors (LQ, WH, and SK) have given their final approval for the publication of this manuscript and agree to be accountable for all aspects of the work, ensuring that any questions related to the accuracy or integrity are thoroughly investigated and resolved. All authors have read and approved the final manuscript.

Funding

The authors would like to acknowledge the financial support received for this paper's research, writing, and publication. This work was made possible through grants from the Natural Science Fund of the Chongqing Municipal Education Commission in China, under the reference numbers KJQN202200502.

Data availability

The dataset used during the current study is available on reasonable request from LQ.

Declarations

Ethics approval and consent to participate

The Universiti Malaya Ethics Committee reviewed and approved this study, reference number UM.TNC2/UMREC-2042. This study adhered to the guidelines set forth in the Declaration of Helsinki, and written informed consent was secured from the parents.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Faculty of Physical Education and Health Science, Chongqing Normal University, Chongqing, China. [2]Faculty of Sports and Exercise Science, Universiti Malaya, Kuala Lumpur, Malaysia. [3]Tokyo Gakugei University, Tokyo, Japan.

Received: 27 April 2024 Accepted: 29 October 2024
Published online: 11 November 2024

References

1. World Health Organisation. Physical activity. 2022. https://www.who.int/news-room/fact-sheets/detail/physical-activity.
2. Tarp J, Brønd JC, Andersen LB, Møller NC, Froberg K, Grøntved A. Physical activity, sedentary behavior, and long-term cardiovascular risk in young people: A review and discussion of methodology in prospective studies. J Sport Health Sci. 2016;5:145–50.
3. Whooten R, Kerem L, Stanley T. Physical activity in adolescents and children and relationship to metabolic health. Curr Opin Endocrinol Diabetes Obes. 2019;26:25–31.
4. Faigenbaum AD, MacDonald JP, Stracciolini A, Rebullido TR. Making a strong case for prioritizing muscular fitness in youth physical activity guidelines. Curr Sports Med Rep. 2020;19:530–6.
5. Yang X, Zhai Y, Zhang J, Chen J-Y, Liu D, Zhao W-H. Combined effects of physical activity and calcium on bone health in children and adolescents: a systematic review of randomized controlled trials. World J Pediatr. 2020;16:356–65.
6. Orsso CE, Silva MIB, Gonzalez MC, Rubin DA, Heymsfield SB, Prado CM, et al. Assessment of body composition in pediatric overweight and obesity: a systematic review of the reliability and validity of common techniques. Obes Rev. 2020;21:e13041.
7. Palenzuela-Luis N, Duarte-Clíments G, Gómez-Salgado J, Rodríguez-Gómez JÁ, Sánchez-Gómez MB. Questionnaires assessing adolescents' self-concept, self-perception, physical activity and lifestyle: a systematic review. Children. 2022;9:91.
8. Sember V, Jurak G, Kovač M, Morrison SA, Starc G. Children's physical activity, academic performance, and cognitive functioning: a systematic review and meta-analysis. Front Public Health. 2020;8:307.
9. Marquez DX, Aguiñaga S, Vásquez PM, Conroy DE, Erickson KI, Hillman C, et al. A systematic review of physical activity and quality of life and well-being. Transl Behav Med. 2020;10:1098–109.
10. Guimarães JP, Fuentes-García JP, González-Silva J, Martínez-Patiño MJ. Physical activity, body Image, and its relationship with academic performance in adolescents. Healthcare. 2023;11:602.
11. Muntaner-Mas A, Morales JS, Martínez-de-Quel Ó, Lubans DR, García-Hermoso A. Acute effect of physical activity on academic outcomes in school-aged youth: a systematic review and multivariate meta-analysis. Scand J Med Sci Sports. 2024;34:e14479.
12. Visier-Alfonso ME, Sánchez-López M, Álvarez-Bueno C, Ruiz-Hermosa A, Nieto-López M, Martínez-Vizcaíno V. Mediators between physical activity and academic achievement: a systematic review. Scand J Med Sci Sports. 2022;32:452–64.
13. World Health Organisation. WHO guidelines on physical activity and sedentary behaviour. 2020. https://www.who.int/publications/i/item/9789240015128. Accessed 28 Dec 2023.

14 Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1.6 million participants. Lancet Child Adolesc Health. 2020;4:23–35.
15. Liu Y, Ke Y, Liang Y, Zhu Z, Cao Z, Zhuang J, et al. Results from the China 2022 report card on physical activity for children and adolescents. J Exerc Sci Fit. 2023;21:1–5.
16. Katzmarzyk PT, Barreira TV, Broyles ST, Champagne CM, Chaput J-P, Fogelholm M, et al. Relationship between lifestyle behaviors and obesity in children ages 9–11: results from a 12-country study. Obesity. 2015;23:1696–702.
17. Owen CG, Nightingale CM, Rudnicka AR, Sattar N, Cook DG, Ekelund U, et al. Physical activity, obesity and cardiometabolic risk factors in 9- to 10-year-old UK children of white European, South Asian and black African-Caribbean origin: the child heart and health study in England (CHASE). Diabetologia. 2010;53:1620–30.
18. Zhu Z, Tang Y, Zhuang J, Liu Y, Wu X, Cai Y, et al. Physical activity, screen viewing time, and overweight/obesity among Chinese children and adolescents: an update from the 2017 physical activity and fitness in China—the youth study. BMC Public Health. 2019;19:1–8.
19. Tarasenko YN, Howell BM, Studts CR, Strath SJ, Schoenberg NE. Acceptability and feasibility of physical activity assessment methods for an Appalachian population. J Community Health. 2015;40:714–24.
20. Vanhelst J. Quantification de l'activité physique par l'accélérométrie. Rev DÉpidémiologie Santé Publique. 2019;67:126–34.
21. Marasso D, Lupo C, Collura S, Rainoldi A, Brustio PR. Subjective versus objective measure of physical activity: a systematic review and meta-analysis of the convergent validity of the physical activity questionnaire for children (PAQ-C). Int J Environ Res Public Health. 2021;18:3413.
22. Eckard ML, Kuwabara HC, Van Camp CM. Using heart rate as a physical activity metric. J Appl Behav Anal. 2019;52:718–32.
23. van Hees VT, Thaler-Kall K, Wolf K-H, Brønd JC, Bonomi A, Schulze M, et al. Challenges and opportunities for harmonizing research methodology: raw accelerometry. Methods Inf Med. 2016;55:525–32.
24. Montoye AHK, Moore RW, Bowles HR, Korycinski R, Pfeiffer KA. Reporting accelerometer methods in physical activity intervention studies: a systematic review and recommendations for authors. Br J Sports Med. 2018;52:1507–16.
25. Erdim L, Ergün A, Kuğuoğlu S. Reliability and validity of the Turkish version of the physical activity questionnaire for older children (PAQ-C). Turk J Med Sci. 2019;49:162–9.
26. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport. 2000;71:1–14.
27. Ndahimana D, Kim E-K. Measurement methods for physical activity and energy expenditure: a review. Clin Nutr Res. 2017;6:68.
28. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. Prog Cardiovasc Dis. 2015;57:387–95.
29. Chinapaw MJM, Mokkink LB, van Poppel MNM, van Mechelen W, Terwee CB. Physical activity questionnaires for youth: a systematic review of measurement properties. Sports Med. 2010;40:539–63.
30. Kowalski KC, Crocker PRE, Kowalski NP. Convergent validity of the physical activity questionnaire for adolescents. Pediatr Exerc Sci. 1997;9:342–52.
31. Martínez-Gómez D, Martínez-de-Haro V, Pozo T, Welk GJ, Villagra A, Calle ME, et al. Fiabilidad y validez del cuestionario de actividad física PAQ-A en adolescentes españoles. Rev Esp Salud Pública. 2009;83:427–39.
32. Andarge E, Trevethan R, Fikadu T. Assessing the physical activity questionnaire for adolescents (PAQ–A): specific and general insights from an Ethiopian context. BioMed Res Int. 2021;2021:e5511728.
33. Bervoets L, Van Noten C, Van Roosbroeck S, Hansen D, Van Hoorenbeeck K, Verheyen E, et al. Reliability and validity of the Dutch physical activity questionnaires for children (PAQ-C) and adolescents (PAQ-A). Arch Public Health. 2014;72:1–7.
34. Koh D, Zainudin NH, Zawi MK. Validity and reliability of the modified physical activity questionnaire for adolescents (PAQ-A) among Malaysian youth. Int J Hum Mov Sports Sci. 2020;8:355–60.
35. Pratanaphon S, Longlalerng K, Jitmana J, Chaikla K, Nankanya T, Pirunsan U, et al. Content validity and psychometric characteristics of the Thai translated version of the physical activity questionnaire for children (PAQ-C) and adolescents (PAQ-A). J Assoc Med Sci. 2020;53:84–91.
36. Rahayu A, Sumaryanti S, Arovah NI. The validity and reliability of the physical activity questionnaires (PAQ-A) among Indonesian adolescents during online and blended learning schooling. Phys Educ Theory Methodol. 2022;22:173–9.
37. Wyszyńska J, Matłosz P, Podgórska-Bednarz J, Herbert J, Przednowek K, Baran J, et al. Adaptation and validation of the physical activity questionnaire for adolescents (PAQ-A) among Polish adolescents: cross-sectional study. BMJ Open. 2019;9:e030567.
38. Li X, Wang Y, Li X, Li D, Sun C, Xie M, et al. Reliability and validity of physical activity questionnaire for adolescents (PAQ-A) in Chinese version [青少年体力活动问卷(PAQ-A) 中文版的修订及信效度研究]. J Beijing Sport Univ. 2015;28:63–7.
39. Dong X, Ding M, Chen W, Liu Z, Yi X. Relationship between smoking, physical activity, screen time, and quality of life among adolescents. Int J Environ Res Public Health. 2020;17:8043.
40. Dong X, Ding L, Zhang R, Ding M, Wang B, Yi X. Physical activity, screen-based sedentary behavior and physical fitness in Chinese adolescents: a cross-sectional study. Front Pediatr. 2021;9:722079.
41. Guo D. Factors affecting physical activity of Chinese adolescents in the perspective of social ecological modelling [社会生态学模型视域下我国青少年身体活动影响因素研究] [Doctoral dissertation]. Tianjin: Tianjin University of Sport; 2022.
42. Guo K. Study on the relationship between physical education environment, exercise intentions and physical activity of junior high school students [学校体育环境、锻炼意向与初中生身体活动的关系研究] [Doctoral dissertation]. Shanghai: Shanghai University of Sport; 2019.
43. Li X, Li X, Wang Z. Assessment of physical activity level and the influencing factors among middle school students in the main urban area of Beijing [北京市主城区中学生身体活动水平及其影响因素]. Chin J Sch Health. 2018;39:993–996.
44. Zhang Z. Analysis of physical activity levels and influencing factors of children and adolescents in Nanjing municipality [南京市儿童青少年体力活动水平及影响因素分析]. Chin J Sch Health. 2018;39:1885–8.
45. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health. 2018;6:149.
46. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34–42.
47. Comrey AL, Lee HB. A first course in factor analysis. 2nd ed. USA: Psychology Press; 2013.
48. Danielsoper. Free a-priori sample size calculator for structural equation models - free statistics calculators. https://www.danielsoper.com/statcalc/calculator.aspx?id=89. Accessed 22 Feb 2024.
49. Kowalski KC, Crocker PRE, Donen RM. The physical activity questionnaire for older children (PAQ-C) and adolescents (PAQ-A) manual. Saskatoon: University of Saskatchewan; 2004.
50. Hurley AE, Scandura TA, Schriesheim CA, Brannick MT, Seers A, Vandenberg RJ, et al. Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives. J Organ Behav. 1997;18:667–83.
51. Tharwat A. Principal component analysis-a tutorial. Int J Appl Pattern Recognit. 2016;3:197–240.
52. Meyers LS, Gamst G, Guarino AJ. Applied multivariate research: design and interpretation. 3rd ed. USA: Sage publications; 2016.
53. Nunnally JC. An overview of psychological measurement. In: Wolman BB, editor. Clinical Diagnosis of Mental Disorders: A Handbook. Boston, MA: Springer, US; 1978. p. 97–146.
54 Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 7th ed. Upper Saddle River, NJ, USA: Pearson Education International; 2010.
55. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. Eur Bus Rev. 2019;31:2–24.
56. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res. 1981;18:39–50.
57. Huang D, Yang Q, Zhang X. Dilemma of reform and development path of school physical education in China under the strategy of "healthy China" ["健康中国"战略下我国学校体育的改革困境与发展路径]. Sports Cult Guide. 2018;3:103–107.
58. Lam LW. Impact of competitiveness on salespeople's commitment and performance. J Bus Res. 2012;65:1328–34.
59 Setyowati A, Chung M-H, Yusuf A, Haksama S. Psychometric of the curiosity and exploration inventory-ii in Indonesia. J Public Health Res. 2020;9(3):1745.

60. Hair JF, Alamer A. Partial least squares structural equation modeling (PLS-SEM) in second language and education research: guidelines using an applied example. Res Methods Appl Linguist. 2022;1:100027.
61. Polit DF, Beck CT. Nursing research: principles and methods. Philadelphia: Lippincott Williams & Wilkins; 2004.
62. Chen F. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equ Model Multidiscip J. 2007;14:464–504.
63. Rutkowski L, Svetina D. Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. Educ Psychol Meas. 2014;74:31–57.
64. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equ Model Multidiscip J. 2002;9:233–55.
65. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw Hill; 1994.
66. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Dev Rev. 2016;41:71–90.
67. Bajamal E, Robbins LB. Psychometric properties of the Arabic version of physical activity questionnaire for adolescents (PAQ-A). Appl Nurs Res. 2023;69:151660.
68. Janz KF, Lutuchy EM, Wenthe P, Levy SM. Measuring activity in children and adolescents using self-report: PAQ-C and PAQ-A. Med Sci Sports Exerc. 2008;40:767–72.
69. Voss C, Dean PH, Gardner RF, Duncombe SL, Harris KC. Validity and reliability of the physical activity questionnaire for children (PAQ-C) and adolescents (PAQ-A) in individuals with congenital heart disease. PLoS ONE. 2017;12:e0175806.
70. Schumaher MLN, Milani D, Alexandre NMC. Psychometric properties evaluation of the psychological empowerment instrument in a Brazilian context. J Nurs Manag. 2019;27:404–13.
71. Hyde JS. The gender similarities hypothesis. Am Psychol. 2005;60:581–92.
72 de Souza AC, Alexandre NMC, de Brito Guirardello E. Psychometric properties in instruments evaluation of reliability and validity. Epidemiol E Serviços Saúde. 2017;26:649–59.

Publisher's Note