

Content validity and test-retest reliability with principal component analysis of the translated Malay four-item version of Paffenbarger physical activity questionnaire

Fazlisham Binti Ghazali and Siti Nurhafizah Saleeza Ramlee
OSHA, College of Medical Sciences, Cyberjaya University, Cyberjaya, Malaysia

Najib Alwi

Management and Science University, Shah Alam, Malaysia, and

Hazuan Hizan

Sultan Idris Education University, Tanjong Malim, Malaysia

Abstract

Purpose – This study aimed to develop the construct validity for the Malay version of the Paffenbarger physical activity questionnaire (PPAQ) by adapting the original questionnaire to suit the local context.

Design/methodology/approach – The PPAQ was adopted and translated into the Malay language and modified to reach good content agreement among a panel of experts. A total of 65 participants aged 22–55 years old, fluent and literate in the Malay language were selected. Principal component analysis (PCA) was used to investigate construct validity. Reliability of this adapted instrument was analyzed according to types of variables.

Findings – The panel of experts reached a consensus that the final four items chosen in the adapted Malay version of PPAQ were valid and supported by a good content validity index (CVI). In total, two domains consonant with the operational domain definition were identified by PCA. Based on scores from intensity and duration of exercise, the study further divided the group into who were physically active and those who chose the unstructured physical activity. Relative reliability after a 14-day interval demonstrated moderate strength of agreement with an acceptable range of measurement error.

Research limitations/implications – PPAQ has been used worldwide but was less familiar in the local context. The Malay-four item PPAQ will provide the locally validated version of physical activity questionnaire. In addition, the authors have improved the original PPAQ by dividing the question items into two distinct domains which will effectively identify those who are physically active and those who are involved in unplanned exercise. Nevertheless, further research is recommended in bigger and heterogeneous samples along with a number of reliability tests.

Practical implications – To the authors' knowledge, this is the first study to assess internal structure of the four-item version of PPAQ. This analysis successfully identified two components with eigenvalue more than one in the Malay four-item PPAQ. Based on this, the authors were able to separate pool of population into two groups, which are physically active and unplanned exercise (involved in unstructured exercise). The ability of the validated

© Fazlisham Binti Ghazali, Siti Nurhafizah Saleeza Ramlee, Najib Alwi and Hazuan Hizan. Published in *Journal of Health Research*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

Declaration of conflicting interests: The author(s) declared no potential conflicts of interest concerning the research, authorship and/or publication of this article.

Funding support for this study was provided by a grant from the Cyberjaya University College of Medical Sciences: Grant number CRG /01/03/2018.



questionnaire to divide the population into various intensities of physical activity is a novel one, which may be useful in many public health studies where high intensity of physical activity; hence, greater energy expenditure is associated with increased longevity, better health benefit and improved cognitive function.

Social implications – In addition, the second domain “unplanned exercise” was successfully grouped together. Implication of the unplanned exercise component is to identify pool of population with active lifestyle awareness and choose the unstructured exercise instead of vigorous and formal exercising. Even though the amount of intensity and duration of incidental exercise does not reach recommended public health recommendation, it has been proven that preferred healthier lifestyle is positively associated with better cognition in later life.

Originality/value – The adapted Malay version of PPAQ has sound psychometric properties and could assist in differentiating groups of population based on their physical activity.

Keywords Paffenbarger physical activity, Malay-translated physical activity, Test–retest reliability, Content validity, Principal component analysis

Paper type Research paper

Introduction

Producing an accurate measurement of physical activity is important for detecting important health associations or effects. Moreover, the choice of an appropriate physical activity measurement tool depends upon the application for which it is intended [1]. We aimed to develop a reliable tool for physical activity measurement to be adapted to primary care in the Malaysian setting.

The Paffenbarger physical activity questionnaire (PPAQ) has been developed to suit the changing terms and guidelines for physical health. The PPAQ was developed by Dr. Ralph Seal Paffenbarger to assess physical activity via questionnaires [2]. Since then, it has been extensively tested for its reliability and validity in large population studies. The current format of PPAQ consists of eight questions that measure not only sedentary lifestyle but also energy expenditure through a physical activity index [3]. A recent study showed that PPAQ is more adept at capturing vigorous activity as it uses more descriptive terms and proper physiological definitions of physical activity intensity [4].

This study aimed to translate and validate the PPAQ which has been used in the Common Cold Project [5] to provide a reliable questionnaire to measure the level of physical activity adapted to the local primary care setting.

Methodology

Study design

To validate the questionnaire, a cross-sectional study was conducted in selected private hospitals in the area of Hulu Langat, Selangor, Malaysia. A total of 65 participants who were staff at the respective hospitals and were literate and fluent in the Malay language were selected using convenient sampling. Subjects were constructed to answer the modified Malay version of PPAQ which took about 15–20 minutes to complete. All recruited participants gave consent prior to completing the questionnaire on two occasions, 14 days apart.

Ethical consideration

This study was approved by the Ethics Committee of the Cyberjaya University College of Medical Sciences (Reference number CUCMS/CRERC/FR/023). Permission to carry out the research was granted by the General Manager and Chief Executive Officer of the respective hospitals.

Sample size

The sample size calculation for this study was based on the suggestion by Viechtbauer [6], for studies of similar nature.

$$n = \frac{\ln(1-y)}{\ln(1-\pi)}$$

where

n = number of sample size,

y = confidence level (95%) and

π = probability for non-responses to occur (0.05).

It was anticipated that problems that might occur would be minor such as nonresponses or item misinterpretation. Hence, it was decided that, if such difficulties are presented themselves with at least $\pi = 0.05$ probability (i.e. in at least 1 out of 20 participants), it would be good to detect this problem during the validation process. Accordingly, from the above equation, 60 participants needed to be screened to achieve 95% confidence that one or more such problem cases would be encountered.

Measure and procedure

Instrument modification and operational domain definition. It was imperative that the translated version of measurement was clear to respondents, and they perceived the same meaning as what researchers intended to achieve from the questionnaire. Therefore, in this adapted questionnaire, the content was developed and forward translated to Malay through an expert review.

Question 1. In this section, participants were asked if they engaged in any REGULAR physical activity that was long enough to work up a sweat. If the answer was yes, the next question requested them to detail the number of times per week. Physical activity was defined as any bodily movement produced by skeletal muscle that required energy. Translated into the Malay language this was “*Aktivitifizikaldidefinasikansebagai pergerakan badan yang memerlukan tenaga*”. The word skeletal muscle translated into “*ototrangka*” in Malay was removed for its rarity of medical term usage among the nonmedical Malay population.

The word REGULAR in the original Paffenbarger questionnaire was replaced with more specific term that is “engaged at least once a week,” which was translated into “*sekurang-kurangnyasekaliseminggu*” in Malay. By establishing consistency and frequency in a week, it would be possible to identify the physically active compared to the more sedentary.

Sweating is commonly associated with physical endurance with a significant linear relationship between sweat excretion and physical intensity [7]. Sweating sooner or more profusely has been a good indicator of physical activity intensity [8]. A question to assess the physical activity that induces sweating would identify those who are physically fit.

Questions 2 and 3. These questions assessed the subject’s lifestyle by identifying how many stairs they climbed up each day and the distances they walked on average. Sperandio’s study showed that walking less than 500 meters per week was the best predictor of physical inactivity [9] as it provided a better research metric for epidemiologic research and better public health targets than walking duration [10].

On the other hand, climbing stairs, an underrated exercise, has been proven to benefit an individual’s health [11, 12] and predicted longevity [13] as well as lowering blood pressure and improving fitness [14]. There is no universal consensus about the ideal number of stairs, but 8900–9900 steps per week are recommended [15].

Question 4.

(1) *Seven day recall*

This question was about sports or recreational activity in the past week. The seven-day recall contrasted with the original Paffenbarger physical activity question requesting details of such activity in the past year. Due to limitations in human memory, it was deemed best to keep the reporting interval relatively short.

Kjellsson's experiment showed that the overall level of recall error increased with the length of the recall period [16]. Masse and de Niet's literature reviews showed that seven-day recall can be validly ranked to identify those who are physically active and is sensitive enough to detect changes in physical activity behavior [17]. Therefore, this modified questionnaire required only a short-term recall of seven days.

(2) *Restriction to sports and recreational activities*

In this study, the specific activity of any sports or recreational activities was used as a heading under question 4.

Question 5 and 6. The effectiveness of public health campaigns depends on people to know the intensity, duration and frequency of physical activity performed [18]. The WHO suggested adults aged 18–64 should do at least 150 minutes of physical activity at moderate intensity or 75 minutes at vigorous intensity throughout the week to achieve the desired health outcomes. Also, for reasons of practicality, raw data of all components of this complex behavior which include the type (intensity), duration and frequency of physical activity were converted into energy expenditure, i.e. the metabolic equivalent of tasks (METs). Therefore, under questions 5 and 6, we asked about the frequency and duration of physical activity performed.

Translation and back translation. It is imperative that the translated version of measurement was well understood by respondents, and that they perceived the same meaning as what researchers intended to achieve from the questionnaire. Hence, the modified Malay version of the PPAQ was translated into the Malay language by a sports scientist (HH) who was also well versed in both the Malay and English languages. The questionnaire was then back translated into English by an independent professional translator. Another independent professional translator reviewed the back-translated version against the original PPAQ and concluded that no further modification was necessary.

An accredited professional translator then checked the Malay translated version to ensure the terms used were correct and culturally appropriate. The final Malay version was harmonized for any language errors by all the experts until an acceptable translation was developed.

Content validity. A total of four professional bilingual senior sports science lecturers with over five years' research study in the English language medium were requested to determine if the items fully and sufficiently represented the targeted domain.

All four specialists were initially contacted by email and phone. They were provided with a formal invitation letter from Cyberjaya University, including details of the research and instructions. Attached to this was a set of questionnaires in the Malay language with an empty box for them to score each domain on a Likert scale.

The four experts rated the content validity of each test in relation to the five tasks in the rating protocol. The scale was scored as follows: 1 = test not being relevant; 2 = somewhat relevant; 3 = quite relevant and 4 = highly relevant. Grades 3 and 4 were considered acceptable. Apart from assessing the content, the four experts were invited to comment in more detail in boxes on the side of each question.

Subjects understanding of the modified questionnaire (cognitive interview). The final Malay version was pretested on ten respondents randomly picked from the public who fulfilled the criteria of being fair-minded and literate in Malay. They were aged between 20 and 40 years old with an equal mix of genders. The objective was to identify any words and grammatical errors that might affect the comprehension of the respondents. This also included an examination of respondents' cognitive ability to recall the information and assessment of the format and wording to elicit appropriate responses and whether respondents gave socially desirable answers.

Subjects were instructed to share their thoughts about each question and to describe their thought processes before answering each question. Participants were also invited to suggest

alternative wording or sentences if they wished. In this session, the examiner read out the questions, and the subjects were answered with minimal interference from the examiner.

At the end of the session, participants were requested to provide more feedback about the length of questions and their clarity. All ten participants agreed that the questions were reasonable, and they were able to recall events pertinent to the questions asked.

Test-retest reliability. Participants were informed that they were required to complete the questionnaire twice at 14 days apart. The researcher was present during the completion period to assist participants if required. All 65 volunteers completed the test-retest assessment.

Analyses and results

Sociodemographic data

A total of 65 respondents ranging from 22 to 55 years old were with mean and standard deviation (SD) of 29.49 and 5.54 years, respectively. Females and Malays were dominant in gender and ethnicity. Most participants had studied beyond secondary school with 47.9 % studying beyond further education level for further 3.5 years.

Statistical analysis

In content validity tests, the initial content validity index (CVI) was used to analyze agreement between four experts judging the relevance of question items used. Further, construct validity test, principal component analysis (PCA) was done by using SPSS version 19. Where in the reliability test, the analysis was divided into two, i.e. analysis of continuous and categorical data; for continuous data, intraclass correlation coefficient (ICC), paired *t*-test and Bland-Altman plot were used to examine the agreement between two tests at two different times. In addition, standard error mean (SEM), minimal detectable change (MDC) and minimal important difference (MID) were used to demonstrate the absolute reliability of the questionnaire. Moreover, agreement between categorical data was observed from weighted kappa.

Validity test. Content validity index. A panel of four experts reached a consensus that the final items in all six questions were valid to be used. An item-level CVI (I-CVI) was computed by dividing the total number of experts giving a rating of 3 or 4 (relevant) by the total number of experts in which all items scored a rating of 1, as presented in [Table 1](#).

Construct validity. Construct validity was done using confirmatory and exploratory factor analysis with a factor loading of 0.4 or more considered good.

(1) KMO and Bartlett's test of sphericity

Bartlett's test of sphericity resulted in 0.707, which reached statistical significance, supporting the factorability of the correlation matrix [19]. The null hypothesis could be rejected, and the alternate hypothesis that there may be a statistically significant interrelationship between variables was accepted. Hence, factor analysis was considered as an appropriate technique for further analysis of the data.

(2) Confirmatory and exploratory factor analysis

From this analysis, [Table 2](#), two components have been identified with eigen values of more than 1.0 suggesting that dividing the questionnaire into two components was most appropriate.

In further analysis, orthogonal rotation (varimax) was to delineate further the two components with an assumption that what was explained by one factor was independent of information from other factors. Factor rotation made it easier for further interpretation of components.

Rotated component matrix sorted six variables into two overlapping groups each with a loading factor of 0.4 or more. There were blanks in the matrix where weights were

Table 1.
Content validity index
of Malay
version PPAQ

Question item (Malay version)	Question item (English version)	Expert 1	Expert 2	Expert 3	Expert 4	Content validity index (CVI)- item level
Q1. Adakah anda membeli atau melakukan mana-mana aktiviti fizikal seperti berjalan pantas... aktiviti yang mengeluarkan peluh?	Q1. Do you engage in any regular physical activity like brisk walking, i.e. long enough to work up a sweat?	3	4	4	4	4/4 = 1.00
Q2. Berapakah jumlah anak tangga yang anda naiki pada setiap hari?	Q2. How many stairs do you climb up every day?	3	4	4	4	4/4 = 1.00
Q3. Berapakah jumlah berjalanan dalam purata setiap hari?	Q3. How much of a distance do you walk per day?	3	4	4	3	4/4 = 1.00
Q4. Senaraikan aktiviti sukatan rekreasi yang anda ambil bahagian dalam tempoh masa seminggu yang lepas. Kami hanya berminat dengan aktiviti yang aktif	Q4. List down any sport or recreational activities you participated in during the past week. We are only interested in the time you were physically active	3	4	4	4	4/4 = 1.00
Q5. Berapa kali dalam seminggu anda menjalankan aktiviti tersebut?	Q5. How frequent do you perform the activity in one week?	4	4	4	4	4/4 = 1.00
Q6. Purata masa setiap aktiviti?	Q6. How long do you do the activity per session?	4	4	4	4	4/4 = 1.00

less than 0.4 (Table 3). The factor column represented the rotated factors that were extracted out of the total factor. These are the core factors, which will be used as the final factor after data reduction.

The first component suggests that the mode of intensity and duration is highly correlated with each other, which explains about 51% of the variability in the performance of this physical activity questionnaire. The second component consisting of the number of stairs climbed and walking distance per day explained 22% of variance from PCA, as presented in Table 2. Surprisingly, the second component successfully delineated the two question items together which comprised the stairs climbed per day and walking distances per day with a higher loading factor.

(3) Internal consistency test

Cronbach's alpha was used to measure the internal consistency of the scale. As from the factor analysis calculated earlier, two components were extracted out from this scale (Tables 2 and 4).

In this analysis, all items in component 1 had a moderately high corrected item scale correlation. On the other hand, there was no correlation at all between climbing stairs and walking distances in the second component which is expected considering that both questions were not considered related to each other. The final Malay version PPAQ kept all the questions in view that it makes clinical sense to retain them in the respective components.

Reliability test. An agreement between continuous data, i.e. walking distances per day and stairs climbed per day of Malay version PPAQ at two different times of measurement were analyzed using ICC (two-way random effects, absolute agreement and single rater) for relative reliability, paired t-test and Bland–Altman diagram for systemic bias. Furthermore, the Bland–Altman plot was useful to provide the limit of agreement and to detect outliers possibly caused by errors of measurement [20]. SEM, MDC and MID were used to estimate minimal scores that are not due to error [21]. In contrast, categorical data in this reliability test were examined by using weighted kappa, which is more helpful to provide strength of agreement between two measures.

Component	Total	Initial eigen values		Extraction sums of squared loadings		
		% of variance	Cumulative %	Total	% of variance	Cumulative %
1	3.035	50.581	50.581	3.035	50.581	50.581
2	1.349	22.488	73.069	1.349	22.488	73.069
3	0.571	9.516	82.585			
4	0.477	7.944	90.529			
5	0.323	5.383	95.912			
6	0.245	4.088	100.000			

Table 2. Principle component analysis of Malay version PPAQ

	Component 1 Structured physical activity	Component 2 Unstructured physical activity
Malay version: intensity of physical activity Q4	0.919	
Malay version: duration of physical activity Q6	0.870	
Malay version: frequency of physical activity Q5	0.752	
Malay version: involves in any physical activity Q1	0.544	0.456
Malay version: climbing stairs/day Q2		0.852
Malay version: walking distances/day (m) Q3		0.797

Table 3. Rotated component matrix: factor loadings (>0.4) for Malay version PPAQ

Note(s): Rotation method: varimax with Kaiser normalization
a. Rotated converged in three iterations

Table 4.
Cronbach's alpha on each component and their proposed names for Malay version PPAQ

Component	Cronbach's alpha	Items	Cronbach's alpha if item deleted	Corrected item-total correction
Physical activity (component 1)	0.774	Intensity of physical activity (Q4)	0.591	0.853
		Duration of physical activity (Q6)	0.732	0.665
		Frequency of physical activity (Q5)	0.719	0.578
		Involve in any physical activity(Q1)	0.799	0.468
Unplanned exercise (component 2)	0.083	Climbing stairs/day (Q2)	0.002	0.454
		Walking distances /day (meters)(Q3)	0.420	0.454

Determine relative reliability for continuous data.

- (1) Intraclass correlation coefficient (ICC)

Table 3 shows the acceptable test–retest reliability with ICC ranging 0.534 to 0.623 for both climbing stairs and walking distances per day.

Determine systemic bias for continuous data.

- (1) Paired *t*-test

We also found that there were no significant differences for means at 14 days interval with both *p*-value > 0.05 and agreed not to reject the null hypothesis that there was no statistically significant difference between the two tests.

- (2) Bland–Altman plot

Potential error of measurement was further analyzed by using the Bland–Altman diagram which addresses if there is any systematic difference between two sets of measurements as well as to identify possible outliers (see Figure 1).

Each sample was represented on the graph by conveying the mean value of the two assessments (x-axis) and the difference between the two assessments (y-axis). The mean difference was the estimated bias, and the SD of the differences measured the fluctuations around this mean (outliers being above 1.96 SD difference).

Determine absolute reliability for continuous data.

- (1) Standard error of mean (SEM) and minimal detectable change (MDC₉₀)

The findings demonstrate that although test–retest reliability (relative reliability) for the clinical tests was excellent, there was still a substantial degree of variability of performance for individual participants from one test session to the next (absolute reliability). The SEM and MDC₉₀ were calculated to objectify these findings.

SEM was calculated based on the formula:

$$SEM = SD_{\text{difference mean}} \times \sqrt{(1 - ICC)}$$

In accordance, SEM was based on the assumption of normal distribution, and probabilities of the normal curve could be applied to SEM values. In total, 68% probability that repeated questions for climbing stairs and walking distances will be within ±37.7 and ± 572.6 of the mean score on the first day of assessment, respectively. Thus, a 96% probability that repeated measures will give the values ranging ±75.4 for climbing stairs per day (2 × SEM) and ±1145.2 meters per day for walking distances (2 × SEM).

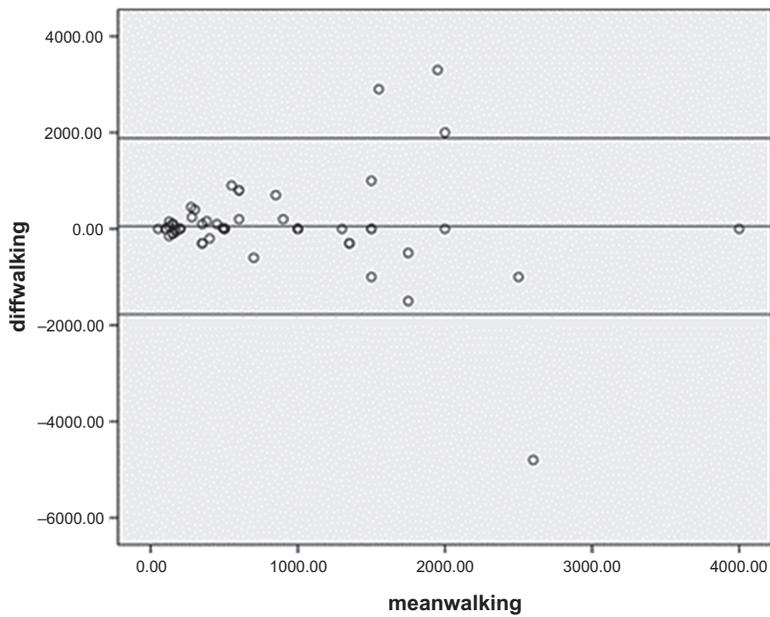
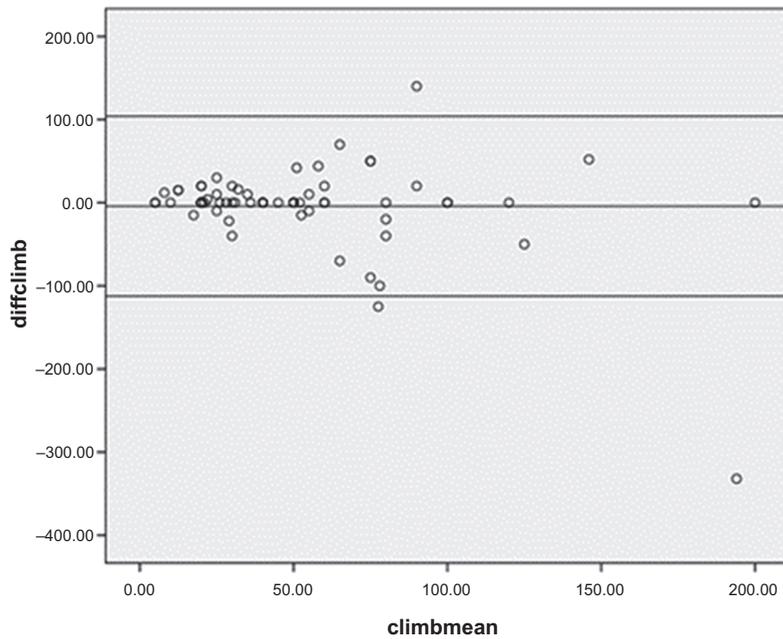


Figure 1.
Bland-Altman plots
showing limits of
agreement between
two sets of
measurement at
14 days interval for
numbers of stairs
climbed per day (left-
hand side) and distance
of walking per day
(right-hand side)

A 90% value as a confidence interval for MDC was used. Using the formula of $MDC_{90} = SEM \times 1.645$ (Z score at 90% confidence interval) $\times \sqrt{2}$, the resulted values for MDC_{90} are shown in Table 3. Both MDC_{90} values for climbing stairs and walking distances were out of range from changes of means across the two-time points. The overlapped test-retest scores with the interval of MDC_{90} value indicated that the changes were likely due to random measurement error.

(2) Minimal Important difference (MID)

This study is the first to determine the measurement error of PPAQ which is an indication of the accuracy of the measurement instrument. COSMIN guidelines proposed that the interpretation of SEM should be based on the value of the MID [22]. However, the true purpose of MID which is to represent the smallest change in score that is considered a relevant outcome is not going to be utilized. Instead, MID was to assess statistical reliability, i.e. measurement errors relying on other statistical measures like SD, SEM and effect size [23]

1SEM: $1 \times$ standard error of means; SD: standard deviation and ICC: intraclass correlation coefficient

Determine reliability of categorical data.

(1) Weighted kappa

The observed percentage of agreement implies the proportion of ratings where the raters agree, and the expected percentage is the proportion of agreements that is expected to occur by chance as a result of the raters scoring randomly. Hence, kappa is the proportion of agreements that is observed between raters, after adjusting for the proportion of agreements that takes place by chance [24].

By using the formula of

$$\text{Kappa}(K) = \frac{P_o - P_c}{1 - P_c}$$

where P_o = observed agreement and

P_c = proportion of agreement by chance.

We were able to generate values of kappa as shown in Tables 5–7.

Many scholars agreed that it is important to retain the hierarchical nature of the categories.

Therefore, further analysis of the ordinal data, weighted kappa was used to reflect the degree of agreement in terms of their seriousness, as shown in Table 5. In this analysis, quadratic weighting was preferred over linear as the variation coefficients of the former increases with the number of categories, which will be a more desirable weighting scheme given the hierarchical nature of categories.

Discussion

This study aimed to translate PPAQ into the Malay language. The Malay version PPAQ had good interrater reliability and internal structure. The panel of experts reached a consensus that the final items in both domains were valid to be used with item CVI reached a total mutual agreement.

To our knowledge, this is the first study to assess the internal structure of PPAQ. Our analysis successfully identified two components with eigen values more than one in the Malay version PPAQ. The ability of the validated questionnaire to divide the population into various intensities of physical activity is a novel one, which may be useful in many public

health studies where high intensity of physical activity; hence, greater energy expenditure is associated with increased longevity, better health benefit and improved cognitive function.

In addition, the second domain “unplanned exercise” was successfully extracted with Q2 and Q3 grouped under principal component analysis.

Analysis of measurement errors in this study was divided into two parts according to the type of variables. In continuous data, which is the unplanned exercise component, we found that the self-reported Malay version PPAQ has fair relative reliability within 14 days of interval.

Limitations of this study

The Malay version of PPAQ will provide a locally validated version of the physical activity questionnaire. Future studies in bigger and heterogeneous samples along with more reliability tests are encouraged to evaluate the validity of this instrument with more objective measures for example accelerometer as in this study; we only measured the reliability and content validity of the translated version for PPAQ. These future studies are particularly important in view of the limitations of subjective measurement to accurately identify those who need further recommendations for health activity.

Table 5. Average scores of day 1 and 14, paired *t*-test, intraclass correlation coefficient (ICC) with significant level $p \leq 0.001$, standard error of mean (SEM) and minimal detectable change (MDC) at 90% CI

Malay version PPAQ	D1 of test mean (SD)	D14 of test mean (SD)	Difference mean (SD)	<i>p</i> -value (paired <i>t</i> -test)	ICC ($p \leq 0.001$)	SEM	MDC ₉₀
Climbing stairs/ day	50.08 (40.14)	54.22 (56.23)	-4.138 (55.20)	0.188	0.534	37.7	88
Walking distances/ day	798.77 (843.96)	745.38 (930.94)	53.385 (932.49)	0.873	0.623	572.6	1337

Method	Mid calculation	Mid climbing stairs per day	Mid walking distances per day (meter)
1SEM	$SD_{\text{baseline}} \times \sqrt{1 - \text{ICC}}$	27	518
Empirical rule effect size	$0.08 \times 6 \times SD_{\text{difference}}$	26	448
Cohen's effect size	$0.5 \times SD_{\text{difference}}$	28	466
$0.5 \times SD$	$0.5 \times SD_{\text{baseline}}$	20	422

Table 6. Distribution-based estimates of the minimal importance difference (MID)

Category	Observed agreement (P_o)	Chance agreement (P_c)	Un weighted Cohen's kappa (95% CI)	Quadratic weighted kappa (95% CI)
Physical activity index score	0.778	0.561	0.494 (0.260–0.728)	0.613 (0.316–0.909)
Physical activity intensity	0.677	0.371	0.487 (0.306–0.667)	0.660 (0.374–0.945)
Physical activity frequency	0.723	0.380	0.5534 (0.378–0.729)	0.603 (0.161–0.1.00)
Physical activity duration	0.846	0.469	0.7103 (0.545–0.875)	0.778 (0.662–0.895)

Table 7. Proportions of agreement of physical activity index scores, physical intensity, frequency and duration

Conclusion

PPAQ instrument has been used worldwide but is less familiar in the local region of Malaysia. Lack of its translated version and psychometric analysis makes this study imperative as a starting point for further research. Our statistical analysis successfully identified and delineated two major components in accordance with our operational domain definition with fair internal consistency. Hence, the six items were compressed into a four-item questionnaire. Further research is recommended in bigger and heterogeneous samples along with more reliability tests.

References

1. Bassett DR Jr. Validity and reliability issues in objective monitoring of physical activity. *Res Q Exerc Sport*. 2000 Jun; 71(Suppl 2): 30-6. doi: [10.1080/02701367.2000.11082783](https://doi.org/10.1080/02701367.2000.11082783).
2. Lee IM, Matthews CE, Blair SN. The legacy of Dr. Ralph Seal Paffenbarger, Jr. - past, present, and future contributions to physical activity research. *Pres Counc Phys Fit Sports Res Dig*. 2009 Mar; 10(1): 1-8.
3. Paffenbarger RS Jr, Hyde RT, Wing AL, Hsieh CC. Physical activity, all-cause mortality, and longevity of college alumni. *N Engl J Med*. 1986 Mar; 314(10): 605-13.
4. Tonstad S, Herring P, Lee J, Johnson JD. Two Physical activity measures: Paffenbarger physical activity questionnaire versus Aerobics Center Longitudinal Study as predictors of adult-onset type 2 diabetes in a follow-up study. *Am J Health Promot*. 2018 May; 32(4): 1070-7. doi: [10.1177/0890117117725282](https://doi.org/10.1177/0890117117725282).
5. Carnegie Mellon University. The common cold project. Carnegie Mellon University. Pittsburgh cold study 3. [Cited 2018 Jan]. Available from: <https://www.cmu.edu/common-cold-project/measures-by-study/health-practices/physical-activity/index.html#cahsq>.
6. Viechtbauer W, Smits L, Kotz D, Budé L, Spigt M, Serroyen J, Crutzen R. A simple formula for the calculation of sample size in pilot studies. *J Clin Epidemiol*. 2015 Nov; 68(11): 1375-9. doi: [10.1016/j.jclinepi.2015.04.014](https://doi.org/10.1016/j.jclinepi.2015.04.014).
7. Holmes N, Miller V, Bates G, Zheo Y. The effect of exercise intensity on sweat rate and sweat sodium loss in well trained athletes. *J Sci Med Sport*. 2011; 14: e112. doi: [10.1016/j.jsams.2011.11.234](https://doi.org/10.1016/j.jsams.2011.11.234).
8. Shibasaki M, Crandall CG. Mechanisms and controllers of eccrine sweating in humans. *Front Biosci (Schol Ed)*. 2010 Jan; 2: 685-96. doi: [10.2741/s94](https://doi.org/10.2741/s94).
9. Sperandio EF, Arantes RL, Silva RPD, Matheus AC, Lauria VT, Bianchim MS, Romiti M, Gagliardi ARDT, Dourado VZ. Screening for physical inactivity among adults: the value of distance walked in the six-minute walk test. A cross-sectional diagnostic study. *Sao Paulo Med J*. 2016 Jan-Feb; 134(1): 56-62. doi: [10.1590/1516-3180.2015.00871609](https://doi.org/10.1590/1516-3180.2015.00871609).
10. Williams PT. Distance walked and run as improved metrics over time-based energy estimation in epidemiological studies and prevention; evidence from medication use. *PLoS ONE*. 2012; 7(8): e41906. doi: [10.1371/journal.pone.0041906](https://doi.org/10.1371/journal.pone.0041906).
11. Shenassa ED, Frye M, Braubach M, Daskalakis C. Routine stair climbing in place of residence and body mass index: a pan-European population based study. *Int J Obes (Lond)*. 2008 Mar; 32(3): 490-4. doi: [10.1038/sj.ijo.0803755](https://doi.org/10.1038/sj.ijo.0803755).
12. Meyer P, Kayser B, Kossovsky MP, Sigaud P, Carballo D, Keller PF, Eric Martin X, Farpour-Lambert N, Pichard C, Mach F. Stairs instead of elevators at workplace: cardioprotective effects of a pragmatic intervention. *Eur J Cardiovasc Prev Rehabil*. 2010 Oct; 17(5): 569-75. doi: [10.1097/HJR.0b013e3283338a4dd](https://doi.org/10.1097/HJR.0b013e3283338a4dd).
13. Lee IM, Paffenbarger RS Jr. Associations of light, moderate, and vigorous intensity physical activity with longevity. The Harvard Alumni Health Study. *Am J Epidemiol*. 2000 Feb; 151(3): 293-9. doi: [10.1093/oxfordjournals.aje.a010205](https://doi.org/10.1093/oxfordjournals.aje.a010205).

14. Andersen LL, Sundstrup E, Boysen M, Jakobsen MD, Mortensen OS, Persson R. Cardiovascular health effects of internet-based encouragements to do daily workplace stair-walks: randomized controlled trial. *J Med Internet Res*. 2013 Jun; 15(6): e127. doi: [10.2196/jmir.2340](https://doi.org/10.2196/jmir.2340).
15. Tudor-Locke C. Steps to better cardiovascular health: how many steps does it take to achieve good health and how confident are we in this number? *Curr Cardiovasc Risk Rep*. 2010 Jul; 4(4): 271-6. doi: [10.1007/s12170-010-0109-5](https://doi.org/10.1007/s12170-010-0109-5).
16. Kjellsson G, Clarke P, Gerdttham UG. Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. *J Health Econ*. 2014 May; 35: 34-46. doi: [10.1016/j.jhealeco.2014.01.007](https://doi.org/10.1016/j.jhealeco.2014.01.007).
17. Masse LC, de Niet JE. Sources of validity evidence needed with self-report measures of physical activity. *J Phys Act Health*. 2012 Jan; 9(Suppl 1): S44-55. doi: [10.1123/jpah.9.s1.s44](https://doi.org/10.1123/jpah.9.s1.s44).
18. Wicker P, Frick B. The relationship between intensity and duration of physical activity and subjective well-being. *Eur J Public Health*. 2015 Oct; 25(5): 868-72. doi: [10.1093/eurpub/ckv131](https://doi.org/10.1093/eurpub/ckv131).
19. Cerny BA, Kaiser HF. A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behav Res*. 1977 Jan 1; 12(1): 43-7.
20. Watson PF, Petrie A. Method agreement analysis: a review of correct methodology. *Theriogenology*. 2010 Jun; 73(9): 1167-79. doi: [10.1016/j.theriogenology.2010.01.003](https://doi.org/10.1016/j.theriogenology.2010.01.003).
21. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther*. 2006 May; 86(5): 735-43.
22. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, De Vet HC. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010 May; 19(4): 539-49. doi: [10.1007/s11136-010-9606-8](https://doi.org/10.1007/s11136-010-9606-8).
23. Copay AG, Subach BR, Glassman SD, Polly DW, Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*. 2007 Sep-Oct; 7(5): 541-6. doi: [10.1016/j.spinee.2007.01.008](https://doi.org/10.1016/j.spinee.2007.01.008).
24. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005 Mar; 85(3): 257-68.

Corresponding author

Fazlsham Binti Ghazali can be contacted at: adarhusni@gmail.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com