Dear Author/Editor,

Greetings, and thank you for publishing with SAGE. Your article has been copyedited, and we have a few queries for you. Please respond to these queries when you submit your changes to the Production Editor.

Thank you for your time and effort.

Please assist us by clarifying the following queries:

<table>
<thead>
<tr>
<th>No</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Please provide degree(s) for all authors.</td>
</tr>
<tr>
<td>2</td>
<td>Please check that all authors are listed in the proper order; clarify which part of each author’s name is his or her surname; and verify that all author names are correctly spelled/ punctuated and are presented in a manner consistent with any prior publications.</td>
</tr>
<tr>
<td>3</td>
<td>Please confirm whether the given conflict of interest statement is accurate.</td>
</tr>
<tr>
<td>4</td>
<td>Please confirm whether the given funding statement is accurate.</td>
</tr>
</tbody>
</table>
Reliability and Validity of the Malay International Physical Activity Questionnaire (IPAQ-M) Among a Malay Population in Malaysia

Anne Hin Yee Chu1 and Foong Ming Moy1

Abstract
The International Physical Activity Questionnaire (IPAQ) was developed to assess the physical activity patterns in populations. The authors aim to examine the reliability and validity of the Malay version of IPAQ (IPAQ-M). The IPAQ-M was self-administered twice at a 1-week interval to assess its test–retest reliability. Criterion validity was assessed between the IPAQ-M and a 7-day physical activity log (PA-Log). A total of 81 Malay adults participated in the study. Intra-class correlation coefficients (ICC), kappa (κ), correlation coefficients (ρ), and Bland–Altman plot were used for data analyses. The ICC scores revealed moderate to good correlations (ICC = 0.54-0.92; P < .001) on items categorized by intensities and domains and a κ of 0.73 for total activity. Validity results from the PA-Log were statistically significant (P < .001) across intensities and domains (ρ = 0.67-0.98). The IPAQ-M demonstrated good reliability and validity for the evaluation of physical activity among this Malay population.

Keywords
IPAQ, Malay, physical activity, reliability, validity

Introduction
Physical activity has been recognized internationally as a key factor contributing to the maintenance and improvement of health.1 There is substantial evidence showing that physical activity is an important determinant against the development of chronic diseases such as noninsulin-dependent diabetes mellitus, osteoporosis, stroke, hypertension, colon cancer, obesity, depression, and coronary heart disease.2-4 These chronic diseases were found to be more commonly developed in individuals who engage in little or no physical activity, as compared with those who engage regularly in physical activity.5

1University of Malaya, Kuala Lumpur, Malaysia

Corresponding Author:
Anne Hin Yee Chu, Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia
Email: annechu@siswa.um.edu.my
Physical inactivity was reported as the fourth leading risk factor for global mortality (6% of deaths globally). In Malaysia, there is a sharp rise in the obesity rate over the past decade, which was believed to be associated with a reported 60% of the physically inactive Malaysian adult population, with 55% men and 65% women reported as inactive. This phenomenon is contributing to the causal factors underlying noncommunicable diseases seen practically in all developing countries.

To support, promote, and monitor physical activity in the population, valid and reliable tools are needed. Questionnaires, diaries, or more objective measures such as accelerometers and pedometers are the commonly used instruments in assessing physical activity. Of all the measuring methods, questionnaires are the most widely used instruments in large-scale epidemiological studies owing to their relatively low cost.

One of the most important questionnaires developed by a multinational working group in the late 1990s is the International Physical Activity Questionnaire (IPAQ), which is available in a short form for surveillance and in a long form for collecting more detailed physical activity information, and both forms are available in a number of languages. Although the long form has been rigorously tested for reliability and validity in a number of countries, it has not been validated in Malay population. Therefore, our present study fills this gap by translating and validating a Malay version of IPAQ long form to be used in the Malay population.

**Methods**

Participants were a group of Malay employees working in a public university in Kuala Lumpur, Malaysia. Malays make up the largest ethnic group working in the campus, and they were native Malay speakers. This sample was recruited through convenient sampling (mail invitation) due to the high level of cooperation needed from the participants. Inclusion criteria for recruitment were the following: *(a)* aged 18 years and older, *(b)* self-identified race as Malay, *(c)* absence of physical illness or disabilities that would limit daily physical activities such as walking, and *(d)* the ability to read and write well enough to record physical activities. Written informed consent was given by all participants. Ethics clearance was obtained from the Medical Ethics Committee of the university (Reference Number: MEC 782.18).

Based on a previous study examining the reliability of the Chinese version of the IPAQ long form, the test–retest reliability (Intraclass correlation coefficient [ICC]) of all categories of activities (vigorous, moderate, walking, and total physical activity [PA]) ranged from 0.74 to 0.97. Using “Sample Size Tables for Clinical Studies” (Table 14.5: Sample sizes required to observe a given ICC, $\rho$, using the confidence interval [CI] approach, for an assumed ICC = 0.75, a desired 95% CI width of 0.2), 75 subjects were required for reliability test. As for the sample size calculation of validity test, based on a study on the assessment on the validity of IPAQ long form, Spearman correlation coefficient ($\rho$) was found to range from 0.47 to 0.67 for home/garden domain, leisure domain, work domain, and total PA. Using “Sample Size Tables for Clinical Studies” (Table 12.1: Sample sizes for detecting a statistically significant correlation coefficient, with a correlation coefficient ($\rho$) of 0.4, significant level of .05, and power of 0.80), 47 subjects were required.

The participants were instructed to fill in the Malay version of the IPAQ (IPAQ-M) on day 1 and record their daily activities on a physical activity log (PA-Log) for 7 consecutive days. The physical activity log was adapted from the physical activity log based on a previous study. Participants were requested to fill in the PA-Log at the end of each day, starting from the day when they first completed the IPAQ-M. They were also instructed not to alter their daily habits throughout the measurement period. The participants were contacted for the second time to collect the completed PA-Log after 1 week of the first appointment, and then they were requested to fill in the second IPAQ-M, which was identical to the first one. Reliability was conducted by...
requesting each participant to fill in the IPAQ-M 1 week after first taking it, using a test–retest design based on recent reviews.\textsuperscript{20,21} Criterion validity was measured by comparing data collected from the second IPAQ-M and PA-Log.

The self-administered, 7-day period of IPAQ long form was used in this study. Appropriate cultural adaptations and translation were made by following the procedures and guidelines provided by the International Consensus Group (available at http://www.ipaq.ki.se). The IPAQ was translated from the original English version to Malay by 2 bilingual translators; the translated version was then reviewed by 5 bilingual people who were similar to the intended users to ensure the translation would be acceptable to monolingual people. The Malay version was then back-translated into English by 2 different translators and reviewed by another 5 bilingual people. The meanings of the 2 translated and back-translated versions were found to be comparable, and the translated questionnaire was pilot tested to reduce any confusion or uncertainty prior to data collection. There was some confusion in distinguishing between moderate and vigorous activities as reported by the participants, which was consistent with other studies.\textsuperscript{13,21} Hence, a few more practical examples representing moderate and vigorous activities, which were relevant to Malaysian culture, were given in addition to the original IPAQ, based on the Compendium of Physical Activities.\textsuperscript{22} The appropriate cultural adaptations were described and encouraged by the IPAQ research committee to suit the Malaysian sociocultural context.\textsuperscript{23}

For the analysis of IPAQ-M data, the following metabolic equivalent of task (MET) values was used: walking = 3.3 METs, moderate activity = 4.0 METs, and vigorous activity = 8.0 METs. The MET-minutes per week (MET-min week\textsuperscript{-1}) was calculated as follows: minutes of activity/day \times days per week \times MET level. The total amount of physical activity scores obtained from the IPAQ-M was classified into 3 categories for each individual—low, moderate, or high physical activity level—according to the scoring protocol on the IPAQ Web site guidelines.\textsuperscript{23}

The PA-Log covers all activities categorized into different domains (home, at work, sitting, moderate leisure, vigorous leisure, transportation, and “other” activities) in Malay language on 1 page for each day of the 7-day recording period. At the end of each day, the participants were required to record the amount of time (\geq 10 minutes) spent on specific types of activities they were engaged in. The participants were also requested to report the time spent sitting at work/home. After assessing the duration and types of activity spent in different domains, METs were assigned to each activity using published Compendium of Physical Activities,\textsuperscript{19} and were further processed into estimates of energy expenditure (MET \times min). Based on the assigned MET values, all self-reported activities were classified as light (<3 METs), moderate (3-5.99 METs), or vigorous (\geq 6 METs). The weekly total time spent in each intensity level was then calculated from the 7 completed daily logs as a product of MET-min week\textsuperscript{-1}. Total scores were presented as domain-specific scores and activity-specific subscores. Data collected from PA-Log were then used to compare with the questionnaire.

Data was entered and analyzed using SPSS for Windows version 15.0. The significance level was set at $P < .05$. Test–retest reliability of the IPAQ-M was analyzed using Spearman’s correlation coefficients. Reliability was further assessed by a single-measure ICC (1-way random-effects model). ICC represents the proportion of total variance accounted for by the variability between measures. Cohen’s $\kappa$ coefficient was used to examine the physical activity classification agreement of test–retest reliability and validity. Bland and Altman plots\textsuperscript{24} were constructed to assess the validity of total activity.

**Results**

A total of 81 Malay employees participated in the study, and there were slightly more women (52%) than men (48%) in the sample. The age of the study participants ranged from 20 to 55 years, with a mean age of 33 $\pm$ 9.7 years. About half of the respondents had completed
tertiary education, while 42% had graduated from secondary school. There were no significant gender differences ($P > .05$) in age, education, or time spent in different intensities of physical activity or other domains measured by the IPAQ-M and PA-Log. Consequently, data for both sexes were combined for all further analyses.

Classification of the participants into 3 activity categories (low, moderate, and high) was assigned based on their physical activity scores (data not shown). High prevalence of moderate to high physical activity was found among participants in the present study. From the first IPAQ-M, 56% of the participants were classified in the high physical activity category, 37% were in the moderately active category, and only 7% in the low physical activity category. From the second IPAQ-M, 57% of the participants were reported as active, 37% were moderately active, and 6% were inactive. From the PA-Log data, 53% of the participants were classified as highly active, 41% as moderately active, and the remaining 6% as inactive.

Table 1 presents data from different domains and intensities of the IPAQ-M and PA-Log used in the analysis. It was shown that both men and women were physically active in terms of total activity measured by the IPAQ-M (3218 ± 2258 MET-min week$^{-1}$ for first IPAQ-M and 3568 ± 2201 MET-min week$^{-1}$ for second IPAQ-M) and PA-Log (3440 ± 2447 MET-min week$^{-1}$). On the basis of estimated energy expenditure derived from the IPAQ-M, activity during leisure time was the most common form of activity.

The assessment of test–retest reliability for IPAQ-M between the first and second administrations is shown in Table 2. The ICC scores revealed moderate to good correlations in all group activities categorized by intensity and domain and ranged from 0.54 to 0.92, all of which were highly significant ($P < .001$). When expressed by intensity levels of the activity, good ICC values (0.84-0.92) were revealed, except for walking activity, which yielded an ICC of 0.54, 95% CI = 0.37-0.68, but was still considered as moderately correlated. When categorized according to the domains, the ICC was greatest for leisure-time activity (0.90; 95% CI = 0.84-0.94), work and domestic activity correlations were lower, but were still good, ranging from 0.74 to 0.80, and lowest for active transport activity (0.61; 95% CI = 0.44-0.72). The ICC was also good for the sitting domain (0.84; 95% CI = 0.77-0.90). The reliability of the physical activity classification between IPAQ-M was good for the overall group ($\kappa = 0.73$; 95% CI = 0.58-0.87).

The correlation coefficients of the IPAQ-M compared with the subjective PA-Log are shown in Table 2. Results showed good correlation between activities in terms of both intensity levels and domains, ranging from 0.67 to 0.98 ($P < .001$). Vigorous activity from the PA-Log correlated most strongly with IPAQ-M ($\rho = 0.98$; $P < .001$), followed by leisure-time activity ($\rho = 0.96$; $P < .001$). Total activity scores were also significantly correlated to PA-Log scores ($\rho = 0.91$; $P < .001$). The correlation was lowest with walking activity; however, the correlation was still

### Table 1. Descriptive Data From IPAQ-M and PA-Log at Given Intensities and Domains (N = 81)

<table>
<thead>
<tr>
<th>PA Measure</th>
<th>IPAQ-M (First Administration)</th>
<th>IPAQ-M (Second Administration)</th>
<th>PA-Log</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>Median (Range)</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>Total activity</td>
<td>3218 ± 2258</td>
<td>2667 (66-10232)</td>
<td>3568 ± 2200</td>
</tr>
<tr>
<td>Vigorous activity</td>
<td>1100 ± 1415</td>
<td>720 (0-7200)</td>
<td>1207 ± 1561</td>
</tr>
<tr>
<td>Moderate activity</td>
<td>1379 ± 1164</td>
<td>885 (0-4840)</td>
<td>1489 ± 1190</td>
</tr>
<tr>
<td>Walking activity</td>
<td>739 ± 747</td>
<td>545 (0-3960)</td>
<td>872 ± 631</td>
</tr>
<tr>
<td>Work</td>
<td>827 ± 820</td>
<td>605 (0-5307)</td>
<td>831 ± 749</td>
</tr>
<tr>
<td>Transport</td>
<td>265 ± 377</td>
<td>132 (0-1980)</td>
<td>315 ± 383</td>
</tr>
<tr>
<td>Domestic</td>
<td>834 ± 849</td>
<td>540 (0-4440)</td>
<td>923 ± 821</td>
</tr>
<tr>
<td>Leisure</td>
<td>1293 ± 1518</td>
<td>891 (0-8160)</td>
<td>1499 ± 1676</td>
</tr>
<tr>
<td>Sitting</td>
<td>3186 ± 724</td>
<td>3120 (1680-5160)</td>
<td>3266 ± 792</td>
</tr>
</tbody>
</table>

Abbreviations: IPAQ-M, International Physical Activity Questionnaire–Malay version; PA-Log, physical activity log; PA, physical activity; MET, metabolic equivalent task.
significant (ρ = 0.67; P < .001). Sitting hours per week was also significantly positively correlated with PA-Log (ρ = 0.91; P < .001). Validity of the overall activity classification systems has been tested and was found to have an almost perfect agreement (κ = 0.89; 95% CI = 0.79-0.98).

In the Bland–Altman plot (see Figure 1), a small mean bias of 128.1 (standard deviation [SD] = 783.74) was seen between the total activity mean and the difference data from IPAQ-M and PA-Log, but the 95% limits of agreement (LOA) for total activity scores between the 2 methods were large, ranging from 1439.4 (lower LOA) to 1695.5 (upper LOA). There was no significant difference between the IPAQ-M and PA-Log total scores, with most of the values falling between validity intervals (±2 SD).

### Discussion

The idea of our study is to adopt and adapt a questionnaire that is conceptually equivalent to the original IPAQ and allows comparison of physical activity data across countries. To the best of our knowledge, this is the first reliability and validation study on the Malay version of the IPAQ long form (last 7 days) among Malaysian adults.

We found evidence of good reliability on the repeated IPAQ-M. Similar to our results, Craig et al.\(^1\) reported a good Spearman’s correlation coefficient of 0.80 on the reported total activity from the multinational test–retest reliability of the IPAQ long form. It was also noted that reliabilities analyzed by their intensity levels were good, but the repeated recall of vigorous activities were generally better than moderate activities.\(^1\) In another study conducted among French Canadians, high test–retest reliability (ICC = 0.72-0.93) was also observed, in which the authors suggested there is a possibility that the recalls were made from the first time of administration to the second, thereby decreasing measurement error.\(^2\) Our study obtained a higher kappa index (κ = 0.73) as compared with the finding reported by Roman-Viñas et al (κ = 0.61) in Barcelona for total activity,\(^2\) which further supports the IPAQ-M as being good for reporting the physical activity level of the participants.

The considerable variations in the repeatability seen in the recalled subcomponents (vigorous, moderate, and walking activity) and the total activity may be explained by the true variations in the study participants’ daily activities. In agreement with the results from other studies,\(^1,2\) the reliability of low and moderate activities appeared to be not as good as vigorous activity. It seems

### Table 2. Reliability of IPAQ-M and Validity Between PA-Log and IPAQ-M

<table>
<thead>
<tr>
<th>IPAQ-M (MET·min·week(^{-1}))</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC</td>
<td>95% CI</td>
</tr>
<tr>
<td>Total activity</td>
<td>0.92</td>
<td>0.87-0.95</td>
</tr>
<tr>
<td>Vigorous activity</td>
<td>0.84</td>
<td>0.76-0.89</td>
</tr>
<tr>
<td>Moderate activity</td>
<td>0.86</td>
<td>0.79-0.91</td>
</tr>
<tr>
<td>Walking activity</td>
<td>0.54</td>
<td>0.37-0.68</td>
</tr>
<tr>
<td>Work</td>
<td>0.74</td>
<td>0.62-0.82</td>
</tr>
<tr>
<td>Transport</td>
<td>0.61</td>
<td>0.44-0.72</td>
</tr>
<tr>
<td>Domestic</td>
<td>0.8</td>
<td>0.70-0.87</td>
</tr>
<tr>
<td>Leisure</td>
<td>0.9</td>
<td>0.84-0.94</td>
</tr>
<tr>
<td>Sitting</td>
<td>0.84</td>
<td>0.77-0.90</td>
</tr>
</tbody>
</table>

Abbreviations: IPAQ-M, International Physical Activity Questionnaire–Malay version; ICC, intraclass correlation coefficient; MET, metabolic equivalent task; CI, confidence interval.

\(^a\) \(P < .001\).
that this is because low or moderate activities are often accumulated on an everyday basis, plus the diversity and number of these activities is massive, thus leading to a poorer recall. In contrast, vigorous activity such as different types of high-intensity exercise is much more structured and is much easier to recall, and thus better reliability was reported for vigorous activity. Although there were occurrences of variations between the reported activities, the changes were not significant.

The positive correlations for the IPAQ-M in assessing different domains were consistent with other studies, indicating minimal reactivity during the 7 days of test–retest. In our study, leisure-time activity was found to have the highest ICC of 0.90, possibly due to the regular exercise patterns that the participants engaged in and it is thus easier for them to recall during the 7-day measurement period. However, the lowest correlation was reported in active transport activity, which covers time spent walking. The lowest correlation reliability coefficients were reported for time spent walking, similar to those observed by Roman-Viñas et al. This may be explained by the assumption that the IPAQ long form requires individuals to answer 3 times on their time spent walking, making it difficult to divide the weekly overall time spent walking in different domains. During the interviews in our study, some of the participants were confused and tended to sum up the time spent walking in the working environment and walking as a mode of transportation, or they were likely to report it twice in 2 different domains. These questions should hence be operationalized or better defined to make recall easier.

It has been a concern that the interval for test–retest should be long enough to reduce the effects of memory but short enough to diminish the likelihood of systematic alterations. Although the choice of this interval is somewhat arbitrary, in practice a period of 2 to 14 days apart is
considered adequate. Our reliability results indicate an adequate stability of the IPAQ-M with a period of 7 days between the interviews, and the recall of the respondents’ physical activity was considered stable as well.

The validity correlations for all domains are high between the IPAQ-M and PA-Log. Consistent with the results reported by Hagströmer et al., a significant relationship was found between the PA-Log and the IPAQ for all domains except active transport domain. High validity was also reported on leisure-time activity as demonstrated in our study.

The validity comparison between the IPAQ-M and PA-Log on the overall reported activity in our study using \( \kappa \) was very good, probably because PA-Log allowed individualized MET values for each reported activity and allowed capture of the activities accurately, thus producing consistent total scores. The Bland–Altman plot shows good agreement/validity for total activity where a small difference between the means of total weekly activity was observed. Yet, it should be noted that the large 95% limits of agreement can be due to individual errors or a few outliers that produced higher activity levels.

On the contrary, our results contradicted those presented by Macfarlane et al., in which they found no significant correlations between the IPAQ long form and PA-Log. Macfarlane et al explained that their participants might have failed to fully comply with the PA-Log protocols or they did not record their activity at the end of each day on a regular basis. The difference seen in the present study could be due to the fact that our participants were represented mostly by clerical staff working in the university, where their job duties and activities are more routine and hence easier for them to report the activities they have performed and thus reducing error caused by cognitive recall.

It should also be highlighted that none of the respondents in our study engaged in bicycling activity as a mode of transportation. With the increased availability of private automobiles, people tend to replace bicycling with driving motorized vehicles such as a car or motorcycle as a more appropriate transport mode, even just for short distances. There were also no suitable bicycle lanes and cycle paths available here in Kuala Lumpur, coupled with considerable social safety reasons.

PA-Log was the choice of our criterion measure not only due to its low cost but also due to the potential to provide valid estimates of activity energy expenditure and time spent at different intensity levels/domains. Although there is an increasing use of the accelerometers and pedometers in the IPAQ validation studies, these tools are known to underestimate energy expenditure. Other objective tools such as the heart rate monitor or the doubly labeled water methods can be an alternative, but cost is a concern.

A further limitation is that the participants in our study comprised a relatively small convenient sample that were mostly made up of individuals with higher education level from the urban setting; therefore, the validation of the IPAQ-M should be further tested among the Malays in the rural parts of Malaysia. However, our study is probably the first to validate the IPAQ long form in Malay language in our country. This will provide a valid and reliable tool in measuring the levels and distribution of physical activity among the Malay population in the urban setting.

**Conclusion**

Our results show that the IPAQ-M has good reliability and validity for assessing physical activity in healthy Malay adults. The findings of this study therefore support the use of the IPAQ-M for epidemiological research among the Malay population.

**Acknowledgments**

We would also like to thank all participants and colleagues for their support to this study.
Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a research grant (RG051/09HTM) from the University of Malaya.

References