Test–retest reliability of the Brunel Lifestyle Physical Activity Questionnaire

CrossMark

Massimo M. Vencato [a], [1], Costas I. Karageorghis [a], [*], [1], Alan M. Nevill [b], [2], David-Lee Priest [c], [3]

[a] Brunel University London, UK
[b] University of Wolverhampton, UK
[c] City College Norwich, UK

A B S T R A C T

Objectives: To establish the test–retest reliability of planned physical activity (PPA) and unplanned physical activity (UPA) components of the Brunel Lifestyle Physical Activity Questionnaire (BLPAQ). To provide evidence of the BLPAQ's stability using the proportion of agreement (PoA) method over a 5-week period.
Design: Test–retest over a 5-week period using three diverse samples of adults.
Methods: The 277 participants were subdivided into three adult samples: gymnasium users ($n = 80$), undergraduate students ($n = 111$), and university staff members ($n = 86$). They were asked to complete the test–retest measure in their places of exercise, study, or work respectively.
Results: Correlation coefficients between test–retest administrations were calculated for each participant group and intraclass correlations were calculated for each item. Pearson's product-moment correlations ranged from $r = 0.95$ to $r = 0.96$ for the PPA subscale and $r = 0.93$ to $r = 0.98$ for the UPA subscale. Intraclass correlations ranged from $R = 0.52$ to $R = 0.99$ for PPA and $R = 0.87$ to $R = 0.99$ for UPA. Fisher's z tests indicated that the test–retest correlation coefficients for the BLPAQ subscales were, on the whole, significantly stronger than those of older, comparable subscales from lifestyle physical activity questionnaires. The PoA analysis for each item revealed that the test–retest administrations were in high agreement (>95%).
Conclusions: Overall, the PPA and UPA factors of the BLPAQ demonstrated high reliability and stability. The present study also illustrates the utility of PoA analysis in establishing the stability of physical activity measures.

© 2017 Published by Elsevier Ltd.

1. Introduction

Over the last three decades, it has become apparent that the intensity of exercise needed to promote health benefits is considerably lower than the intensity needed to improve physical fitness (American College of Sports Medicine (ACSM), 2013, pp. 5–12). A growing number of scientific studies have demonstrated the favourable effects of low-to-moderate intensity physical activity (PA) on cardiometabolic health (Karmali & Lloyd-Jones, 2014).

Owing to their wide availability, low cost, and limited imposition on the respondent, PA questionnaires are advantageous in researching PA behaviour when compared to other direct and/or objective measurement tools (Helmerhorst, Brage, Warren, Besson, & Ekelund, 2012; Matthews, Steven, George, Sampson, & Bowles, 2012). The need for valid and reliable moderate-intensity lifestyle physical activity (LPA) measures has been widely recognised as a priority for the continuous advancement of this field of research endeavour (Aguilar-Farías, Brown, Olds, & Peeters, 2015). Before such measures can be employed for epidemiological research, their reliability must first be assessed (Warren et al., 2010). The rationale for using three differing participant groups in the present study was to develop a measure that is suitable for use with diverse age, ethnic and socio-economic groups. Notably, these personal factors are associated with the amount of PA undertaken during leisure time (Davies, Spence, Vandelanotte, Caperchione, & Mummery,

* Corresponding author. Department of Life Sciences, Brunel University London, UB8 3PH, UK.
E-mail address: costas.karageorghis@brunel.ac.uk (C.I. Karageorghis).
[1] Department of Life Sciences, Brunel University London, UK.
[2] Faculty of Education, Health and Wellbeing, University of Wolverhampton, UK.
[3] Department of Psychology, City College Norwich, UK.

2012; Lee & Im, 2010; Peels et al., 2013).

The principal aim of the present study was to ascertain the reliability and stability of the Brunel Lifestyle Physical Activity Questionnaire (BLPAQ; Karageorghis, Vencato, Chatzisarantis, & Carron, 2005). A secondary aim was to provide supporting evidence for the stability of the BLPAQ as a measure of LPA using proportion of agreement analysis (PoA), which is a relatively novel statistical technique used to assess the agreement/differences between two sets of scores (Nevill, Lane, Kilgour, Bowes, & Whyte, 2001). The PoA method was used to ascertain the proportion of test–retest differences that lie between ±1 for all items of the BLPAQ. This statistical analysis has been recommended for establishing the stability of a questionnaire that uses a 5-point Likert-type scale (Nevill et al., 2001). Further, the "item-by-item" PoA may also be used to identify "rogue" items in the initial stages of psychometric measure development (Nevill et al., 2001). The BLPAQ measure consists of two factors proposed to reflect the planned and unplanned components of LPA (Dunn, Andersen, & Jakicic, 1998). Vencato, Karageorghis, Priest, and Nevill (2017) recently demonstrated the criterion validity of the BLPAQ with British leisure centre users and two reference measures: the Baecke Questionnaire of Habitual Physical Activity (BQHPA; Baecke, Burema, & Frijters, 1982) and Godin's Leisure-Time Exercise Questionnaire (GLTEQ; Godin & Shephard, 1985).

Given the generally more stable and recollectable nature of the constructs tapped by the BLPAQ relative to those tapped by the reference measures (BQHPA and GLTEQ), it was hypothesized that both the PPA and UPA factors of the BLPAQ would show higher reliability scores. Notably, the two reference measures of PA do not reflect the theoretically important distinction between planned and unplanned PA (Hallal et al., 2012; Macfarlane, Lee, Ho, Chan, & Chan, 2006). The secondary hypothesis was that both the PPA and UPA factors would demonstrate high (>95%) PoA values (Nevill et al., 2001).

2. Methods

2.1. Participants

The initial sample ($N = 337$) was reduced to 277 as 60 participants were excluded on the grounds that they failed to complete the retest. The sample was subdivided as follows: Subsample 1 comprised 80 regular gym users at a local authority-run leisure centre (42 women and 38 men, $M_{age} = 38.8$ years, $SD = 17.7$ years). Subsample 2 comprised 111 undergraduate students (66 women and 45 men, $M_{age} = 24.4$ years, $SD = 4.9$ years). Subsample 3 comprised 86 university staff members (44 women and 42 men, $M_{age} = 41.1$ years, $SD = 11.9$ years).

2.2. Study design and procedures

The project was granted institutional ethical approval and all participants provided written informed consent. Subsequent to the initial administration of the BLPAQ at their respective recruitment locations (subsample 1: local authority gymnasium; subsamples 2 and 3: university campus), participants were invited to complete the retest measure after a 5-week period. This time gap was considered of a sufficient length to ensure that participants could not recall their BLPAQ responses (Wendel-Vos, Schuit, Saris, & Kromhout, 2003), and of sufficient brevity to prevent seasonal changes in PA from influencing the results (O'Connell, Griffiths, & Clemes, 2014). On the second administration, a revised question format was used wherein participants were asked to report the average amount of PPA and UPA that they had completed over the preceding 5-week period.

2.3. Measures

The BLPAQ is a published and validated instrument that measures PPA and UPA (Karageorghis et al., 2005; Vencato et al., 2017). PPA is measured by use of six items that tap the intensity, frequency, and duration of such activity (e.g., "How many times in a normal week do you engage in planned physical activity?"). UPA is measured by use of three items that tap only the intensity and duration of such activity (e.g., "How vigorously do you engage in these other forms of physical activity"). Items are rated using a 5-point continuous closed numerical scale relating to a "normal" week. The initial development of the BLPAQ indicated that the instrument exhibited internal consistency (Cronbach α estimates of 0.90 for the PPA subscale and 0.68 for the UPA subscale, which had only three items) and factorial validity (Comparative Fit Index = 0.94; Standardized Root Mean Residual = 0.05; Akaike Information Criterion = 54.74) (Karageorghis et al., 2005; Tabachnick & Fidell, 2013, p. 774).

2.4. Data analysis

Correlation coefficients between test–retest administrations of the BLPAQ were calculated to assess reliability using Pearson correlations (one-tailed). The resulting coefficients were tested statistically against the original test–retest coefficients of the reference measures (i.e., those established in the original validation studies by Baecke et al., (1982), Jacobs et al., (1993), Pols et al., (1995), and Godin & Shephard (1985)). This was accomplished by use of Fisher's z transformation test, which computes confidence intervals to facilitate difference testing between correlations (Fisher, 1915).

To fortify the assessment of test–retest reliability, intraclass correlations (Cohen's κ) were computed at item level. Subsequently, a PoA analysis was performed for each item as recommended by Nevill and colleagues (Nevill et al., 2001). The findings from the PoA analyses were entered into a single-sample Wilcoxon Signed-Ranks test (r_{ES}: Wilcoxon, 1945) to determine whether there was a departure from the hypothesized median (0) for each item's score. The test–retest variations from the median (0) for each item-score were transformed into the percentage (%) of agreement for each item composing the two subsamples of the BLPAQ.

3. Results

PPA and UPA scores showed significant test–retest correlations for the entire sample and each subsample (range: $r = 0.93–0.98$, $p < 0.01$; variance explained = 86.5–96.0%; see Table 1). Table 1 presents the Pearson correlation results relative to PPA and UPA for the entire sample and each subsample (range $r = 0.88–0.98$; $p < 0.01$; variance explained range = 86.5–96.0%). In each subsample, correlation coefficients for the female participants were significantly greater than those of the males.

Intraclass correlations indicated that there was acceptable test–retest reliability at item level (see Table 1). Specifically, one of the Cohen's κ coefficients was 0.52, which is considered "fair", two were in the range 0.60–0.74, which is considered "good", and the remainder (i.e., 97.2%) were >0.75, which is considered "excellent" (Cicchetti, 1994). Fisher's z tests to compare the magnitude of test–retest Pearson or Spearman correlations between BLPAQ subscales and those from the reference measures indicated that against the BQHPA, the BLPAQ exhibited significantly larger correlations in 97.2% of comparisons (see Table S1–Table S4). Similarly, against the GLTEQ, the BLPAQ exhibited significantly larger correlations in 93.1% of comparisons (see Table S5–Table S8).

Table 1
Pearson's one-tailed product-moment correlation and intraclass coefficient correlation results for PPA and UPA across the entire sample ($N = 277$) and for subsample 1 ($n = 80$), subsample 2 ($n = 111$), and subsample 3 ($n = 86$) by gender.

	Entire sample			Subsample 1			Subsample 2			Subsample 3		
	Both	F	M	Both	F	M	Both	F	M	Both	F	M
	($N = 277$)	($n = 152$)	($n = 125$)	($n = 80$)	($n = 42$)	($n = 38$)	($n = 111$)	($n = 66$)	($n = 45$)	($n = 86$)	($n = 44$)	($n = 42$)
PP-MC-PPA	0.96**	0.97*	0.94**	0.95**	0.95**	0.97**	0.96**	0.99*	0.91**	0.95**	0.98*	0.92**
ICC-Item 1	0.94**	0.97**	0.90**	0.93**	0.97**	0.88**	0.95**	0.98**	0.90**	0.94**	0.96**	0.91**
ICC-Item 2	0.94**	0.96**	0.92**	0.95**	0.93**	0.97**	0.93**	0.98**	0.87**	0.94**	0.97**	0.92**
ICC-Item 3	0.89**	0.95**	0.82**	0.93**	0.96**	0.91**	0.89**	0.99**	0.71**	0.72**	0.84**	0.52*
ICC-Item 4	0.95**	0.97**	0.91**	0.97**	0.97**	0.97**	0.96**	0.98**	0.91**	0.89**	0.95**	0.80**
ICC-Item 5	0.98**	0.99**	0.97**	0.98**	0.98**	0.98**	0.98**	0.99**	0.95**	0.98**	0.99**	0.96**
ICC-Item 6	0.96**	0.96**	0.96**	0.97**	0.98**	0.97**	0.96**	0.97**	0.95**	0.94**	0.93**	0.94**
PP-MC-UPA	0.96**	0.98*	0.93**	0.98**	0.97**	0.98**	0.93**	0.99*	0.88**	0.96**	0.99*	0.94**
ICC-Item 1	0.97**	0.99**	0.94**	0.98**	0.99**	0.98**	0.97**	0.99**	0.92**	0.97**	0.99**	0.93**
ICC-Item 2	0.94**	0.96**	0.93**	0.96**	0.95**	0.96**	0.92**	0.96**	0.87**	0.95**	0.96**	0.95**
ICC-Item 3	0.97**	0.99**	0.95**	0.98**	0.98**	0.98**	0.97**	0.99**	0.93**	0.97**	0.99**	0.95**

Note. F = female; M = male; PPA = planned physical activity; UPA = unplanned physical activity; PP-MC = Pearson's product-moment correlations (r; one-tailed); ICC = intraclass coefficient correlations (κ).
*$p < 0.05$. **$p < 0.01$.

Each item demonstrated PoA scores above the minimum threshold as recommended by Nevill and colleagues (Nevill et al., 2001). PPA factor items were also tested for significance of deviation from the median between test–retest administrations of the questionnaire using the single-sample Wilcoxon Signed-Ranks test. These analyses yielded only small negative effect sizes (r_{ES} range = 0.00 to −0.14). When re-administered, item 3 was found to deviate from the median ($T = 188.50$, $p = 0.001$, $r_{ES} = -0.14$; see Table 2). PoA scores by gender appear in Table 3.

Results for each subsample are displayed in Table 4. In respect of subsample 1, The PoA for all the items composing the PPA ranged from 96.3% to 100%, and the single-sample Wilcoxon test yielded only small negative effect sizes (r_{ES} range = −0.02 to −0.22). Only item 3 showed a significant deviation ($T = 12.00$, $p = 0.007$, $r_{ES} = -0.21$) from the median. For subsample 2, the PoA for all the PPA factor items ranged from 95.5% to 100%, and the Wilcoxon test yielded only small negative effect sizes (r_{ES} range = −0.01 to −0.11). None of the items showed any significant deviation ($p > 0.05$) from the median. In respect of subsample 3, the item scores ranged from 96.5% to 98.9%. Tests for the deviation from the median yielded small effect sizes (r_{ES} range = 0.00 to −0.15).

All UPA factor items were above the minimum threshold recommended by Nevill et al. (2001); therefore, they were considered to display high PoA levels between test and retest administrations. Specifically, the combined sample ($n = 277$) yielded PoA that ranged from 98.2% to 99.6% (see Table 1). The Wilcoxon analyses

yielded only small negative effect sizes (r_{ES} range = −0.03 to −0.05). None of these items were found to have deviated significantly ($p > 0.05$) from the median at the re-administration of the BLPAQ. For both female and male subgroups, all of the UPA factor items were above the minimum recommended threshold (Nevill et al., 2001).

The results for each subsample appear in Table 4. For subsample 1, the PoA for each item was 100%. All items were tested for the significance of deviation from the median between the two administrations. These analyses yielded only small effect sizes (r_{ES} range = 0.00 to −0.08), and none of the items significantly deviated ($p > 0.05$) from the median between test–retest administrations. Findings for subsample 2 indicated that all the UPA factor items ranged from 97.3% to 99.1%, and the single-sample Wilcoxon test produced only small effect sizes (r_{ES} range = 0.00 to −0.06). None of the items significantly deviated ($p > 0.05$) from the median. Results for subsample 3 in respect of PoA indicated that all the UPA factor items ranged from 97.7% to 100%. Tests for the deviation from the median yielded small effect sizes (r_{ES} range = 0.00 to −0.20), with only item 2 having scored significantly lower ($T = 0.00$, $p = 0.008$, $r_{ES} = -0.20$) at the second administration of the BLPAQ, and exhibited a 100% agreement between the two administrations.

4. Discussion

The principal aim of the present study was to examine the

Table 2
Proportion of agreement analysis results for planned and unplanned physical activity items across the entire sample ($N = 277$).

BLPAQ items	Score range		Test 1 - Test 2		T	r_{ES}	% (±1)	≤1	0 diff	≥1
	Min	Max	Mean	SD						
Entire sample										
PPA Item 1	−2	2	0.03	0.45	202.50	−0.04	97.10	13	246	18
PPA Item 2	−3	2	−0.02	0.49	−230.00	−0.03	97.10	16	245	16
PPA Item 3	−2	2	0.09	0.47	188.50	−0.14*	97.10	9	237	31
PPA Item 4	−2	3	0.01	0.41	217.00	−0.01	98.60	15	247	15
PPA Item 5	−2	1	−0.01	0.35	−232.50	−0.01	99.60	16	246	15
PPA Item 6	−1	3	0.04	0.36	121.00	−0.06	98.20	11	251	15
UPA Item 1	−2	2	−0.01	0.36	−129.00	−0.03	98.50	14	253	10
UPA Item 2	−1	2	0.02	0.33	148.00	−0.05	99.60	11	250	16
UPA Item 3	−2	2	−0.01	0.37	−118.00	−0.03	98.20	13	254	10

Note. Min = minimum; Max = maximum; T = single sample Wilcoxon signed-ranks test; r_{ES} = direction of effect size; diff = difference; PPA = planned physical activity; UPA = unplanned physical activity.
*$p < 0.01$.

Table 3
Proportion of agreement analysis results for planned and unplanned physical activity items for female (n = 152) and male (n = 125) participants.

BLPAQ items	Score range		Test 1 - Test 2		T	r_{ES}	% (± 1)	≤1	0 diff	≥1
	Min	Max	Mean	SD						
Female participants										
PPA Item 1	−2	1	−0.05	0.29	−3.00	−0.11*	98.7	6	145	1
PPA Item 2	−3	2	0.01	0.42	41.00	−0.02	98.0	5	139	8
PPA Item 3	0	2	0.10	0.34	0.00	−0.20**	98.6	0	139	13
PPA Item 4	−2	2	0.00	0.33	13.50	−0.01	98.1	4	145	3
PPA Item 5	−1	1	0.01	0.26	22.00	−0.04	100	4	142	6
PPA Item 6	−1	3	0.00	0.36	−21.00	−0.01	98.7	6	143	3
UPA Item 1	−1	1	−0.01	0.20	−7.00	−0.05	100	4	146	2
UPA Item 2	−1	1	0.00	0.23	18.00	0.00	100	4	144	4
UPA Item 3	−1	1	0.00	0.23	18.00	0.00	100	4	144	4
Male participants										
PPA Item 1	−2	2	0.11	0.57	78.50	−0.14*	95.2	7	101	17
PPA Item 2	−3	2	−0.06	0.56	−69.00	−0.07	96.0	11	106	8
PPA Item 3	−2	2	0.09	0.60	126.00	−0.10	95.2	9	98	18
PPA Item 4	−1	3	0.02	0.50	126.50	−0.02	99.2	11	102	12
PPA Item 5	−2	1	−0.03	0.44	−94.50	−0.05	99.2	12	104	9
PPA Item 6	−1	2	0.08	0.45	37.50	−0.13*	97.6	5	108	12
UPA Item 1	−2	2	−0.02	0.49	−78.00	−0.02	96.8	10	107	8
UPA Item 2	−1	2	0.05	0.42	66.50	−0.08	99.2	7	106	12
UPA Item 3	−2	2	−0.03	0.49	−48.00	−0.04	96.0	9	110	6

Note. Min = minimum; Max = maximum; T = single-sample Wilcoxon signed-ranks test; r_{ES} = direction of effect size; diff = difference; PPA = planned physical activity; UPA = unplanned physical activity.
*$p < 0.05$. **$p < 0.01$.

Table 4
Proportion of agreement analysis results for planned and unplanned physical activity items for subsamples 1 (n = 80), 2 (n = 111) and 3 (n = 86).

BLPAQ items	Score range		Test 1 - Test 2		T	r_{ES}	% (±1)	≤1	0 diff	≥1
	Min	Max	Mean	SD						
Subsample 1										
PPA Item 1	−1	2	0.09	0.53	48.00	−0.12	97.60	6	63	11
PPA Item 2	−3	1	−0.04	0.56	−45.50	−0.04	97.60	7	66	7
PPA Item 3	−1	2	0.16	0.51	12.00	−0.21*	96.30	2	66	12
PPA Item 4	−2	1	0.03	0.39	18.00	−0.05	98.80	3	71	6
PPA Item 5	−1	1	0.04	0.34	15.00	−0.08	100	3	71	6
PPA Item 6	−1	2	−0.01	0.41	−25.00	−0.02	98.8	6	70	4
UPA Item 1	−1	1	0.00	0.28	10.50	0.00	100	3	74	3
UPA Item 2	−1	1	−0.04	0.34	−15.00	−0.08	100	6	71	3
UPA Item 3	−1	1	−0.03	0.32	−13.50	−0.06	100	5	72	3
Subsample 2										
PPA Item 1	−2	2	0.01	0.42	13.00	−0.01	96.40	3	104	4
PPA Item 2	−2	2	−0.01	0.46	16.00	−0.07	95.50	4	103	4
PPA Item 3	−2	2	0.03	0.37	16.50	−0.05	98.20	3	102	6
PPA Item 4	−1	2	−0.03	0.34	−20.00	−0.06	99.10	7	101	3
PPA Item 5	−1	1	−0.05	0.31	−18.00	−0.10	100	8	100	3
PPA Item 6	−1	2	0.05	0.35	7.00	−0.11	98.20	2	103	6
UPA Item 1	−2	2	−0.04	0.38	−18.50	−0.06	98.20	7	101	3
UPA Item 2	−1	2	0.02	0.36	27.50	−0.04	99.10	5	100	6
UPA Item 3	−2	2	0.00	0.41	−22.00	0.00	97.30	4	102	5
Subsample 3										
PPA Item 1	−2	2	−0.01	0.39	−12.50	−0.02	97.70	4	79	3
PPA Item 2	−3	1	−0.02	0.46	−25.00	−0.02	98.90	5	76	5
PPA Item 3	−2	2	0.12	0.54	38.50	−0.15	96.50	4	69	13
PPA Item 4	−1	3	0.05	0.51	25.00	−0.06	97.70	5	75	6
PPA Item 5	−2	1	0.00	0.41	33.00	0.00	98.90	5	75	6
PPA Item 6	−1	3	0.06	0.47	10.50	−0.08	97.70	3	78	5
UPA Item 1	−2	2	0.00	0.41	18.00	0.00	97.70	4	78	4
UPA Item 2	0	1	0.08	0.28	0.00	−0.20*	100.00	0	79	7
UPA Item 3	−2	2	−0.02	0.38	−8.00	−0.04	97.70	4	80	2

Note. Min = minimum; Max = maximum; T = single-sample Wilcoxon signed-ranks test; r_{ES} = direction of effect size; diff = difference; PPA = planned physical activity; UPA = unplanned physical activity. Subsample 1 consisted of gym users, Subsample 2 of undergraduate students and Subsample 3 of university staff.
*$p < .01$.

reliability of the BLPAQ over a 5-week period among three diverse samples of adults. It was expected that the reliability scores of the BLPAQ would be greater than those of the older questionnaires used for comparative purposes: the Baecke Questionnaire of Habitual Physical Activity (BQHPA; Baecke et al., 1982) and the Godin's Leisure Time Exercise Questionnaire (GLTEQ; Godin & Shephard, 1985). Diagnostic criteria indicate that a very high reliability coefficient (e.g., r = 0.85–0.90) would be required when scores of a single person are considered, but relatively moderate values (e.g., r = 0.50–0.70) are acceptable when assessing two

groups of participants (Streiner, Norman, & Cairney, 2014, p. 228).

Results of the Pearson product-moment correlations showed high correlation coefficients between test–retest administrations of the BLPAQ (range $r = 0.93–0.98$, all at $p < 0.01$, range of variance explained = 77.4–98.0%; see Table 1). Together, both subscales of the BLPAQ explain 92.2% of the variance, thus demonstrating the significance of these factors in ascertaining LPA behaviour among the general population. Accordingly, when re-administered following a 5-week period, the BLPAQ explained a greater percentage of the variance than the BQHPA (Baecke et al., 1982) and the GLTEQ (Godin & Shephard, 1985).

Similarly, when examining test–retest by use of intraclass correlations, Cohen's κ statistic indicated that, with the exception of good and fair agreement for item 3 of the PPA ("In general, what is the duration of each session of pre-planned physical activity that you engage in?") among males in subsamples 2 and 3, respectively, all other items exhibited excellent agreement (cf. Cicchetti, 1994). This provides strong evidence at subscale and item level of the test–retest reliability of the BLPAQ factors. The slightly lower Cohen's κ scores for item 3 might be attributed to the fact that men in subsamples 2 and 3 who were, on the whole, less active than their counterparts in subsample 1, also did not have the leisure centre time structure that served to assist them in recalling the precise duration of their engagement in PA.

To further test the primary hypothesis, Fisher's z test was used to facilitate comparison between the magnitude of test–retest correlations (Pearson's product moment and Spearman's rank order, as appropriate) between BLPAQ subscales and those from the reference measures (BQHPA and GLTEQ). These analyses indicated that, on average, the BLPAQ exhibited significantly larger correlations in 95.2% of comparisons against BQHPA and GLTEQ subscales (see Table S1–Table S8). Collectively, the evidence accumulated from the test–retest analyses led us to accept the primary research hypothesis.

Although the agreement scores for each item of the BLPAQ may not be compared directly to those of other PA questionnaires, the secondary research hypothesis can be accepted given that both constructs of the BLPAQ have been found to exhibit high stability (>95%; Nevill et al., 2001). Three of the four indices of the BQHPA (i.e., leisure, work, and total PA) reflect similar activities to those measured collectively by the PPA and UPA factors (Matthews et al., 2012). In the study by Baecke and colleagues, test–retest Pearson coefficients for the BQHPA factors ranged from $r = 0.74$ (Leisure index, variance explained = 54.8%) to $r = 0.88$ (Work index, variance explained = 77.4%; Baecke et al., 1982).

Jacobs, Ainsworth, Hartman, and Leon (1993) indicated that using a shorter time period between test and retest (i.e., one month) led to significant ($p < 0.05$) Spearman rank-order correlation results that ranged from $r_s = 0.78$ (Work index, variance explained = 60.8%) to $r_s = 0.93$ (Total PA index, variance explained = 86.5%). As the Jacobs et al.'s study employed a sample of university students and staff, a comparison can be made with the findings from subsamples 2 and 3 employed in the present study (see Table 1). Such results provide additional evidence that the BLPAQ factors were able to account for a greater percentage of the variance in LPA behaviour than those of the BQHPA.

In the current study, three PPA and three UPA item scores were higher for women than for men (see Table 3). These findings indicate that the BLPAQ may be more sensitive to gender variations than the BQHPA (Baecke et al., 1982). A possible explanation for these variations is that females and males may perceive LPA as part of their leisure time and/or working lives (e.g., household activities, walking children to and from school, afterschool activities, etc.; Department of Health, Physical Activity, Health Improvement and Protection (DHPAHIP), 2016). This differentiation requires further

research. Sports profiles may also differ by gender; for example, men are more likely to participate in sports involving high impact (e.g., rugby, soccer, etc.; Sisson, McClain, & Tudor-Locke, 2008), whereas women more often participate in walking-related activities (Speck & Harrell, 2003).

Previous BQHPA research over longer test–retest periods (i.e., 5–11 months; Pols et al., 1995) found that replicability was slightly higher after five as opposed to 11 months, but not significantly so ($p > 0.05$). It was not reported whether or not the two groups differed from each other at both measurement points. It is plausible that the demanding schedule of activities imposed by the researchers may have created selection bias towards health-oriented individuals, thus leading to overestimation of the BQHPA's reliability. It was concluded that the BQHPA may be a more accurate measure for men than women due to its lesser emphasis on domestic tasks.

The three components of the GLTEQ (i.e., Light PA, Moderate PA, and Total PA) reflect activities that are comparable to the PPA and UPA factors of the BLPAQ. However, the reliability of the BLPAQ exceeds that of the GLTEQ (Adams et al., 2005). The correlation coefficients and the variances explained for Light PA, Moderate PA and Total PA were $r = 0.48$ (variance explained = 23.0%), $r = 0.46$ (21.2%), and $r = 0.74$ (54.8%) respectively (Godin & Shephard, 1985). Notably, these findings were obtained over a 2-week time span, which suggests that the stability of the BLPAQ is not only greater than that of the GLTEQ, but also demonstrable over a broader timeframe.

The 2-week and 1-month retest observations for the GLTEQ indicated high correlations for strenuous PA $r = 0.94$ (variance explained = 88.4%; Godin & Shephard, 1985), and $r_s = 0.84$ (variance explained = 70.6%; Jacobs et al., 1993). These results were probably influenced by high-intensity activities (estimated METs > 7.0), which are more easily recalled over a narrow timeframe (Shephard, 2003). Reliability decreases with the length of recall period, partly due to seasonal and/or temporal variations in PA patterns (Shephard, 2003). Comparable data from the College Alumnus Physical Activity Questionnaire (Paffenbarger, Wing, & Hyde, 1978) found correlation coefficients of $r = 0.72$ at 1-month, falling to $r = 0.3$ and to $r = 0.4$ over 8–12 months (Jacobs et al., 1993).

A further aim of the present study was to ascertain the stability of the BLPAQ using the PoA, a relatively novel method in establishing test–retest stability of a PA measure. Due to the high sample size (>100) needed to yield meaningful results (Nevill et al., 2001), the comparison between gender groups was only possible when the three subsamples were merged (see Table 3). In respect of both the female and male samples, each PPA item was above the minimum threshold (Nevill et al., 2001). Results for the entire sample showed that PPA item 3 (pertaining to the duration of each session) deviated from the median ($T = 188.50$, $p = 0.001$, $r_{ES} = -0.14$; see Table 2), particularly among women (see Table 3). The initial administration of the BLPAQ may have sensitized participants to the amount of time they were engaged in PPA, which could have motivated them to increase this prior to the retest. Alternatively, they may have over-reported their time expended in a particular activity owing to social desirability, a bias that may be more prevalent among female respondents (Adams et al., 2005). It is notable that the intraclass correlations (Table 1) showed no such issues for women in respect of PPA item 3, albeit that a minor weakness was identified in the same item among men in subsamples 2 and 3. This is perhaps illustrative of the benefit of using both reliability and stability measures in the developmental phase of PA questionnaires (see Nevill et al., 2001).

Two other items showed a positive deviation from the median in respect of men (PPA item 1: $T = 78.50$, $p = 0.031$, $r_{ES} = -0.14$; PPA

item 6: $T = 37.50$, $p = 0.048$, $r_{ES} = -0.13$; see Table 3). Female participant responses indicated a positive deviation from the median for PPA item 3 ($T = 0.00$, $p = 0.001$, $r_{ES} = -0.20$; see Table 3). Item 1 assesses how many times a week one engages in PPA, item 3 ascertains the duration of each PPA session, whereas, item 6 concerns how vigorously one engages in PPA. Both of these items were scored higher at the retest, possibly due to more favourable weather conditions, which, according to extant research (O'Connell et al., 2014), may have facilitated more vigorous outdoor activity outdoors. The scores for each UPA item were above the minimum PoA threshold (Nevill et al., 2001). Subsample 3 recorded a deviation from the median for item 2 ($T = 0.00$, $p = 0.008$, $r_{ES} = -0.20$; see Table 4), which relates to the intensity level of unplanned activities. It may have been that the seasonality factor, which could have influenced the responses to item 6 in the PPA construct, may also have a bearing on this item.

Participants in subsample 1 were gym users who exercised regularly; consequently, their test–retest scores exhibited higher correlation coefficients than participants from the other two subsamples. Thus, it is plausible that their perception of the intensity level for activities included in PPA and UPA constructs might have differed from those of the general population. The findings relating to subsamples 1 (gym users) and 2 (university students) may have greater generalizability with comparable subgroups in the population. Contrastingly, subsample 3 (university staff) may be more readily compared with PA trends relating to the general UK population.

The BQHPA and GLTEQ have seldom been used in UK-based research (Boreham et al., 2004; Dugdill, Graham, & McNair, 2005; Vencato et al., 2017). The limited evidence associated with these instruments precludes detailed comparison with the present findings. Although there are no reported differences in PA behaviour across English regions, there are some differences across the UK home nations (DHPAHIP, 2016). Therefore, the present findings cannot necessarily be generalized to the entire UK population. The absence of UK-based studies that have employed the PoA method to ascertain the stability of existing PA measures prevents the comparison of the present findings with an appropriate criterion measure.

The reliability coefficients reveal the extent to which the present results are free from error variance. The high correlation coefficients may suggest that not only were the majority of participants able to differentiate between planned and unplanned activities, but also significantly recall low-to-moderate intensity PA, a factor found to compromise the reliability of other PA measures (Shephard, 2003). There was a small random error that may have resulted from the examinations period for the university students and summer vacation period for the university staff. Additionally, the elevated daily temperatures experienced during the summer months (June–August mean range 20.4–22.8 °C; Wikipedia, 2016) could have influenced some of the planned and unplanned activities (O'Connell et al., 2014).

5. Conclusions and recommendations

Collectively, the evidence presented herein coupled with previous work on the validity and reliability of the BLPAQ (Karageorghis et al., 2005; Vencato et al., 2017), points to the fact that it is an instrument suitable for use by researchers and practitioners in the physical activity domain. The high level of reliability and stability for the BLPAQ subscales evidenced in the present study over a 5-week period would suggest that future work should address test–retest reliability and stability using a longer timeframe (e.g., 3 months; cf. Baecke et al., 1982) and across seasons (e.g., winter and spring). Researchers should also re-administer

reference measures concurrently with the main measure of interest in order to compare reliability and stability using the same participant pool (cf. Jacobs et al., 1993).

The novel use of the PoA method described herein has highlighted the necessity to reconsider the established validation procedures for PA questionnaires owing to limited data pertaining to their stability (Nevill et al., 2001). The present PoA analyses show that the PPA and UPA factors of the BLPAQ exhibit a high level of stability. The results also provide initial evidence that the BLPAQ has the capacity to detect low-to-moderate intensity LPA over a short time-span. Overall, the BLPAQ possesses some distinct advantages when compared to many other PA measures: it is relatively brief in length (9 items), and can be completed in just 3–5 min. Accordingly, the BLPAQ represents a progressive and promising method by which to estimate compliance with the current PA guidelines for health in large populations (see e.g., ACSM, 2013). Additionally, the BLPAQ could be used by health professionals who wish to assess the efficacy of LPA interventions.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.psychsport.2017.07.003.

References

Adams, S. R., Matthews, C. E., Ebbeling, C. B., Hardt, J., Gorber, S. C., & Tremblay, M. (2005). The effect of social desirability and social approval on self-reports of physical activity. *American Journal of Epidemiology, 161*, 389–398. http://dx.doi.org/10.1186/1479-5868-5-56.

Aguilar-Farías, N., Brown, W. J., Olds, T. S., & Peeters, G. G. (2015). Validity of self-report methods for measuring sedentary behaviour in older adults. *Journal of Science and Medicine in Sport, 18*, 662–666. http://dx.doi.org/10.1016/j.jsams.2014.08.004.

American College of Sports Medicine. (2013). *ACSM's guidelines for exercise testing and prescription* (9th ed.). Philadelphia, PA: Lippincott, Williams & Wilkins.

Baecke, J. A., Burema, J., & Frijters, J. E. (1982). A short questionnaire for the measurement of habitual physical activity in epidemiological studies. *The American Journal of Clinical Nutrition, 36*, 936–942.

Boreham, C. A., Ferreira, I., Twisk, J. W., Gallagher, A. M., Savage, M. J., & Murray, L. J. (2004). Cardiorespiratory fitness, physical activity, and arterial stiffness: The Northern Ireland young hearts project. *Hypertension, 44*, 721–726. http://dx.doi.org/10.1161/01.HYP.0000144293.40699.9a.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290. http://dx.doi.org/10.1037/1040-3590.6.4.284.

Davies, C. A., Spence, J. C., Vandelanotte, C., Caperchione, C. M., & Mummery, W. K. (2012). Meta-analysis of internet-delivered interventions to increase physical activity levels. *International Journal of Behavioral Nutrition and Physical Activity, 9*, 52–63. http://dx.doi.org/10.1186/1479-5868-9-52.

Department of Health, Physical Activity, Health Improvement and Protection. (2016, March, 11). Start active, stay active: A report on physical activity from the four home countries' chief medical officers. https://www.gov.uk/government/publications/start-active-stay-active-a-report-on-physical-activity-from-the-four-home-countries-chief-medical-officers Accessed 05.05.16.

Dugdill, L., Graham, R. C., & McNair, F. (2005). Exercise referral: The public health panacea for physical activity promotion? A critical perspective of exercise referral schemes; their development and evaluation. *Ergonomics, 48*, 1390–1410. http://dx.doi.org/10.1080/00140130500101544.

Dunn, A. L., Andersen, R. E., & Jakicic, J. M. (1998). Lifestyle physical activity interventions: History, short-and long-term effects, and recommendations. *American Journal of Preventive Medicine, 15*, 398–412. http://dx.doi.org/10.1016/S0749-3797(98)00084-1.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika, 10*, 507–521. http://dx.doi.org/10.2307/2331838.

Godin, G., & Shephard, R. J. (1985). A simple method to assess exercise behavior in the community. *Canadian Journal of Applied Sport Sciences, 10*, 141–146.

Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W., Ekelund, U., et al. (2012). Global physical activity levels: Surveillance progress, pitfalls, and prospects. *The Lancet, 380*, 247–257. http://dx.doi.org/10.1016/S0140-6736(12)60646-1.

Helmerhorst, H. H. J., Brage, S., Warren, J., Besson, H., & Ekelund, U. (2012). A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *International Journal of Behavioral Nutrition and Physical Activity, 9*, 103–157. http://dx.doi.org/10.1186/1479-5868-9-103.

Jacobs, D. R., Jr., Ainsworth, B. E., Hartman, T. J., & Leon, A. S. (1993). A simultaneous

evaluation of 10 commonly used physical activity questionnaires. *Medicine and Science in Sports and Exercise, 25,* 81–91.

Karageorghis, C. I., Vencato, M. M., Chatzisarantis, N. L. D., & Carron, A. V. (2005). Development and initial validation of the Brunel lifestyle physical activity questionnaire. *British Journal of Sport Medicine, 39,* e23. http://dx.doi.org/10.1136/bjsm.2004.014258.

Karmali, K. N., & Lloyd-Jones, D. M. (2014). Achieving and maintaining cardiovascular health across the lifespan. *Current Epidemiology Reports, 1,* 75–81. http://dx.doi.org/10.1007/s40471-014-0011-7.

Lee, S. H., & Im, E. O. (2010). Ethnic differences in exercise and leisure time physical activity among midlife women. *Journal of Advanced Nursing, 66,* 814–827. http://dx.doi.org/10.1111/j.1365-2648.2009.05242.x.

Macfarlane, D. J., Lee, C. C., Ho, E. Y., Chan, K. L., & Chan, D. (2006). Convergent validity of six methods to assess physical activity in daily life. *Journal of Applied Physiology, 101,* 1328–1334. http://dx.doi.org/10.1152/japplphysiol.00336.2006.

Matthews, C. E., Steven, C. M., George, S. M., Sampson, J., & Bowles, H. R. (2012). Improving self-reports of active and sedentary behaviors in large epidemiologic studies. *Exercise and Sport Sciences Reviews, 40,* 118–126. http://dx.doi.org/10.1097/JES.0b013e31825b34a0.

Nevill, A. M., Lane, A. M., Kilgour, L. J., Bowes, N., & Whyte, G. P. (2001). Stability of psychometric questionnaires. *Journal of Sports Sciences, 19,* 273–278. http://dx.doi.org/10.1080/026404101750158358.

O'Connell, S. E., Griffiths, P. L., & Clemes, S. A. (2014). Seasonal variation in physical activity, sedentary behaviour and sleep in a sample of UK adults. *Annals of Human Biology, 41,* 1–8. http://dx.doi.org/10.3109/03014460.2013.827737.

Paffenbarger, R. S., Jr., Wing, A. L., & Hyde, R. T. (1978). Physical activity as an index of heart attack risk in college alumni. *American Journal of Epidemiology, 108,* 161–175.

Peels, D. A., Bolman, C., Golsteijn, R. H. J., de Vries, H., Mudde, A. N., van Stralen, M. M., et al. (2013). Long-term efficacy of a printed or a Web-based tailored physical activity intervention among older adults. *International Journal of Behavioral Nutrition and Physical Activity, 10,* 104–113. http://dx.doi.org/10.1186/1479-5868-10-104.

Pols, M. A., Peeters, P. H. M., Bueno-De-Mesquita, H. B., Ocke, M. C., Wentink, C. A., Kemper, H. C., et al. (1995). Validity and repeatability of a modified Baecke questionnaire on physical activity. *International Journal of Epidemiology, 24,* 381–388. http://dx.doi.org/10.1093/ije/24.2.381.

Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *British Journal of Sports Medicine, 37,* 197–206. http://dx.doi.org/10.1136/bjsm.37.3.197.

Sisson, S. B., McClain, J. J., & Tudor-Locke, C. (2008). Campus walkability, pedometer-determined steps, and moderate-to-vigorous physical activity: A comparison of 2 university campuses. *Journal of American College Health, 56,* 585–592. http://dx.doi.org/10.3200/JACH.56.5.585-592.

Speck, B. J., & Harrell, J. S. (2003). Maintaining regular physical activity in women: Evidence to date. *Journal of Cardiovascular Nursing, 18,* 282–291. http://dx.doi.org/10.1097/00005082-200309000-00007.

Streiner, D. L., Norman, G. R., & Cairney, J. (2014). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford, UK: Oxford University Press.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (5th ed.). Harlow, UK: Pearson Education.

Vencato, M. M., Karageorghis, C. I., Priest, D.-L., & Nevill, A. M. (2017). Concurrent validity and cross-validation of the Brunel Lifestyle Physical Activity Questionnaire. *Journal of Medicine and Science in Sport, 20,* 766–770.

Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of physical activity – a review of methodologies with reference to epidemiological research: A report of the exercise physiology section of the European association of cardiovascular prevention and rehabilitation. *European Journal of Cardiovascular Prevention & Rehabilitation, 17,* 127–139. http://dx.doi.org/10.1097/HJR.0b013e32832ed875.

Wendel-Vos, G. W., Schuit, A. J., Saris, W. H., & Kromhout, D. (2003). Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *Journal of Clinical Epidemiology, 56,* 1163–1169. http://dx.doi.org/10.1016/S0895-4356(03)00220-8.

Wikipedia. (2016). Climate of the United Kingdom. https://en.wikipedia.org/wiki/Climate_of_the_United_Kingdom Accessed 05.05.16.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics, 1,* 80–83.