Psychology of Sport and Exercise 33 (2017) 24–30

Contents lists available at ScienceDirect

Psychology of Sport and Exercise

journal homepage: www.elsevier.com/locate/psychsport

Test–retest reliability of the Brunel Lifestyle Physical Activity Questionnaire

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

ARTICLE INFO

Article history:
Received 21 December 2016
Received in revised form
21 June 2017
Accepted 7 July 2017
Available online 8 July 2017

Keywords:
Gender
Planned physical activity
Proportion of agreement
Unplanned physical activity

ABSTRACT

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.

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.

http://dx.doi.org/10.1016/j.psychsport.2017.07.003
1469-0292/

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).