

Review

# Measurement Characteristics of the Most Common Used International Physical Activity Questionnaires for Adults Across EU-28 in National Language Versions: Systematic Review and Meta Analysis

Vedrana Sember<sup>1\*</sup>, Kaja Meh<sup>1</sup>, Maroje Sorić<sup>1,2</sup>, Gregor Starc<sup>1</sup>, Paulo Rocha<sup>3</sup> and Gregor Jurak<sup>1</sup>

1 University of Ljubljana, Faculty of Sports, Ljubljana, Slovenia; [vedrana.sember@fsp.uni-lj.si](mailto:vedrana.sember@fsp.uni-lj.si); [kaja.meh@fsp.uni-lj.si](mailto:kaja.meh@fsp.uni-lj.si); [gregor.starc@fsp.uni-lj.si](mailto:gregor.starc@fsp.uni-lj.si); [gregor.jurak@fsp.uni-lj.si](mailto:gregor.jurak@fsp.uni-lj.si)

2 University of Zagreb, Faculty of Kinesiology, Zagreb, Croatia; [maroje.soric@kif.unizg.hr](mailto:maroje.soric@kif.unizg.hr)

3 Portugese Institute of Sport and Youth, Lisbon, Portugal; [paulo.rocha@ipdj.pt](mailto:paulo.rocha@ipdj.pt)

\* Correspondence: [vedrana.sember@fsp.uni-lj.si](mailto:vedrana.sember@fsp.uni-lj.si); tel.:(V.S.) +386 51 268263

Received: date; Accepted: date; Published: date

**Abstract:** This review and meta-analysis (PROSPERO registration number: CRD42020138845) critically evaluates test-retest reliability, concurrent validity and criterion validity of different physical activity (PA) levels of three most-common used international PA questionnaires (PAQs) in national language versions of EU-28: IPAQ-SF, GPAQ, and EHIS-PAQ. In total, 4969 abstracts were read, 287 full-text articles were identified as relevant to the study objectives, and 19 studies were included in the final meta-analysis. The studies' results and quality were evaluated using the Quality Assessment of Physical Activity Questionnaires (QAPAQ) checklist. Results indicate that only five EU-28 countries validated national language versions of selected PAQs. Meta analysis revealed that assessment of moderate-to-vigorous PA (MVPA) is the most relevant PA level outcome, since no publication bias in any of measurement properties was detected while test-retest reliability was the moderately-high ( $r_w=0.74$ ), and moderate for both criterion ( $r_w=0.41$ ) and concurrent validity ( $r_w=0.72$ ). Reporting of methods and results of the studies was poor with overall moderate risk of bias with total score of 0.43. In conclusion, where only self-reporting is feasible, assessment of MVPA with selected PAQs in EU-28 adult population is recommended for policymaking.

**Keywords:** validity; reliability; European Union; measurement properties; language version; IPAQ; GPAQ; EHIS-PAQ.

## 1. Introduction

Increasing the level of physical activity (PA) has become one of the priorities of public health policies in most developed countries of the world [1]. Over the last thirty years, we have witnessed an accelerated increase in the quantity of interventions to increase PA worldwide, although with limited effects [2,3]. Creating optimal policies and planning effective interventions aimed at increasing PA is not possible without reliable data on the prevalence of physical inactivity. Hence, numerous global authorities have called for concerted efforts in PA surveillance [4–7]. Conversely, how to execute PA monitoring is not entirely clear. Firstly, although methods for the assessment of PA are numerous, but given the complex nature of the PA, none of the currently available methods can assess all PA dimensions.

Based on the literature review, we can classify scientific methods for determining PA as direct observation or objectively assessed PA and indirect or subjectively assessed PA [8,9]. Large PA surveillance systems have, until recently, relied solely on PA questionnaires (PAQs) as one of the subjective assessment of PA. Questionnaires are easy to apply in large groups of individuals and are therefore the basic method of assessing PA in large epidemiological studies. However, this method

is subject to recall bias, which typically leads to overestimation of PA [10]. Therefore, some of the large PA surveillance systems have recently begun to rely on objective assessment by accelerometers to monitor activity levels [11]. Although the validity of accelerometers has been tested in numerous settings [12–14] and despite the fact that accelerometers have proved to be more reliable and valid than PA questionnaires [15–17], several shortcomings have to be noted, such as underestimation of energy expenditure during uphill walking, cycling, load carrying, etc. [18]. Additionally, other important issues for large surveillance systems might be costs [19], demanding data reduction procedures, obtrusiveness of the devices, which reduces compliance and increases non-wear time [20], specialized training required for assessors and the need for the physical proximity of participants. On the other hand, the advancement in technology has led to the development of commercial activity monitors for personal use. Recent evidence on accuracy of these devices indicates that this technology could be a very useful tool for surveillance systems [12,21–27], however at the moment PAQs still prevail [28–30].

Secondly, in designing a monitoring system for PA, a harmonised approach using a single, international instrument is preferred to enable cross-country comparisons. However, because PA is a behaviour, cultural environment should be taken into account when PA is assessed by questionnaires. Namely, most of PAQs rely on a person remembering activities they participated in or self-estimates of the intensity of the recalled PA. Therefore, the cultural context and country specific types of PA are very important for interpretation of questions and consequently for the content validity of a PAQ. Although some reviews and meta-analysis of measurement properties of PAQs have already been published [31–34], there is still lack of knowledge addressing this issue on European population which is very multi-national, multi-cultural and multi-linguistic.

Within project EUPASMOS, which aims to establish PA, sedentary behaviour patterns and sport participation monitoring framework in the European Union (EU) member states, we searched for studies describing measurement characteristics of nationally adapted versions of the most commonly used international PAQs across EU: i) International Physical Activity Questionnaire - Short form (IPAQ-SF), which is the most commonly used and validated PAQ [35]; ii) Global Physical Activity Questionnaire (GPAQ), which is used in more than 100 countries [36] and iii) European Health Interview Survey – Physical Activity Questionnaire (EHIS-PAQ), which was conducted within EU by all member states and is the only questionnaire that allows comparison of PA in different areas [29].

Selected PAQs have some common features, but many specifics. IPAQ is an instrument, which was developed to establish a standardized and culturally adaptable measurement tool for measuring PA in different cultural areas of the world [38]. The short form of IPAQ (IPAQ-SF) comprises nine items [36]. IPAQ-SF is an open-ended questionnaire, last 7-days recall, available in English and many other languages, covers four domains of PA (leisure time PA, domestic activities, work-related PA, transport-related PA) at each of four types of PA (sitting, walking, moderate-intensity activities and vigorous-intensity activities) [39]. The outcome of the IPAQ-SF is METmin/week and PA category score. Nevertheless long version of IPAQ (IPAQ-L) is slightly more reliable, it has proven to be too long and incomprehensible compared to IPAQ-SF [40], making its use more user-friendly. GPAQ was developed under the auspices of World Health Organization and uses a typical week recall, is somewhat shorter than IPAQ-L and includes different domains of PA. GPAQ has 16 questions and covers three domains of PA (work, transport and leisure) and sedentary behaviour [41]. GPAQ can differentiate between two intensities of PA (vigorous and moderate) [36]. Both GPAQ and IPAQ were designed to compare PA levels in different cultural settings around the world. On the other hand, EHIS-PAQ is EU-specific questionnaire within European Health Interview Survey. EHIS-PAQ is a domain-specific questionnaire with last 30-days recall, which covers three domains of PA (work-related, transport-related and leisure time), includes eight questions and distinguishes between aerobic and muscle-strengthening PA [37].

Therefore, the purpose of this systematic review and meta analysis is to critically appraise, compare and summarize the measurement properties (reliability, criterion validity, construct validity) of most commonly used PAQs for adults in EU-28 national language versions, taking the methodological quality of these studies, as well as the quality of evidence into account.

## 2. Materials and Methods

The meta-analysis was performed and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines [31,32]. The present work was registered at the International Prospective Register for Systematic Reviews, identification code CRD42020138845.

### 2.1. Search strategy

Identical search strategy was employed in PubMed, SportDiscus, Scopus, Dart and ResearchGate databases, examining studies describing measurement properties of three international PAQs through April and May 2018 and through April and May 2020. We used the following search string “name of the questionnaire e.g. IPAQ AND (valid\* OR reliab\* OR repeat\* OR reproducib\* OR assess\* OR measure\*)”. Additional studies were identified by manually searching the reference lists of the full papers identified during the search. Grey literature was additionally reviewed through ResearchGate, Google Scholar and Mendeley, using only keyword “name of the questionnaire e.g. IPAQ AND valid\* and through personal communication of members of the research team with other scientists. Additional literature that corresponds to the eligibility criteria of the present review was also obtained through online questionnaire on platform 1KA (University of Ljubljana, Faculty of Social Sciences) with the help of World Health Organization within EUPASMOS project activities. National HEPA focal points were asked to report on any national research, reports and doctoral thesis, published in their national languages that examined the measurement properties of any of the three PAQs included in this study. All articles generated from the initial search were stored on Mendeley reference management software & researcher network (Elsevier, Amsterdam, Netherlands) which removed duplicate references.

### 2.2. Eligibility criteria

Studies included in the present review had to be performed in healthy adults (18 yrs. old or older) and carried out in one of the European Union's countries (28 countries included) and published in one of the 24 EU's official languages. Studies were included if they were examining the following PAQs: IPAQ, GPAQ and EHIS-PAQ. and reported on: (i) PAQ translation protocol, (ii) mode of administration (interview, self-administered) and (iii) reliability or (iv) concurrent validity or (v) criterion validity of included PAQs. Studies performed in special populations (e.g. participants with specific medical conditions) were excluded.

Reliability refers to the consistency of a response either across multiple tests within a single assessment, generally called internal consistency, or across multiple assessments, known as test-retest or stability reliability [44]. The time interval between the test and retest must have been described and short enough that subject's PA cannot change, but long enough to prevent recall [35]. For PA assessment during current or past week, 1 day to 3 months recall period was considered appropriate, whereas for lifetime PA assessment, 1 day to 12 months recall period was considered appropriate [35].

### 2.3. Risk of bias assessment

The assessment of the risk of bias of included studies was conducted using the criteria, previously used by Sneek [45] and Sember [46], which includes the criterion of power calculations. Each study received '0' (does not meet the criterion) or '1' (meets the criterion) for each criterion based on an analysis of the reporting in the original article. Methodological quality was assessed following QAPAQ checklist [47], which was developed specifically for qualitative assessment of PA questionnaires. Risk of bias assessment and methodological quality was performed by two independent reviewers (V.S. and K.M.).

### 2.4. Data extraction and statistical analysis

Abstract, full-text articles, data extraction and quality assessment were performed by two independent reviewers (V.S. and K.M.) who also checked all databases and identified potential

studies through the search process to identify potentially relevant articles. In case of uncertainty, a third and fourth reviewer (G.J. and G.S.) screened the article. Summary tables of entered data were checked with the trial protocol and latest trial report or publication. Any discrepancies or unusual patterns were checked with the study principal investigator. Hunter-Schmidt estimate was used for reducing the amount of bias and Fisher's  $z$  transformation was applied to samples' correlation to display publication bias [42,43]. We also assessed publication bias with Egger bias test [48] for all PA constructs, separately for reliability, concurrent and criterion validity.

For further analysis correlation ( $r_w$ ) coefficients were determined by Hunter-Schmidt approach [49], which was multiplied by the sample size of each study ( $r_w \times N$ ). The generalizability of  $r_p$  was corrected using an artefact correction and variance sample. For weighted mean ( $r_w$ ), 95% credibility interval:  $CI_w = r_w + 1.96\sqrt{V_p}$  and  $I^2$  and  $Q$  statistics to measure heterogeneity of ES were calculated.

### 2.5. Data synthesis

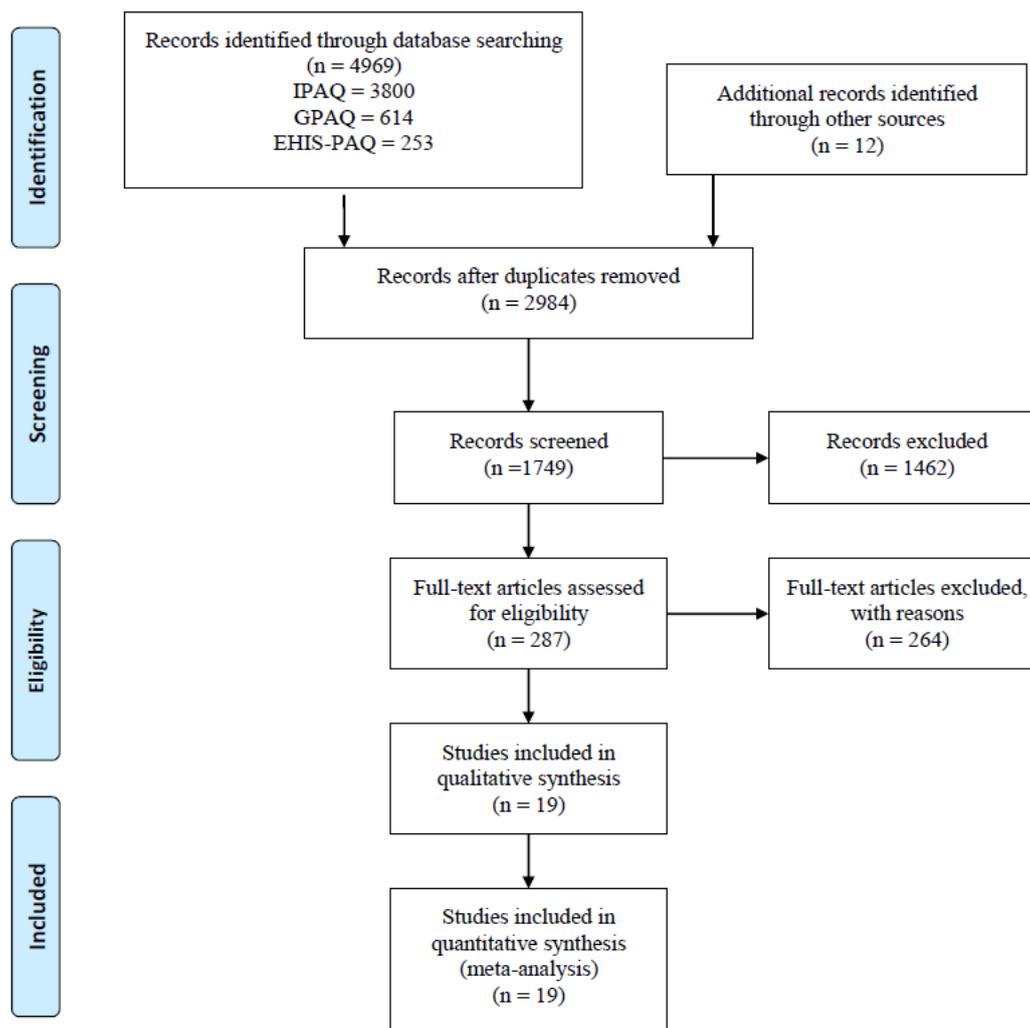
Results of 19 studies for moderate (MPA), vigorous (VPA) and total PA (tPA) were synthesised into four categories: (1) General characteristics of selected studies of PAQs across EU (Table 1); (2) Reliability of PAQs in selected studies across EU (Table 2); (3) Concurrent validity of PAQs in selected studies across EU (Table 2) and (4): Criterion validity of PAQs in selected studies across EU (Table 2).

### 2.6. Grading the level of evidence

Reliability levels of evidence were formulated following van Poppel and colleagues (2010) levels of evidence: (1) adequate time between test and retest and use of interclass correlation (ICC), Kappa or Concordance reliability score  $> 0.7$ ; (2) inadequate time interval between test and retest and use of ICC, Kappa or Concordance reliability score  $< 0.7$ , adequate time interval between test and retest, Pearson/Spearman correlation  $> 0.7$ ; (3) an inadequate time interval between test and retest, Pearson/Spearman correlation  $\leq 0.7$ . Additional grade was given due to number of participants and level of index or correlation. Positive score (+) was given for studies with  $> 50$  participants and the reliability coefficients were  $\geq 0.70$ . Negative (-) score was given for studies with  $< 50$  participants and the reliability coefficients were  $\leq 0.70$ . Pearson and Spearman correlation were considered inadequate due to systematic errors [50], therefore only ICC, Kappa or Concordance were deployed in level (1) of evidence. Validity is the degree to which an instrument measures constructs [47]. Highest level of criterion validity evidence would be comparing PAQs to gold standard – doubly labelled water (DLW) [51]. However, DLW also includes basal metabolic rate and the thermic effects of food, therefore use of other validated instrument are more reliable for obtaining construct validity. This will be done by comparing PAQs to other PAQs (concurrent validity) and accelerometers (criterion validity). For concurrent and criterion validity, research team established following levels of evidence: (1) concurrent validity score  $> 0.8$ ; (2)  $0.8 >$  validity score  $\geq 0.5$ ; (3) concurrent validity score  $< 0.5$ . Positive score (+) was given for studies with  $> 50$  participants and negative (-) score was given for studies with  $< 50$  participants.

## 3. Results

The flow of the review process is shown in Figure 1. In total, 4969 abstracts were read, 287 full-text articles were identified and read and 19 studies were finally included in present review (Figure 1). The characteristics of the included studies are presented in Table 1. We included studies from 17 different EU countries, mostly from United Kingdom (7), Spain (4) and Germany (3). Three studies were cross-national [38,52,53]. Table 1 represents information from all 19 studies included in present review of selected PAQs [36–38,52,54–70] including the country where the study was carried out, sample size, participants age and gender, sample description, modes and means of administration of selected studies.



**Figure 1.** Flowchart showing study identification process.

Altogether 5902 people in 22 different sub studies participated. Age range included participants in all studies was between 18 and 75 years. In 17 of 19 studies participant's gender proportion was included, whereas in two studies gender proportion is unknown [59,60]. Regarding sample description, 13 studies used convenient (68.4%), 4 random (21.05%), 1 quota (5.26%), 1 multistage stratified probability sampling (5.26%). Most of the studies (n=12) used self-administered mode of administration, 4 used interview and 2 telephone interviews. In one study, they used both self-administered questionnaire and interview. All of the studies assessed duration and frequency of physical activity. All interventions, meeting all predefined criteria, are presented in Table 1.

**Table 1.** General characteristics of selected studies of PAQs across EU.

Author (PAQ)	Country	Population**			Sample description	Dimensi on	Setting	Construct			Format		
		Size	Age; (range)	Gender (male, female)				Recall period	No. of Q	Mode and means of administration	Parameters	Scores	Unit of measurement
Baumeister et al. [37] (EHIS-PAQ) <i>German</i>	DE	140	55 (18-79)	<u>73</u> + 67	Random community sample	Sitting, LPA, MPA, VPA	Work-related PA, transport, leisure time, sport activities, HEPA, sedentary	30-days	9	Self-administered Unknown mode	Duration, frequency	MVPA, LPA	Min/day, MET*min
Bull et al. [36] (GPAQ), <i>Portugese</i>	PT	67	18 – 75	<u>17</u> + 50	Prevalence of young participants (18-44, n = 56) Convenient regional sample	Sitting, MPA, VPA	Work-related PA, transport, leisure time, sedentary	7-days	19	Interview Unknown mode	Duration, frequency	VPA, MPA, TPA, sedentary	Min
Cámara et al. [71] (GPAQ), <i>Spanish</i>	ES	163	70 (67–75)	<u>67</u> + 96	Older adults from IMPACT65+ study	Sitting	Sedentary time		1	Interview Face to face	Duration, frequency	Sedentary time	Min
Cleland et al. [66] (GPAQ), <i>English</i>	UK	22	46	<u>8</u> + 14	Random national sample	Sitting, MPA, VPA	Work-related PA, transport, leisure time, sedentary	7-days	16	Self-administered Unknown mode	Duration, frequency	MVPA, sedentary	Min/day
Craig et al. [38] (IPAQ-SF), <i>German, English, Finnish, Dutch, Portugese, Sweedish</i>	Cross-national: AU, UK, FI, NL, PT, SE	2115: 200 SE1 50 SE2; 149 UK1 101 UK2 88 FI 196 PT 74 NL	47 41 35 41 56 35 33	<u>77</u> + 123 <u>22</u> + 28 <u>68</u> + 81 <u>38</u> + 63 <u>43</u> + 45 <u>96</u> + 100 <u>34</u> + 40	Specific populations Convenient samples, but collectively, the participants represented a wide range of age, education, income, and activity levels	Sitting, MPA, VPA	Leisure time PA, domestic and gardening activities, work-related PA, transport-related PA	7-days	9	Self-administered Unknown modes	Duration, frequency	Categorical measure of % min/week	Min/week
Ekelund et al. [67] (IPAQ-SF), <i>Sweedish</i>	SE	185	42 (20 – 69)	<u>93</u> + 92	Workers and students Convenient regional sample	Sitting, MVPA	Leisure time PA, domestic and gardening activities, work-related PA, transport-related PA	7-days	7	Telephone interview	Duration, frequency	MVPA	MET min/day. MET min/week
Kalvenas et al. [69] (IPAQ-SF), <i>Lithuanian</i>	LT	92 †	18 – 69	reliability <u>29</u> + 63 validity <u>23</u> + 58	Employees of university and private company Convenient sample from urban area	Sitting, MPA, VPA	Leisure time PA, domestic and gardening activities, work-related PA, transport-related PA	7-days	9	Self-administered Unknown mode	Duration, frequency	VPA, MPA+walking, MPA, WPA, sitting, TPA	
Kastelic et al. [63] (GPAQ), <i>Slovenian</i>	SI	42	M 39 F 50	<u>37</u> + 5	Crane operators and office workers Convenient sample	Sitting, MPA, VPA	Work-related PA, transport, leisure time, sedentary	7-days	16	Interview Unknown mode	Duration, frequency	sedentary	Min/day
Kleinauskiene et al. [70] (IPAQ-SF), <i>Lithuanian</i>	LT	92	18-69	<u>29</u> + 63	Convenient sample from Kaunas city	Sitting, MPA, VPA	Leisure time PA, domestic and gardening activities, work-related PA,	7-days	9	Self-administered	Duration, frequency	MET min/week	MET, min/week

Laeremans et al. [52] (GPAQ), <i>German, Spanish, English</i>	Cross-national: B, ES, UK	122: 41 B; 41 ES; 40 UK	35	<u>55</u> + 67	Random regional sample	Sitting, MPA, VPA	transport-related PA Work-related PA, transport, leisure time, sedentary	7-days	16	Self-administered Online	Duration. frequency	MPA, MVPA, VPA, sedentary	MET min/week
Milton et al. [56] (GPAQ), <i>English</i>	UK	240	18-64	<u>119</u> + 121	Quota sample from across England, Scotland and Wales	Sitting, MPA, VPA	Work-related PA, transport, leisure time, sedentary Leisure time PA, domestic and gardening activities, work-related PA, transport-related PA	7-days	16	Telephone interview	Duration. frequency	MVPA	Min/day
Murphy et al. [57] (IPAQ-SF), <i>English</i>	IE	155 <sup>#</sup>	23	<u>69</u> + 86	Students Convenient sample	Sitting, MPA, VPA	Work-related PA, transport, leisure time, sedentary	7-days	9	Self-administered Unknown mode	Duration. frequency	MVPA as % in PA population	Min/week
Novak et al. [72] (GPAQ), <i>German</i>	AT	50	25	<u>39</u> + 11	Students Convenient sample	Sitting, Total PA, VPA	Work-related PA, transport, leisure time, sedentary	7-days	16	Self-administered Unknown mode	Duration. frequency	Total PA, VPA, sedentary	Min/week
Rivière et al. [62] (GPAQ), <i>French</i>	FR	87 <sup>##</sup>	30	<u>25</u> + 67	Medical personnel and students, convenience sample	Sitting, MPA, VPA	Work-related PA, transport, leisure time, sedentary	7-days	16	Interview and self-administered Unknown mode	Duration. frequency	LPA, VPA, TPA, MVPA	Min/day
Rudolf et al. [73] (GPAQ), <i>German</i>	DE	54	28	<u>23</u> + 31	University students Convenience sample	MPA, VPA, Sitting	Work-related PA, transport, leisure time, sedentary	7 days	16	Self-administered Online	Duration, frequency	MPA, VPA, sedentary	Min/day
Rütten et al. [53] (IPAQ-SF), <i>German, Finnish, French, Italian, Dutch, Spanish, English</i>	Cross-national: B, FI, FR, DE, I, NL, ES, UK	951: 100 B; 127 FI; 91 FR; 223 GR; 98 I; 86 N; 128 S; 98 UK	>18	unknown	Random sample	Sitting, MPA, VPA	Leisure time PA, domestic and gardening activities, work-related PA, transport-related PA	7-days	9	Interview Face-to face	Duration. frequency	VPA, MPA, sedentary	Min/week, MET
Scholes et al. [60] (IPAQ-SF), <i>English</i>	UK	1252	> 16	Unknown	Multistage stratified probability sampling	Sitting, MPA, VPA	Leisure time PA, domestic and gardening activities, work-related PA, transport-related PA	7-days	9	Self-administered Pen and paper	Duration. frequency	Categorical MVPA	Min/week
Taylor et al. [64] (IPAQ-SF), <i>English</i>	UK	49	27	<u>11</u> + 38	Students and university staff Convenient sample	Sitting, MPA, VPA	Leisure time PA, domestic and gardening activities, work-related PA,	7-days	9	Self-administered Online	Duration. frequency	MPA, MVPA	MET min/day

Vinas et al. [58] (IPAQ-SF), <i>Spanish</i>	ES	24	41	<u>26</u> + 29	Convenient sample 91% of the participants had a high level of education	Sitting, MPA, VPA	transport-related PA Leisure time PA, domestic and gardening activities, work-related PA, transport-related PA	7-days	9	Self-administered (Catalan version) Unknown mode	Duration. frequency	Min/day
--	----	----	----	----------------	---	-------------------------	---	--------	---	--	------------------------	---------

Notes: AT – Austria; ; B – Belgium; D – Denmark; DE – Germany; ES - Spain; FI – Finland; FR – France; GR – Greece; I – Italy; IE – Ireland; LT – Lithuania; NL – The Netherlands; NO – Norway; PT – Portugal; SE – Sweden; SI – Slovenia; UK – United Kingdom; VPA- vigorous PA; MPA-moderate to vigorous PA; WPA- walking PA; TPA – total PA; LPA – light PA; \*age was presented by mean or median; \*\* population (size, age, gender) was presented only for European country, nevertheless comparisons were made cross-national; #92 reliability and 81 validity; ## 133 reliability and 155 validity; ### 68 reliability and 87 criterion validity.

Table 2 represents information from 8 studies regarding reliability of PAQs in selected studies across EU [37,38,57,59,62,64,69,70], including information about measurement interval, results (Pearson  $r$ , Spearman  $\rho$ , Lin's concordance correlation and Phi coefficient) and quality rating. Most studies assessed test-retest reliability for MPA (30), and the least test-retest reliability for MVPA (5). The information for concurrent validity was reported in 9 PAQ studies across EU [36–38,56,60,62,64,71,72]. Information about comparison method, measured construct, correlation coefficient results and quality rating is shown in Table 2. Most of these studies assessed concurrent validity for tPA (11) and the least for MVPA (6). Table 2 represents information from 13 studies regarding criterion validity of PAQs in selected studies across EU [37,38,52,57,58,62–64,66,67,69,70,73], including information about country where study was carried out, duration of objective assessment, number of valid days and minutes per day, method for validity comparison, cut-off points, epoch length, definition of non-wear time and measured construct. Most studies assessed criterion validity for VPA and tPA (both 11) for MPA, the least studies assessed criterion validity for MPA and MVPA (both 9).

**Table 2.** Results for test-retest reliability, concurrent validity and criterion validity.

Reference (PAQ)	Study	Method	Construct (comparison method)*	Results	Rating
Baumeister et al. [37] (EHIS-PAQ)	DE	TRR	MVPA	ICC = 0.73	1
		CRV	MVPA (ActiGraph GT3X)	ICC = 0.32	3
		CCV	MVPA (IPAQ-L)	ICC = 0.45	2
			MVPA (7-d PAR)	ICC = 0.26	3
Bull et al. [36] (GPAQ)	PT	CCV	VPA (IPAQ-SF)	Spearman $\rho$ = 0.52	2
			MPA (IPAQ-SF)	Spearman $\rho$ = 0.50	2
			tPA (IPAQ-SF)	Spearman $\rho$ = 0.23	3
Cleland et al. [66] (GPAQ)	UK	CRV	MVPA (ActiGraph GT3X)	Spearman $\rho$ = 0.48	3
Craig et al. [38] (IPAQ)	SE 1	TRR	Total PA	Spearman $\rho$ = 0.66	3
			tPA 1st session (IPAQ L7T)	Spearman $\rho$ = 0.6	2
			tPA 2nd session (IPAQ L7T)	Spearman $\rho$ = 0.63	2
	UK1	TRR	tPA	Spearman $\rho$ = 0.87	2
			tPA	Spearman $\rho$ = 0.69	3
	UK2	TRR	tPA (CSA motion detector MTI)	Spearman $\rho$ = 0.40	3
			tPA	Spearman $\rho$ = 0.65	2
			tPA (CSA motion detector MTI)	Spearman $\rho$ = 0.47	3
	FI	TRR	tPA	Spearman $\rho$ = 0.68	2
			tPA 1st session (IPAQ LUS)	Spearman $\rho$ = 0.71	2
			tPA 2nd session (IPAQ LUS)	Spearman $\rho$ = 0.77	2
			tPA	Spearman $\rho$ = 0.49	3
	PT	TRR	tPA 1st session (IPAQ LUS)	Spearman $\rho$ = 0.43	3
			tPA 2nd session (IPAQ LUS)	Spearman $\rho$ = 0.77	2
			tPA	Spearman $\rho$ = 0.49	3
	SE 2	TRR	tPA	Spearman $\rho$ = 0.77	2
tPA (CSA motion detector MTI)			Spearman $\rho$ = 0.02	3	
tPA 1st session (IPAQ LUS)			Spearman $\rho$ = 0.77	2	
tPA 2nd session (IPAQ LUS)			Spearman $\rho$ = 0.87	2	
NL	TRR	tPA	Spearman $\rho$ = 0.85	2	
		tPA (CSA motion detector MTI)	Spearman $\rho$ = 0.32	3	
		tPA 1st session (IPAQ L7T)	Spearman $\rho$ = 0.85	1	
		tPA 2nd session (IPAQ L7T)	Spearman $\rho$ = 0.88	1	
Ekelund et al. [67] (IPAQ)	SE	CRV	MVPA (ActiGraph)	Pearson $r$ = 0.17	3
			tPA (ActiGraph)	Pearson $r$ = 0.34	3
Kalvenas et al. [69] (IPAQ)	LT	TRR	MPA (min/weak)	Spearman $\rho$ = 0.53	3
			VPA (min/weak)	Spearman $\rho$ = 0.67	3
			tPA (min/weak)	Spearman $\rho$ = 0.51	3
			VPA (ActiGraph GT3X)	Spearman $r$ = 0.40	3

			MPA (ActiGraph GT3X)	Spearman $r = -0.03$	3		
			tPA (ActiGraph GT3X)	Spearman $r = -0.11$	3		
Kleinauskiene [70] (IPAQ)	LT	TRR	MPA	Spearman $\rho = 0.35$	3		
			VPA	Spearman $\rho = 0.83$	2		
		CRV	weekly tPA 1st session	Spearman $\rho = 0.27$	3		
			weekly tPA 2nd session	Spearman $\rho = 0.06$	3		
Laeremans et al. [52](GPAQ)	B	CRV	MVPA (SWA)	Spearman $r = 0.56$	2		
			MVPA (SWA)	Spearman $r = 0.64$	2		
			MVPA (SWA)	Spearman $r = 0.55$	2		
			Overall MVPA (SWA)	Spearman $r = 0.54$	2		
		VPA (SWA)	Spearman $r = 0.62$	2			
		VPA (SWA)	Spearman $r = 0.69$	2			
		VPA (SWA)	Spearman $r = 0.59$	2			
		Overall VPA (SWA)	Spearman $r = 0.64$	2			
		MPA (SWA)	Spearman $r = 0.11$	3			
		MPA (SWA)	Spearman $r = 0.34$	3			
		MPA (SWA)	Spearman $r = 0.02$	3			
		Overall MPA (SWA)	Spearman $r = 0.34$	3			
		Murphy et al. [57](IPAQ)	IE	TRR	tPA	ICC = 0.77	2
					MVPA (ActiGraph GT1 M & GT3X)	Spearman $\rho = 0.31$	3
CRV	tPA (ActiGraph GT1 M & GT3X)			Spearman $\rho = 0.28$	3		
Novak et al. [72] (GPAQ)	AT	CCV	VPA (PAQ 24)	Spearman $\rho = 0.51$	2		
			MPA	Spearman $\rho = 0.56$	3		
				ICC = 0.48	3		
			Total VPA	Spearman $\rho = 0.8$	2		
				ICC = 0.84	1		
			Total PA	Spearman $\rho = 0.82$	2		
				ICC = 0.58	2		
			VPA (ActiGraph GT3X)	Spearman $\rho = 0.38$	3		
			VPA (ActiGraph GT3X)	Spearman $\rho = 0.10$	3		
			tPA (ActiGraph GT3X)	Spearman $\rho = 0.24$	3		
			VPA 1st session (IPAQ-LF)	Spearman $\rho = 0.86$	1		
			VPA 2nd session (IPAQ-LF)	Spearman $\rho = 0.76$	1		
			MPA 1st session (IPAQ-LF)	Spearman $\rho = 0.41$	3		
			MPA 2nd session (IPAQ-LF)	Spearman $\rho = 0.58$	2		
			tPA 1st session (IPAQ-LF)	Spearman $\rho = 0.66$	2		
			tPA 2nd session (IPAQ-LF)	Spearman $\rho = 0.67$	2		
			Rütten et al., [53](IPAQ)	B	TRR	tPA	ICC = 0.52
MPA	Spearman $\rho = 0.39$	3					
VPA days	Spearman $\rho = 0.55$	3					
VPA total minutes	Spearman $\rho = 0.44$	3					
tPA Sum MET (moderate, vigorous, walking)	Spearman $\rho = 0.53$	3					
FI	TRR	MPA days				Spearman $\rho = 0.28$	3
		MPA total minutes				Spearman $\rho = 0.55$	3
		VPA days				Spearman $\rho = 0.48$	3
		VPA total minutes				Spearman $\rho = 0.59$	3
		tPA Sum MET (moderate, vigorous, walking)				Spearman $\rho = 0.41$	3
FR	TRR	MPA days				Spearman $\rho = 0.18$	3
		MPA total minutes				Spearman $\rho = 0.28$	3
		VPA days				Spearman $\rho = 0.36$	3
		VPA total minutes				Spearman $\rho = 0.44$	3

			tPA Sum MET (moderate, vigorous, walking)	Spearman $\rho = 0.29$	3
	DE	TRR	MPA days	Spearman $\rho = 0.43$	3
			MPA total minutes	Spearman $\rho = 0.54$	3
			VPA days	Spearman $\rho = 0.51$	3
			VPA total minutes	Spearman $\rho = 0.54$	3
			tPA Sum MET (moderate, vigorous, walking)	Spearman $\rho = 0.39$	3
	I	TRR	MPA days	Spearman $\rho = 0.21$	3
			MPA total minutes	Spearman $\rho = 0.22$	3
			VPA days	Spearman $\rho = 0.41$	3
			VPA total minutes	Spearman $\rho = 0.53$	3
			tPA Sum MET (moderate, vigorous, walking)	Spearman $\rho = 0.14$	3
	NL	TRR	MPA days	Spearman $\rho = 0.40$	3
			MPA total minutes	Spearman $\rho = 0.34$	3
			VPA days	Spearman $\rho = 0.34$	3
			VPA total minutes	Spearman $\rho = 0.41$	3
			tPA Sum MET (moderate, vigorous, walking)	Spearman $\rho = 0.34$	3
	ES	TRR	MPA days	Spearman $\rho = 0.38$	3
			MPA total minutes	Spearman $\rho = 0.32$	3
			VPA days	Spearman $\rho = 0.54$	3
			VPA total minutes	Spearman $\rho = 0.62$	3
			tPA Sum MET (moderate, vigorous, walking)	Spearman $\rho = 0.58$	3
	UK	TRR	MPA days	Spearman $\rho = 0.25$	3
			MPA total minutes	Spearman $\rho = 0.43$	3
			VPA days	Spearman $\rho = 0.47$	3
			VPA total minutes	Spearman $\rho = 0.36$	3
			tPA Sum MET (moderate, vigorous, walking)	Spearman $\rho = 0.50$	3
	All nations	TRR	MPA days	Spearman $\rho = 0.36$	3
			MPA total minutes	Spearman $\rho = 0.39$	3
			VPA days	Spearman $\rho = 0.47$	3
			VPA total minutes	Spearman $\rho = 0.51$	3
			tPA Sum MET (moderate, vigorous, walking)	Spearman $\rho = 0.45$	3
Scholes et al. [60](IPAQ)	ES	CCV	MVPA (PASBAQ) male	Pearson $r = 0.43$	3
			MVPA (PASBAQ) female	Pearson $r = 0.40$	3
			MVPA minutes	Spearman $\rho = 0.67$	2
				ICC = 0.7	1
			Mean MVPA METs	Spearman $\rho = 0.79$	2
				ICC = 0.8	1
Taylor et al. [64](IPAQ)	UK	TRR	MPA total minutes	Spearman $\rho = 0.59$	3
				ICC = 0.57	2
			MPA METs	Spearman $\rho = 0.61$	3
				ICC = 0.58	2
			VPA min	Spearman $\rho = 0.71$	2
				ICC = 0.64	2
			VPA METs	Spearman $\rho = 0.71$	2
				ICC = 0.61	2
		CRV	MVPA METs (ActiGraph GT3X)	Spearman $\rho = 0.08$	3

			MVPA minutes (ActiGraph GT3X)	Spearman $\rho = 0.13$	3
			VPA METs (ActiGraph GT3X)	Spearman $\rho = 0.05$	3
			VPA (ActiGraph GT3X)	Spearman $\rho = 0.04$	3
			MPA METs (ActiGraph GT3X)	Spearman $\rho = 0.11$	3
			MPA (ActiGraph GT3X)	Spearman $\rho = 0.14$	3
			tPA (ActiGraph GT3X)	Spearman $\rho = 0.14$	3
		CCV	MPA MET (OSWEQ)	Spearman $\rho = 0.52$	2
			MPA (OSWEQ)	Spearman $\rho = 0.46$	3
			VPA (OSWEQ)	Spearman $\rho = 0.53$	2
			VPA METs (OSWEQ)	Spearman $\rho = 0.53$	2
			MVPA (OSWEQ)	Spearman $\rho = 0.56$	2
			MVPA METs (OSWEQ)	Spearman $\rho = 0.62$	2
Vinas et al.[58] (IPAQ)	UK	CRV	VPA (ActiGraph)	Spearman $r = 0.38$	3
			tPA (ActiGraph)	Spearman $r = 0.27$	3

Notes: TRR – test retest reliability; CRV – criterion validity; CCT – concurrent validity; AT – Austria; B – Belgium; D – Denmark; DE – Germany; ES - Spain; FI – Finland; FR – France; GR – Greece; I – Italy; IE – Ireland; LT – Lithuania; NL – The Netherlands; NO – Norway; PT – Portugal; SE – Sweden; SI – Slovenia; UK – United Kingdom; VPA- vigorous PA; MVPA-moderate to vigorous PA; TPA – total PA; LPA – light PA.

Based on weighted correlation means, measurement construct test-retest performed the best in construct MVPA ( $r_w=0.74$ ), where 3 associations (of 5) were graded with level of evidence 1 ( $r_w=0.74$ ) and 2 with levels of evidence 2 ( $r_w=0.73$ ); whereas the worst in MPA ( $r_w=0.40$ ) (Table 3), where 28 of 30 associations were graded with level of evidence 3 ( $r_w=0.41$ ) and only 2 with grade 2 ( $r_w=0.58$ ).

Based on weighted correlation means, concurrent validity was best for VPA ( $r_w=0.72$ ), where 4 associations were graded with levels of evidence 1 ( $r_w=0.82$ ) and 5 associations with levels of evidence 2 ( $r_w=0.62$ ) (Table 3). Concurrent validity was the lowest for tPA ( $r_w=0.22$ ), where 9 associations were evaluated with levels of evidence 2 ( $r_w=0.64$ ) and 2 with levels of evidence 3 ( $r_w=0.38$ ). On the other hand, VPA showed the highest validity ( $r_w=0.72$ ), but it should be noted that the Egger test (-5.63) showed significant bias between included correlations coefficients in VPA ( $p<0.0001$ ).

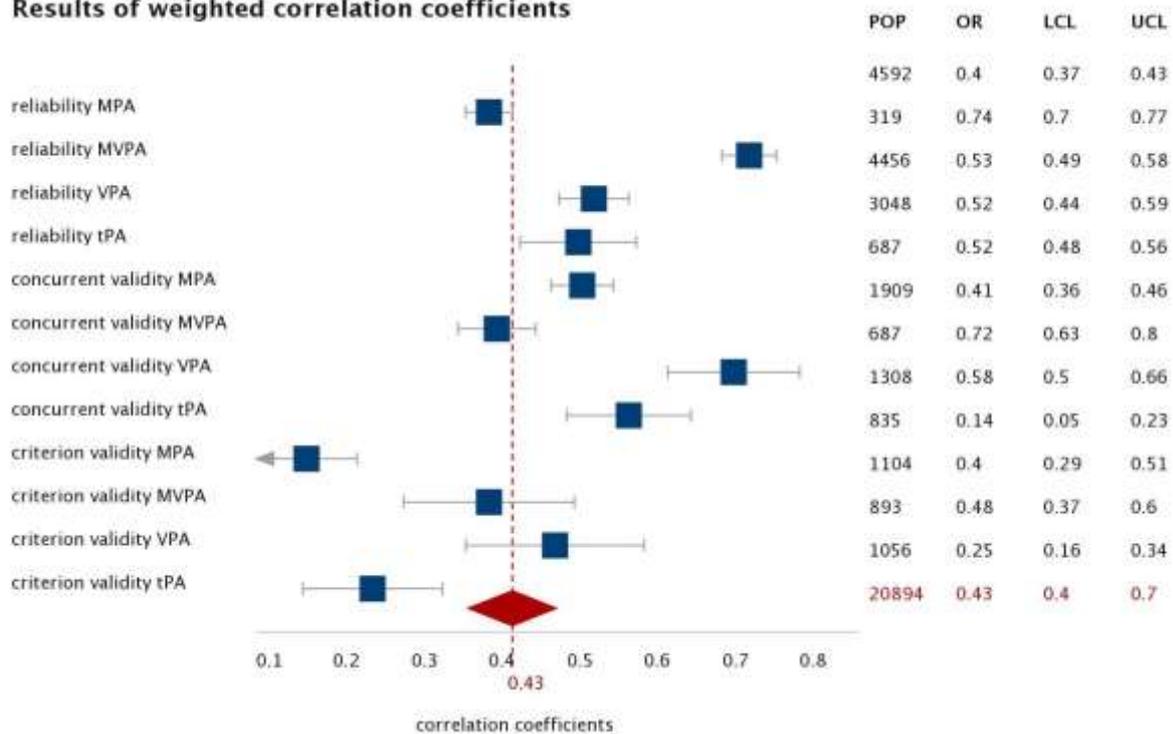
Based on weighted correlation means, measurement construct performed the best for VPA ( $r_w=0.48$ ), where 4 associations was evaluated with level of evidence 2 ( $r_w=0.64$ ) and 7 associations with grade 3 ( $r_w=0.30$ ); whereas the worst criterion validity was noted for MPA ( $r_w=0.14$ ) (Table 3), with all 9 associations graded with level of evidence 3. Once again, although the highest criterion validity was noted for VPA, the Egger test (-5.59) showed significant bias between included correlations coefficients in VPA ( $p<0.0001$ ). Results of weighted correlation coefficients for test-retest reliability, concurrent validity and criterion validity across all included studies stratified by PA intensity are presented in Figure 2.

**Table 3.** Summary results for test-retest reliability, concurrent validity and criterion validity across all included studies stratified by PA intensity.

Measurement characteristic	PA construct	Sample			Population effect				Egger bias test			Heterogeneity		
		N (k)	k	n	Unweighted mean	Weighted mean	95% CI	80% CRI	bias	95% CI	p	I <sup>2</sup> (%)	Q	p
Reliability (test-retest)	MPA	5	30	4592	0.42	0.40	0.37 to 0.43	0.32 to 0.47	0.52	-0.52 to 1.54	0.34	46.34	54.05	0.00
	MVPA	2	5	319	0.74	0.74	0.70 to 0.77	0.74 to 0.74	-0.46	-3.26 to 2.34	0.77	36.45	2.93	0.57
	VPA	3	28	4456	0.57	0.53	0.49 to 0.58	0.39 to 0.67	-0.30	-2.75 to 2.14	0.81	70.41	131.16	0.00
	tPA	5	19	3048	0.55	0.52	0.44 to 0.59	0.33 to 0.71	-0.71	-4.22 to 2.80	0.70	87.52	144.28	0.00
Concurrent validity	MPA	3	9	687	0.51	0.52	0.48 to 0.56	0.52 to 0.52	-2.53	-5.56 to 0.51	0.15	59.10	5.03	0.76
	MVPA	3	6	1909	0.43	0.41	0.36 to 0.46	0.34 to 0.47	0.41	-1.92 to 2.73	0.74	52.33	14.69	0.04
	VPA	3	9	687	0.69	0.72	0.63 to 0.80	0.56 to 0.87	-5.63	-6.80 to -4.46	0.00	84.75	52.47	0.00
	tPA	8	11	1308	0.61	0.58	0.50 to 0.66	0.43 to 0.74	-0.14	-6.47 to 6.20	0.97	55.30	81.92	0.00
Criterion validity	MPA	4	9	835	0.13	0.14	0.05 to 0.23	0.03 to 0.25				47.65	15.28	0.05
	MVPA	6	9	1104	0.38	0.40	0.29 to 0.51	0.19 to 0.60	-4.34	-9.94 to 1.26	0.17	82.09	50.25	0.00
	VPA	6	11	893	0.41	0.48	0.37 to 0.60	0.26 to 0.71	-5.59	-7.38 to -3.81	0.00	82.67	57.68	0.00
	tPA	8	11	1056	0.22	0.25	0.16 to 0.34	0.09 to 0.41	-3.22	-6.55 to 0.11	0.09	66.20	29.56	0.00

N - nr. of participants for selected PA construct and measurement characteristic; k-number of included constructs; n - nr. of participants; CI – confidence interval; CRI – credibility interval; I<sup>2</sup>- I index of heterogeneity; Q- chi-square test of heterogeneity; MPA – moderate PA; MVPA – moderate to-vigorous PA; VPA – vigorous PA; tPA – total PA.

## Results of weighted correlation coefficients



**Figure 2.** Forest plot of weighted correlation coefficients for measurements characteristics stratified by PA intensity.

The Egger bias test [48] provided evidence for publication bias for following measurement characteristics and PA constructs: concurrent validity VPA (bias= -5.63, 95% CI: -6.80 to -4.46,  $p < 0.0001$ ), concurrent validity tPA (bias= -0.14, 95% CI: 6.47 to 6.20,  $p = 0.97$ ), criterion validity VPA (bias= -5.59, 95% CI: -7.38 to -3.81,  $p < 0.0001$ ) and criterion validity tPA (bias= -3.22, 95% CI: -6.55 to 0.11,  $p = 0.09$ ) (Table 3). The results of risk-of-bias assessment are shown in Table 4. Total average risk of bias of all included studies was moderate (0.43). Of the 19 studies, only two were rated as having a low risk of bias ( $\geq 67\%$  of total score) with average of 0.73 of total score; 9 were rated as having moderate risk of bias ( $> 33$  and  $< 67\%$  of the total score) with average of 0.45 of total score and 8 studies were rated as having a high risk of bias ( $\leq 33\%$  of total score) with average of 0.32 of total score. Only 6 studies (33 %) reported power calculations to determine sufficient sample size and only 3 studies met the assumption of randomization, which is actually not so important to determine reliability and validity of questionnaires [74].

**Table 4.** Results of the risk-of bias assessment.

Author (year)	Outcome	R	BC	BV	T	BM	VO	DA	RR	PC	Total
Baumeister (2016) [37]	EHIS * + -	0	0	1	1	0	1	0	1	1	5/9 (0.56)
Bull et al. (2009) [36]	GPAQ +	0	0	1	1	0	1	0	1	0	4/9 (0.44)
Cámara et al 2020 [71]	GPAQ +	0	0	1	0	0	1	0	1	0	3/9 (0.33)
Cleland et al. (2014) [66]	GPAQ -	1	0	1	1	0	1	1	1	1	7/9 (0.78)
Craig et al. (2003) [38]	IPAQ * + -	0	0	0	0	0	1	0	1	0	2/9 (0.22)
Ekelund et al (2005) [67]	IPAQ -	1	0	0	1	0	1	0	0	0	3/9 (0.33)
Kalvenas et al (2016) [69]	IPAQ * -	0	0	0	1	0	1	0	1	0	3/9 (0.33)
Kastelic et al. (2019) [63]	GPAQ -	0	0	0	1	0	1	0	1	0	3/9 (0.33)
Kleinauskienė (2012) [70]	IPAQ * -	0	0	0	1	0	1	0	1	0	3/9 (0.33)
Laeremans et al (2016) [5]	GPAQ -	0	0	1	1	0	1	0	0	0	3/9 (0.33)
Milton et al (2009) [56]	GPAQ +	0	0	1	1	0	1	0	1	0	4/9 (0.44)
Murphy et al (2017) [57]	IPAQ * -	0	0	1	1	0	1	0	1	0	4/9 (0.44)
Novak et al. (2020) [72]	GPAQ +	0	0	0	1	0	1	0	1	1	4/9 (0.44)
Rivière et al. (2016) [62]	GPAQ * + -	0	0	0	1	0	1	0	1	1	4/9 (0.44)
Rudolf et al. (2020) [73]	GPAQ -	0	0	0	1	0	1	1	1	0	4/9 (0.44)
Rütten et al (2003) [53]	IPAQ *	1	0	1	1	0	1	0	1	1	6/9 (0.67)
Scholes et al (2016) [60]	IPAQ +	1	0	0	1	0	1	0	1	0	4/9 (0.44)
Taylor et al (2013) [64]	IPAQ * + -	0	0	0	1	0	1	0	1	1	4/9 (0.44)
Vinas (2012) [58]	IPAQ -	0	0	0	1	0	1	0	1	0	3/9 (0.33)
AVERAGE OF ALL STUDIES		0.21	0	0.42	0.89	0	1	0.11	0.89	0.32	0.43

Legend: R-randomization; BC – Baseline comparable; BV - Baseline values accounted for in analyses; T – timing; BM – blinding of measures; VO – validated outcome measures; DA – dropout analysis; RR – reporting of results; PC – power calculation; Total - total score of the risk of bias (decimal format)\* outcome for test-retest reliability; + outcome for concurrent validity; - outcome for criterion validity.

#### 4. Discussion

This systematic review and meta-analysis investigated the test-retest reliability, concurrent validity and criterion validity of three most common used PAQs across EU-28 in national language versions: IPAQ-SF, GPAQ and EHIS-PAQ. We identified 19 studies that adequately tested selected PAQs in recent 17-year period between 2003 and 2020.

The main findings include: i) GPAQ and IPAQ-SF are one of the most widely used international PAQs in EU [35], resulting in more than 90% of records identified through search strategy; ii) IPAQ and GPAQ are validated for MPA, MVPA and VPA in only 5 countries across EU-28; iii) the assessment of MVPA is the most relevant PA outcome, since no publication bias in any of measurement characteristics was detected and test-retest reliability was the moderately- high ( $r_w=0.74$ ), and moderate for both criterion ( $r_w=0.41$ ) and concurrent validity ( $r_w=0.72$ ); iv) reporting of methods and results of the studies was poor, leading to high risk of bias in 8 studies, moderate risk of bias in 9 studies and resulting in overall moderate risk of bias with total score of 0.43.

Our results revealed that MPA reached lowest overall correlations for all measurement characteristics (reliability  $r_w=0.42$ ; concurrent validity  $r_w=0.51$ ; criterion validity  $r_w=0.13$ ). VPA reached the highest overall correlations (reliability  $r_w=0.53$ ; concurrent validity  $r_w=0.72$ ; criterion validity  $r_w=0.48$ ) but we also found publication bias in concurrent and criterion validity for mentioned PA construct. All measurement characteristics were moderately- high for MVPA (reliability  $r_w=0.74$ ; concurrent validity  $r_w=0.41$ ; criterion validity  $r_w=0.40$ ). Since we did not detect publication bias in any of measurement characteristics for MVPA, we suggest the assessment of MVPA to be the most relevant PA outcome. To the larger extent, research findings indicate that MVPA in particular, positively influence the public health of adult population, which also resulted in the development of recommendations for policymakers to increase PA of the European population [1].

Although there is no single rule of thumb related to adequate sample size, test-retest interval and statistical analysis, academics have recommended the acceptable ratio of survey items and participants to be 1:5 [34,76], the use of ICC and Pearson correlation coefficient [47] and test-retest interval between three and eight days [76]. Based on our qualitative rating, only 8 of 305 PA

constructs within different measurement characteristics received grade 1, 142 constructs were graded with grade 2 and 145 with grade 3. This is a foundation for concern, since more than half constructs did not satisfied preferred recommendations for assessing reliability and validity in PAQs and calls for more rigorous study design in future reliability and validity investigations.

It is promising that the reliability was found to be moderate to high ( $r_w=0.40$  to  $0.74$ ). Of greater importance, time intervals with the exception of two studies [59,77] were within optional range [76] of test-retest interval and ranged mostly between three to eight days. Since the reliability of MVPA and tPA was high even in the two aforementioned studies [77,78] that used one month interval between repeated assessments, this methodological weakness [34] does not hamper the conclusions of this study.

PAQs showed low to moderately-high validity ( $r_w=0.14$  to  $0.48$ ) against measures of objectively measured PA and moderate to high validity against subjective measures of PA (other PAQs). Our results are comparable to results reported previously [32,79] that showed validity of PAQs ranging from 0.1 to 0.50 against objective measures of PA [80]. However, it should be noted that the criterion validity was validated in only 5 different national versions for IPAQ (Ireland, Lithuania, Spain, Sweden and United Kingdom) and 4 different national versions for GPAQ (Austria, Belgium, Spain and United Kingdom) across EU-28. Results are indicating differences in validity between all of versions, therefore other countries assessing PA don't even distinguish how valid their assessed data are. Moreover, factors explaining the variation in the validity of PAQs may relate to differences in qualitative attributes of PAQs, such as recall period and number of items as well as heterogeneity of population. It is well documented, that there are differences in prevalence of overweight and obesity [81] and physical fitness level between different nations and countries [82], which is the governing factor to assess PA with questionnaire. PAQs are assessing the subjective perception of PA, which is conditioned by physical fitness. Accordingly, it is exceptional that only few studies reported the reliability and validity of PAQ related to body mass index [36,66], whereas we haven't found a single study related to physical fitness or fatness. We believe one of important factors effecting variability of PAQs's validity are different physical fitness levels of participants, therefore inclusion of these factors might allow a more objective assessment of PA, as well better international comparability of the PA data. Rather low concurrent validity scores found in our study may be explained by different recall period in investigated PAQs. Subsequently, objective measures of PA are less dependent on long-term variation, and can more accurately capture sporadic and intermittent behaviours [32] which results in higher validity of measured PA construct, but lower criterion validity of PAQs. Moreover, it was often blurred what dimension of PA the PAQ was supposed to measure, which made assessing concurrent validity sometimes impossible. Moreover, it was extremely difficult to assess whether the same or somewhat modified versions of PAQs were used in some studies, and it was not always clear whether the data were derived from a self-report questionnaire or whether the questionnaire was part of an interview [35]. Nevertheless, most of the studies enthusiastically concluded that PAQ is valid, but they did not take into account risk of bias and quality assessment. However, when we applied criteria for risk of bias and quality assessment, we found this conclusions to be over-optimistic which is in concordance with previous reviews [35].

#### 4.1. Limitations

There are several limitations of this study that should be acknowledged: i) although we systematically searched five biggest databases in the field of PA twice and by more investigators, it is possible that not all relevant studies are included in present meta-analysis; ii) the most commonly used PAQs in present review were IPAQ and GPAQ. GPAQ uses a typical week to assess PA data. However, a typical week can be different in many European countries due to weather conditions, yielding different PA level, and iii) even though the quality of each study was assessed, findings from studies with lower quality were given no less importance than the other findings.

## 5. Conclusion

This meta-analysis examined reliability and validity of the most commonly used international PAQs across EU-28 in national language versions and identified the widespread use of GPAQ and IPAQ-SF, which significantly differ in type, duration, recall period and complexity. Our results revealed that the quality of most studies was poor. The reliability was higher for all included PA constructs than validity, whilst MVPA reached highest scores in test-retest reliability and generally performed the best in all measurement characteristics. Thus, where only self-report is affordable due to time availability of participants and resources; assessment of MVPA with GPAQ and IPAQ-SF in European adult population is recommended. When possible, combination of objective measures of PA and self-report will provide the most valid and reliable results of MVPA. Further investigation of measurement characteristics of PAQs should focus on studying differences in validity of different groups of participants according to their physical fitness and/or fatness since results from PAQs represent subjective assessment of PA which is closely related with fitness and fatness. This could contribute to more critical evaluation of PA data, assessed by PAQs.

**Author Contributions:** Conceptualization, M.S., G.S., V.S. and G.J.; methodology, V.S.; software, V.S.; formal analysis, V.S.; investigation, V.S. and K.M.; resources, G.J.; data curation, V.S. and K.M.; writing—original draft preparation, V.S.; writing—review and editing, V.S., P.R., M.S. and G.J.; visualization, V.S.; supervision, G.J. and G.S.; project administration, G.J.; funding acquisition, P.R. and G.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was co-funded by the Erasmus+ Programme of the European Union within the project EUPASMOS No 590662-EPP-1-2017-1-PT-SPO-SCP and Slovenian Research Agency within the Research programme Bio-psycho-social context of kinesiology No P5-0142.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Acknowledgments:** The authors acknowledge the support of the HEPA Europe national focal points and other national representatives who provided information on PAQs validated in their countries.

## References

1. Hallal, P.C.; Andersen, L.B.; Bull, F.C.; Guthold, R.; Haskell, W.; Ekelund, U. Global physical activity levels: surveillance progress, pitfalls, and prospects. *Lancet* **2012**, *380*, 247–257, doi:10.1016/S0140-6736(12)60646-1.
2. De Meester, F.; van Lenthe, F.J.; Spittaels, H.; Lien, N.; De Bourdeaudhuij, I. Interventions for promoting physical activity among European teenagers: a systematic review. *Int. J. Behav. Nutr. Phys. Act.* **2009**, *6*, 82, doi:10.1186/1479-5868-6-82.
3. Baranowski, T. Increasing physical activity among children and adolescents: Innovative ideas needed. *J. Sport Heal. Sci.* **2019**, *8*(1), 1–5. , doi:10.1016/j.jshs.2018.09.011.
4. Andersen, L.B.; Andersen, S.A.; Bachl, N.; Banzer, W.; Brage, S.; Brettschneider, W.-D.; Ekelund, U.; Fogelholm, M.; Froberg, K.; Gil-Antunano, N.P. EU Physical Activity Guidelines: Recommended policy Actions in Support of Health-Enhancing Physical Activity Fourth Consolidated Draft, Approved by the EU-working group. **2008**.
5. Organización Mundial de la Salud Physical activity strategy for the WHO European Region 2016-2025. *Organ. Mund. la Salud* **2015**, *32*, doi:978 92 890 5147 7.
6. Oja, P.; Fogelholm, M.; Martin, B.W. Physical activity recommendations for health: what should Europe do? *BMC Public Health* **2010**, 1–5.
7. European Commission. EU Action Plan on Childhood Obesity 2014-2020 1 : A growing health challenge for the EU. *Com. Eur.* **2014**, 1–68.
8. Kohl, H.W.; Fulton, J.E.; Caspersen, C.J. Assessment of physical activity among children and adolescents: A review and synthesis. *Prev Med* **2000**, *31*(2), 54–76, doi:10.1006/pmed.1999.0542.
9. Welk, G.J.; Corbin, C.B.; Dale, D. Measurement issues in the assessment of physical activity in children. *Res. Q. Exerc. Sport* **2000**, *71*(Sup 2), 59–73, doi:10.1080/02701367.2000.11082788.
10. Helmerhorst, H.H.J.F.; Brage, S.; Warren, J.; Besson, H.; Ekelund, U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. *Int. J. Behav. Nutr. Phys. Act.* **2012**, *9*, 103, doi:10.1186/1479-5868-9-103.
11. Pedišić, Ž.; Bauman, A. Accelerometer-based measures in physical activity surveillance: current practices and issues. *Br J Sport. Med* **2015**, *49*, 219–223, doi:10.1136/bjsports-2013-093407.

12. Ferguson, T.; Rowlands, A. V.; Olds, T.; Maher, C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *Int. J. Behav. Nutr. Phys. Act.* **2015**, *12*, 42, doi:10.1186/s12966-015-0201-9.
13. Corder, K.; Ekelund, U.; Steele, R.M.; Wareham, N.J.; Brage, S. Assessment of physical activity in youth. *J. Appl. Physiol.* **2008**, *105*, 977–987, doi:10.1152/jappphysiol.00094.2008.
14. Gastin, P.B.; Cayzer, C.; Dwyer, D.; Robertson, S. Validity of the ActiGraph GT3X+ and BodyMedia SenseWear Armband to estimate energy expenditure during physical activity and sport. *J. Sci. Med. Sport* **2018**, *21*(3), 291–295, doi:10.1016/j.jsams.2017.07.022.
15. Kohl, H.W.; Cook, H.D.; Van Dusen, D.P.; Kelder, S.H.; Kohl, H.W.; Ranjit, N.; Perry, C.L. *Educating the study body: taking physical activity and physical education to school. Chapter 4: Physical Activity, Fitness, and Physical Education: Effects on Academic Performance*; 2013; Vol. 81; ISBN 9780309283137.
16. Skender, S.; Ose, J.; Chang-Claude, J.; Paskow, M.; Brühmann, B.; Siegel, E.M.; Steindorf, K.; Ulrich, C.M. Accelerometry and physical activity questionnaires—a systematic review. *BMC Public Health* **2016**, *16*, 515, doi:10.1186/s12889-016-3172-0.
17. Westerterp, K.R. Assessment of physical activity: a critical appraisal. *Eur. J. Appl. Physiol.* **2009**, *105*, 823–828, doi:10.1007/s00421-009-1000-2.
18. Sirard, J.R.; Pate, R.R. Physical Activity Assessment in Children and Adolescents. *Sport. Med.* **2001**, *31*, 439–454, doi:10.2165/00007256-200131060-00004.
19. Prince, S.; Adamo, K.; Hamel, M.; Hardt, J.; Gorber, S.; Tremblay, M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int. J. Behav. Nutr. Phys. Act.* **2008**, *5*, 56, doi:10.1186/1479-5868-5-56.
20. Lee, I.-M.; Shiroma, E.J. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *Br J Sport. Med* **2014**, *48*, 197–201, doi:10.1136/bjsports-2013-093154.
21. An, H.-S.; Jones, G.C.; Kang, S.-K.; Welk, G.J.; Lee, J.-M. How valid are wearable physical activity trackers for measuring steps? *Eur. J. Sport Sci.* **2017**, *17*(3), 360–368, doi:10.1080/17461391.2016.1255261.
22. Bai, Y.; Welk, G.J.; Nam, Y.H.; Lee, J.A.; Lee, J.-M.; Kim, Y.; Meier, N.F.; Dixon, P.M. Comparison of consumer and research monitors under semistructured settings. *Med Sci Sport. Exerc* **2016**, *48*, 151–158, doi:10.1249/MSS.0000000000000727.
23. Lee, J.-M.; Kim, Y.-W.; Welk, G.J. TRACK IT: Validity and utility of consumer-based physical activity monitors. *ACSMs. Health Fit. J.* **2014**, *18*(4), 16–21, doi: 10.1249/FIT.0000000000000051.
24. Nelson, M.B.; Kaminsky, L.A.; Dickin, D.C.; Montoye, A.H.K. Validity of consumer-based physical activity monitors for specific activity types. *Med. Sci. Sport. Exerc.* **2016**, *48*(8), 1619–1628, doi:10.1249/MSS.0000000000000933.
25. Sasaki, J.E.; Hickey, A.; Mavilia, M.; Tedesco, J.; John, D.; Keadle, S.K.; Freedson, P.S. Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. *J. Phys. Act. Heal.* **2015**, *12*(2), 149–154, doi:10.1123/jpah.2012-0495.
26. Gomersall, S.R.; Ng, N.; Burton, N.W.; Pavey, T.G.; Gilson, N.D.; Brown, W.J. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and ActiGraph accelerometry. *J. Med. Internet Res.* **2016**, *18*(9), e239, doi:10.2196/jmir.5531.
27. Price, K.; Bird, S.R.; Lythgo, N.; Raj, I.S.; Wong, J.Y.L.; Lynch, C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. *J. Med. Eng. Technol.* **2017**, *41*(3), 208–215, doi:10.1080/03091902.2016.1253795.
28. Baptista, F.; Santos, D.A.; Silva, A.M.; Mota, J.; Santos, R.; Vale, S.; Ferreira, J.P.; Raimundo, A.M.; Moreira, H.; Sardinha, L.B. Prevalence of the Portuguese population attaining sufficient physical activity. *Med. Sci. Sport. Exerc.* **2012**, *44*(3), 466–473, doi:10.1249/MSS.0b013e318230e441.
29. Troiano, R.P.; Berrigan, D.; Dodd, K.W.; Mâsse, L.C.; Tilert, T.; Mcdowell, M. Physical activity in the United States measured by accelerometer. *Med. Sci. Sports Exerc.* **2008**, *40*, 181–188, doi:10.1249/mss.0b013e31815a51b3.
30. Dalene, K.E.; Anderssen, S.A.; Andersen, L.B.; Steene-Johannessen, J.; Ekelund, U.; Hansen, B.H.; Kolle, E. Secular and longitudinal physical activity changes in population-based samples of children and adolescents. *Scand. J. Med. Sci. Sport.* **2018**, *28*, 161–171, doi:10.1111/sms.12876.
31. Kim, Y.; Park, I.; Kang, M. Convergent validity of the International Physical Activity Questionnaire (IPAQ): Meta-analysis. *Public Health Nutr.* **2013**, *16*(3), 440–452, doi:10.1017/S1368980012002996.
32. Bakker, E.A.; Hartman, Y.A.W.; Hopman, M.T.E.; Hopkins, N.D.; Graves, L.E.F.; Dunstan, D.W.; Healy, G.N.; Eijsvogels, T.M.H.; Thijssen, D.H.J. Validity and reliability of subjective methods to assess sedentary behaviour in adults: a systematic review and meta-analysis. *Int. J. Behav. Nutr. Phys. Act.* **2020**, *17*, 1–31, doi:10.1186/s12966-020-00972-1.
33. Lee, P.H.; Macfarlane, D.J.; Lam, T.H.; Stewart, S.M. Validity of the international physical activity

- questionnaire short form (IPAQ-SF): A systematic review. *Int. J. Behav. Nutr. Phys. Act.* **2011**, *8*, 115, doi:10.1186/1479-5868-8-115.
34. Keating, X.D.; Zhou, K.; Liu, X.; Hodges, M.; Liu, J.; Guan, J.; Phelps, A.; Castro-Piñero, J. Reliability and concurrent validity of global physical activity questionnaire (GPAQ): a systematic review. *Int. J. Environ. Res. Public Health* **2019**, *16*(21), 4128, doi:10.3390/ijerph16214128.
  35. van Poppel, M.N.M.; Chinapaw, M.J.M.; Mokkink, L.B.; Van Mechelen, W.; Terwee, C.B. Physical activity questionnaires for adults. *Sport. Med.* **2010**, *40*, 565–600, doi:10.2165/11531930-000000000-00000.
  36. Bull, F.C.; Maslin, T.S.; Armstrong, T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. *J. Phys. Act. Heal.* **2009**, *6*(6), 790–804, doi:10.1123/jpah.6.6.790.
  37. Baumeister, S.E.; Ricci, C.; Kohler, S.; Fischer, B.; Töpfer, C.; Finger, J.D.; Leitzmann, M.F. Physical activity surveillance in the European Union: reliability and validity of the European health interview survey-physical activity questionnaire (EHIS-PAQ). *Int. J. Behav. Nutr. Phys. Act.* **2016**, *13*, 61, doi:10.1186/s12966-016-0386-6.
  38. Craig, C.L.; Marshall, A.L.; Sjöström, M.; Bauman, A.E.; Booth, M.L.; Ainsworth, B.E.; Pratt, M.; Ekelund, U.L.F.; Yngve, A.; Sallis, J.F. International physical activity questionnaire: 12-country reliability and validity. *Med. Sci. Sport. Exerc.* **2003**, *35*(8), 1381–1395, doi:10.1249/01.MSS.0000078924.61453.FB.
  39. Wolin, K.Y.; Heil, D.P.; Askew, S.; Matthews, C.E.; Bennett, G.G. Validation of the international physical activity questionnaire-short among blacks. *J. Phys. Act. Heal.* **2008**, *5*(5), 746–760, doi:10.1123/jpah.5.5.746.
  40. Mannocci, A.; Di Thiene, D.; Del Cimmuto, A.; Masala, D.; Boccia, A.; De Vito, E.; La Torre, G. International Physical Activity Questionnaire: validation and assessment in an Italian sample. *Ital. J. Public Health* **2010**, *7*(4), 369–376, doi:10.2427/5694.
  41. Hoos, M.B.; Plasqui, G.; Gerver, W.-J.M.; Westerterp, K.R. Physical activity level measured by doubly labeled water and accelerometry in children. *Eur. J. Appl. Physiol.* **2003**, *89*, 624–626, doi:10.1007/s00421-003-0891-6.
  42. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; Group, P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* **2009**, *6*(7), e1000097, doi:10.1371/journal.pmed.1000097.
  43. Moher, D.; Shamseer, L.; Clarke, M.; Ghersi, D.; Liberati, A.; Petticrew, M.; Shekelle, P.; Stewart, L.A. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **2015**, *4*, 1, doi:10.1186/2046-4053-4-1.
  44. Patterson, P. Reliability, validity, and methodological response to the assessment of physical activity via self-report. *Res. Q. Exerc. Sport* **2000**, *71*(Sup 2), 15–20, doi:10.1080/02701367.2000.11082781.
  45. Sneek, S.; Viholainen, H.; Syväoja, H.; Kankaapää, A.; Hakonen, H.; Poikkeus, A.-M.; Tammelin, T. Effects of school-based physical activity on mathematics performance in children: a systematic review. *Int. J. Behav. Nutr. Phys. Act.* **2019**, *16*, 109, doi:10.1186/s12966-019-0866-6.
  46. Sember, V.; Jurak, G.; Kovač, M.; Morrison, S.A.; Starc, G. Children's physical activity, academic performance and cognitive functioning: a systematic review and meta-analysis. *Front. Public Heal.* **2020**, *8*, 307, doi:10.3389/fpubh.2020.00307.
  47. Terwee, C.B.; Mokkink, L.B.; van Poppel, M.N.M.; Chinapaw, M.J.M.; van Mechelen, W.; de Vet, H.C.W. Qualitative attributes and measurement properties of physical activity questionnaires. *Sport. Med.* **2010**, *40*, 525–537, doi:10.2165/11531370-000000000-00000.
  48. Egger, M.; Smith, G.D.; Schneider, M.; Minder, C. Bias in meta-analysis detected by a simple, graphical test. *Bmj* **1997**, *315*, 629–634, doi:10.1136/bmj.315.7109.629.
  49. Hunter, J.E.; Schmidt, F.L.; Le, H. Implications of direct and indirect range restriction for meta-analysis methods and findings. *J. Appl. Psychol.* **2006**, *91*(3), 594–612, doi:10.1037/0021-9010.91.3.594.
  50. Teugels, J.L.; Vet, H. Observer reliability and agreement. In Wiley StatsRef: Statistics Reference Online (eds N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri and J.L. Teugels). *Wiley StatsRef Stat. Ref. Online.* **2014**, doi:10.1002/9781118445112.stat04910.
  51. Plasqui, G.; Westerterp, K.R. Physical Activity Assessment With Accelerometers: An Evaluation Against Doubly Labeled Water. *Obesity* **2007**, *15*, 2371–2379, doi:10.1038/oby.2007.281.
  52. Laeremans, M.; Dons, E.; Avila-Palencia, I.; Carrasco-Turigas, G.; Orjuela, J.P.; Anaya, E.; Brand, C.; Cole-Hunter, T.; de Nazelle, A.; Götschi, T. Physical activity and sedentary behaviour in daily life: A comparative analysis of the Global Physical Activity Questionnaire (GPAQ) and the SenseWear armband. *PLoS One* **2017**, *12*(5), e0177765, doi:10.1371/journal.pone.0177765.
  53. Rütten, A.; Vuillemin, A.; Ooijendijk, W.T.M.; Schena, F.; Sjöström, M.; Stahl, T.; Auweele, Y. Vanden; Welshman, J.; Ziemainz, H. Physical activity monitoring in europe. The European Physical Activity Surveillance System (EUPASS) approach and indicator testing. *Public Health Nutr.* **2003**, *6*(4), 377–384, doi:10.1079/PHN2002449.

54. Bloemen, M.A.T.; Backx, F.J.G.; Takken, T.; Wittink, H.; Benner, J.; Mollema, J.; De Groot, J.F. Factors associated with physical activity in children and adolescents with a physical disability: a systematic review. *Dev. Med. Child Neurol.* **2015**, *57*, 137–148, doi:10.1111/dmcn.12624.
55. Fernández Cabrera, T.; Medina Anzano, S.; Herrera Sánchez, I.M.; Rueda Méndez, S.; Fernández Del Olmo, A. Construcción y validación de una escala de autoeficacia para la actividad física. *Rev. Esp. Salud Pública* **2011**, *85*(4), 405–417, doi:10.1590/S1135-57272011000400009.
56. Milton, K.; Bull, F.C.; Bauman, A. Reliability and validity testing of a single-item physical activity measure. *Br. J. Sports Med.* **2011**, *45*, 203–208, doi:10.1136/bjsm.2009.068395.
57. Murphy, J.J.; Murphy, M.H.; MacDonncha, C.; Murphy, N.; Nevill, A.M.; Woods, C.B. Validity and reliability of three self-report instruments for assessing attainment of physical activity guidelines in university students. *Meas. Phys. Educ. Exerc. Sci.* **2017**, *21*(3), 134–141, doi: 10.1080/1091367X.2017.1297711.
58. Vinas, B.R.; Barba, L.R.; Ngo, J.; Majem, L.S. Validación en población catalana del cuestionario internacional de actividad física. *Gac. Sanit.* **2013**, *27*(3), 254–257, doi:10.1016/j.gaceta.2012.05.013.
59. Rütten, A.; Ziemainz, H.; Schena, F.; Stahl, T.; Stiggelbout, M.; Auweele, Y. Vanden; Vuillemin, A.; Welshman, J. Using different physical activity measurements in eight European countries. Results of the European Physical Activity Surveillance System (EUPASS) time series survey. *Public Health Nutr.* **2003**, *6*(4), 371–376, doi:10.1079/PHN2002450.
60. Scholes, S.; Bridges, S.; Fat, L.N.; Mindell, J.S. Comparison of the physical activity and sedentary behaviour assessment questionnaire and the short-form international physical activity questionnaire: an analysis of health survey for England data. *PLoS One* **2016**, *11*(3), e0151647, doi:10.1371/journal.pone.0151647.
61. Resaland, G.K.; Aadland, E.; Moe, V.F.; Aadland, K.N.; Skrede, T.; Stavnsbo, M.; Suominen, L.; Steene-Johannessen, J.; Glosvik, Ø.; Andersen, J.R. Effects of physical activity on schoolchildren's academic performance: the Active Smarter Kids (ASK) cluster-randomized controlled trial. *Prev. Med. (Baltim.)* **2016**, *91*, 322–328, doi:10.1016/j.ypmed.2016.09.005.
62. Rivière, F.; Widad, F.Z.; Speyer, E.; Erpelding, M.-L.; Escalon, H.; Vuillemin, A. Reliability and validity of the French version of the global physical activity questionnaire. *J. Sport Heal. Sci.* **2018**, *7*(3), 339–345, doi:10.1016/j.jshs.2016.08.004.
63. Kastelic, K.; Šarabon, N. Comparison of Self-Reported Sedentary Time on Weekdays with an Objective Measure (activPAL). *Meas. Phys. Educ. Exerc. Sci.* **2019**, *23*(3), 227–236, doi:10.1080/1091367X.2019.1603153.
64. Taylor, N.J.; Crouter, S.E.; Lawton, R.J.; Conner, M.T.; Prestwich, A. Development and validation of the Online Self-reported Walking and Exercise Questionnaire (OSWEQ). *J. Phys. Act. Heal.* **2013**, *10*(8), 1091–1101, doi:10.1123/jpah.10.8.1091.
- 65.
66. Cleland, C.L.; Hunter, R.F.; Kee, F.; Cupples, M.E.; Sallis, J.F.; Tully, M.A. Validity of the global physical activity questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. *BMC Public Health* **2014**, *14*, 1255, doi:10.1186/1471-2458-14-1255.
67. Ekelund, U.; Sepp, H.; Brage, S.; Becker, W.; Jakes, R.; Hennings, M.; Wareham, N.J. Criterion-related validity of the last 7-day, short form of the International Physical Activity Questionnaire in Swedish adults. *Public Health Nutr.* **2006**, *9*(2), 258–265, doi:10.1079/PHN2005840.
68. Finger, J.D.; Gisle, L.; Mimilidis, H.; Santos-Hoeverer, C.; Kruusmaa, E.K.; Matsi, A.; Oja, L.; Balarajan, M.; Gray, M.; Kratz, A.L. How well do physical activity questions perform? A European cognitive testing study. *Arch. Public Heal.* **2015**, *73*, 57, doi:10.1186/s13690-015-0109-5.
69. Kalvenas, A.; Burlacu, I.; Abu-Omar, K. Reliability and validity of the International Physical Activity Questionnaire in Lithuania. *Balt. J. Heal. Phys. Act.* **2016**, *8*(2), 29–41, doi: 10.29359/BJHPA.08.2.03.
70. Kleinauskienė, L. Tarptautinio fizinio aktyvumo klausimyno trumposios lietuviškos versijos (IPAQ-LT) patikimumo ir pagrįstumo nustatymas. (Masters dissertation, Lithuanian Academy of Physical Education). Lithuanian Sports University. 2012.
71. De La Cámara, M.A.; Higuera-Fresnillo, S.; Cabanas-Sánchez, V.; Sadarangani, K.P.; Martinez-Gomez, D.; Veiga, Ó.L. Criterion validity of the sedentary behavior question from the global physical activity questionnaire in older adults. *J. Phys. Act. Heal.* **2020**, *17*, 2–12, doi:10.1123/jpah.2019-0145.
72. Novak, B.; Holler, P.; Jaunig, J.; Ruf, W.; van Poppel, M.N.M.; Sattler, M.C. Do we have to reduce the recall period? Validity of a daily physical activity questionnaire (PAQ24) in young active adults. *BMC Public Health* **2020**, *20*, 72, doi:10.1186/s12889-020-8165-3.
73. Rudolf, K.; Lammer, F.; Stassen, G.; Froböse, I.; Schaller, A. Show cards of the Global Physical Activity Questionnaire (GPAQ)—do they impact validity? A crossover study. *BMC Public Health* **2020**, *20*, 223, doi:10.1186/s12889-020-8312-x.
74. Lameck, W.U. Sampling design, validity and reliability in general social survey. *Int. J. Acad. Res. Bus. Soc. Sci.* **2013**, *3*(7), 212–218, doi:10.6007/IJARBS/v3-i7/27.

75. Sember, V.; Jurak, G.; Kovač, M.; Đurić, S.; Starc, G. Decline of physical activity in early adolescence: A 3-year cohort study. *PLoS One* **2020**, *15*(3), e0229305, doi:10.1371/journal.pone.0229305.
76. Meyers, R.M.; Bryan, J.G.; McFarland, J.M.; Weir, B.A.; Sizemore, A.E.; Xu, H.; Dharia, N. V; Montgomery, P.G.; Cowley, G.S.; Pantel, S. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.* **2017**, *49*, 1779–1784, doi:10.1038/ng.3984.
79. Sallis, J.F.; Saelens, B.E. Assessment of physical activity by self-report: status, limitations, and future directions. *Res. Q. Exerc. Sport* **2000**, *71*(Sup 2), 1–14, doi:10.1080/02701367.2000.11082780.
80. Hansen, B.H.; Børtnes, I.; Hildebrand, M.; Holme, I.; Kolle, E.; Anderssen, S.A. Validity of the ActiGraph GT1M during walking and cycling. *J. Sports Sci.* **2014**, *32*(6), 510–516, doi:10.1080/02640414.2013.844347.
81. Abarca-Gómez, L.; Abdeen, Z.A.; Hamid, Z.A.; Abu-Rmeileh, N.M.; Acosta-Cazares, B.; Acuin, C.; Adams, R.J.; Aekplakorn, W.; Afsana, K.; Aguilar-Salinas, C.A. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet* **2017**, *390*, 2627–2642, doi:10.1016/S0140-6736(17)32129-3.
82. Voss, M.W.; Weng, T.B.; Burzynska, A.Z.; Wong, C.N.; Cooke, G.E.; Clark, R.; Fanning, J.; Awick, E.; Gothe, N.P.; Olson, E.A. Fitness, but not physical activity, is related to functional integrity of brain networks associated with aging. *Neuroimage* **2016**, *131*, 113–125, doi:10.1016/j.neuroimage.2015.10.044.