Validity, reliability, and readability of single-item and short physical activity questionnaires for use in surveillance: a systematic review

Authors:

Antonina Tcymbal[1*], Sven Messing[1], Rachel Mait[1], Roberto Galindo Perez[1], Taiyeba Akter[1], Ivo Rakovac[2], Peter Gelius[1], Karim Abu-Omar[1]

1 Department of Sport Science and Sport, Friedrich-Alexander-Universität Erlangen-Nürnberg, Gebbertstr. 123b, 91058 Erlangen, Germany

2 World Health Organization Regional Office for Europe, Marmorvej 51, DK-2100 Copenhagen, Denmark

* Corresponding author:

Dr. Antonina Tcymbal

antonina.tcymbal@fau.de

Abstract

Background: Accurate and fast measurement of physical activity is important for surveillance. Even though many physical activity questionnaires (PAQ) are currently used in research, it is unclear which of them is the most reliable, valid, and easy to use. This systematic review aimed to identify existing brief PAQs, describe and compare their measurement properties, and assess their level of readability.

Methods: We performed a systematic review based on the PRISMA statement. Literature searches were conducted in six scientific databases in March 2022. Articles were included if they evaluated validity and/or reliability of brief (i.e., with a maximum of three questions) physical activity or exercise questionnaires intended for healthy adults. Due to the heterogeneity of studies, data were summarized narratively. The level of readability was calculated according to the Flesch-Kincaid formula.

Results: In total, 34 articles published in English or Spanish were included, evaluating 31 distinct brief PAQs. The studies indicated moderate to good levels of reliability for the PAQs. However, the majority of results showed weak validity when validated against objective measurements and demonstrated weak to moderate validity when validated against other PAQs. Most of the assessed PAQs met the criterion of being "short," allowing respondents to complete them in less than one minute either by themselves or with an interviewer. However, only 17 questionnaires had a readability level that indicates that the PAQ is easy to understand for the majority of the population.

Conclusions: This review identified a variety of brief PAQs, but most of them were evaluated in only a single study. The methods used to assess measurement properties varied widely across studies, limiting the comparability between different

PAQs and making it challenging to identify a single tool as the most suitable. Furthermore, PAQs employed different concepts for measuring physical activity, necessitating consideration of measurement properties and assessment goals when selecting a specific tool. None of the evaluated brief PAQs allowed for the measurement of whether a person fulfills the main WHO physical activity recommendations. Future development or adaptation of PAQs should prioritize readability as an important factor to enhance their usability.

Keywords: physical activity, questionnaire, measurement properties, validity, reliability, readability, surveillance

Background

It has been demonstrated that regular physical activity (PA) can help improve physical and mental functions as well as reverse some effects of chronic diseases [1]. Regularly engaging in 150 minutes of PA per week [2], alongside following a healthy diet and abstaining from smoking and alcohol consumption, is seen as being key to prevent non-communicable diseases.

However, accurate measurement of PA levels is important to determine the amount of activity needed to improve health and identify links with other health outcomes and behaviors [3]. To this day, self-report questionnaires are the most common measures to collect PA data, often as part of surveillance systems such as the Behavioral Risk Factor Surveillance System [4], the WHO STEPwise approach to noncommunicable disease risk factor surveillance [5], and the European Health Interview Survey [6]. While such self-report questionnaires have seen widespread use because of their

efficiency, they have also been shown to have limitations related to bias and data accuracy [7].

Many different self-report questionnaires have been developed over the years with the Global Physical Activity Questionnaire (GPAQ), the International Physical Activity Questionnaire (IPAQ; also available as a short form, IPAQ-SF), and the European Health Interview Survey Physical Activity Questionnaire (EHIS-PAQ) being the most widely utilized in global PA surveillance [8]. All three questionnaires assess PA across different domains, asking respondents to report their PA in a typical week (GPAQ, EHIS-PAQ) or during the last 7 days (IPAQ-SF). All three are comparatively complex and range from seven items (IPAQ-SF) to 16 (GPAQ). Nevertheless, the measurement properties of these questionnaires are modest. For GPAQ [9], IPAQ-SF [10, 11], and the EHIS-PAQ [12], different studies have demonstrated reasonable reliability but comparatively low validity. In a recent review [8] of IPAQ-SF, GPAQ, and EHIS-PAQ, the questionnaires showed low-to-moderate validity against objective measures of PA such as accelerometers, and moderate-to-high validity against subjectively measured PA such as other questionnaires.

It is well known that questionnaires with many items can increase the response burden on respondents [13, 14], which has resulted in the development of several short self-report instruments. In contrast to the detailed questionnaires mentioned above, the purpose of short PA questionnaires is to simplify and speed up the procedure for assessing PA levels.

In addition, language-related difficulties are currently an important topic in public health. Patient education materials can increase patient compliance, but only if they are written in a language that is easy for the patient to understand [15]. Regarding PA questionnaires, Altschuler et al. [16] found significant gaps between respondents'

interpretations of some PA questions and researchers' original assumptions about what those questions were intended to measure. One of the characteristics of language difficulty is the level of readability, which indicates how easily readers can understand the text. Research of texts used in healthcare consistently shows that materials intended for patients often require a high level of education and are too complicated for the average person [15, 17]. In relation to physical activity questionnaires (PAQs), readability can influence the amount of time a person needs to understand the question and, if the text is too complicated, may potentially decrease the response rate and correctness of the answer.

A number of existing reviews have investigated the measurement properties of PA questionnaires. Van Poppel et al. [18] reviewed the validity and reliability methodology of 85 versions of PAQs with no clear consensus regarding the best questionnaire for PA measurement. Helmerhost et al. [19] studied reliability and objective criterion-related validity of 34 newly-developed and 96 existing PAQs. However, to our knowledge, a dedicated review of short and brief PA questionnaires measurement properties has not yet been performed. Therefore, the aim of this systematic review is to identify existing brief PAQs and to comparatively describe their measurement properties and level of readability.

Methods

This review follows the guidelines of the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement [20] and the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guideline for systematic reviews of patient-reported outcome measures [21].

Information sources and search strategy

A systematic search was performed in six databases (PubMed, Web of Science, Scopus, CINAHL, SportDiscus, PsycInfo) in March 2022. Additionally, reference lists from previous reviews of PA questionnaires and other relevant publications were screened to identify additional studies. A comprehensive search strategy was developed with a combination of keywords in the categories of measured construct (physical activity) and type of instruments (brief/short questionnaire).

The resulting search string was as follows:

("physical activit*" OR "physical inactivit*")

AND

(questionnaire OR measure OR evaluat* OR assess* OR surveillance OR monitor* OR screening)

AND

("single item" OR single-item OR "single question" OR "one item" OR one-item OR "one question" OR "brief questionnaire " OR "short questionnaire " OR "brief assessment" OR "short assessment" OR "two-item*" OR "two-item*" OR "two questions" OR "brief physical activity assessment" OR "Single response" OR "Single-response")

No restrictions were made regarding language or publication date.

Eligibility criteria

Articles were included in the review if they fulfilled the following eligibility criteria:

1. The article described a self-report PA or exercise questionnaire intended for healthy adults.
2. The evaluated questionnaire was brief and included a maximum of three questions.
3. The article investigated one or more measurement properties of the questionnaire.
4. The article was published in a peer-reviewed journal.

Articles were excluded based on the following exclusion criteria:

1. The article focused on a questionnaire that measured only physical inactivity, screen time, or sedentary behavior.
2. The evaluated questionnaire was intended solely for children and adolescents or people with specific conditions.
3. The article did not investigate the measurement properties of the questionnaire.

Study selection

Two reviewers independently screened and selected the relevant articles. First, all articles were screened based on titles and abstracts. If the title and/or abstract indicated that the study fulfilled the inclusion criteria, both reviewers screened the full text for eligibility. When necessary, supplementary files were also reviewed for additional information. Disagreements between the reviewers were discussed within the research team until a consensus was reached.

Records were managed using the Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia; www.covidence.org) and EndNote X9 (Clarivate Analytics, Philadelphia, PA, USA).

Data extraction

Data of included studies were extracted and summarized by one reviewer and verified by a second reviewer to reduce bias and error. Discrepancies were discussed between reviewers to achieve consensus. Extracted information included: publication details (first author, year of publication, country), sample characteristics (number of participants, age category, special health conditions), the measurement tool(s) explored, who assessed PA levels of participants, assessed measurement properties, other measurement tool(s) used as a comparison, reliability test-retest interval, and the results of the study.

Risk of bias assessment

The methodological quality of the individual studies included in this review was assessed with 18 questions based on the Appraisal tool for Cross-Sectional Studies (AXIS) [22]. In order to specifically assess the quality of the tools' validity and reliability measurement properties, some questions were modified based on the COSMIN risk of bias checklist [21] and on suggestions from previous systematic reviews examining PA assessment measures [23]. This quality assessment evaluated articles based on study design, sample size, participant selection process, appropriate blinding, examiner experience, method of measurement, adequate data reporting, internal consistency, and six other categories (see S1 file). Risk of bias assessment was conducted independently by two reviewers, with any discrepancies

resolved through discussion in the research team. Studies were scored 1 if they satisfied the quality element, and 0 if they did not. The summary score (range: 0–18) indicates the risk of bias, with a higher score indicating higher quality and therefore a lower risk of bias.

Data synthesis and analysis of measurement properties

The primary objective of this systematic review was to identify existing short questionnaires suitable for assessing PA levels in surveillance and primary care settings, and to compare their measurement properties. To accomplish this goal, relevant information from the included studies was summarized separately for each questionnaire. In order to evaluate the reliability and validity of the identified questionnaires, a range of tests were employed in the included studies. Some studies reported results for a total questionnaire summary score, while others assessed reliability and validity for specific aspects, intensities, or domains of the questionnaire. Additionally, certain studies examined these measurement properties within subgroups of the test population. Due to the heterogeneity in the methods used and the lack of standardized reporting across studies, a quantitative meta-analysis was not feasible. Consequently, the information from the included studies was summarized narratively, highlighting the key findings for each questionnaire. This narrative synthesis allows for a comprehensive overview of the reliability and validity findings, highlighting strengths and limitations of each questionnaire.

Analysis of length and readability of questionnaires

To determine how quickly questionnaires could be answered and how easy it is to understand the questions, identified PAQs were analyzed for word count and readability level. The expected time the tool would take for self-administration (silent reading speed) and interviewer administration (spoken word speed) was calculated based on the respective questionnaire's word count and English reading speeds established by Brysbaert [24]. The level of readability was calculated according to the Flesch-Kincaid formula, which was chosen because it is the most commonly used tool to calculate the readability level of written health information [25]. The formula has two forms: the Flesch Reading-Ease-Score, and the Flesch–Kincaid Grade Level. The Flesch Reading-Ease-Score test produces a score from 0 to 100, and higher scores indicate material that is easier to read; lower scores mark passages that are more difficult to read. The Flesch–Kincaid Grade Level formula matches the text to the grade level achievement (number of years of education) required to understand the text. We applied both formulas for each questionnaire using an online calculator [26].

Results

Study selection process

The search across six databases resulted in 2,422 publications. After removing 1,255 duplicates, 1,167 articles were screened based on title and abstract. 65 studies were found eligible for full text assessment. One recently published article [27] and one article identified while screening the reference lists of relevant publications [28] satisfied all eligibility criteria and were therefore included in the analysis. Subsequently, 33 full text articles were excluded due to the length of the questionnaire (n=14), the publication type (n=14), a lack of self-reporting on PA and

239 exercise (n=2), and a study design that did not consider validity and reliability (n=3).
240 In total, 34 studies were included for data extraction and methodological quality
241 assessment. A summary of the search results is presented in Figure 1.

242

243 Figure 1. PRISMA flowchart of study eligibility (PA=physical activity)

244

245 *Characteristics of included studies*

246 A summary of the characteristics of the 34 included studies is presented in S1 Table.
247 The included studies were conducted in Western European countries (n=14), the
248 USA (n=10), Australia (n=5), Canada (n= 2), New Zealand (n=2), and Japan (n=1).

249 A total of 114,143 adults were assessed using the 31 unique brief PAQs (with sample
250 sizes ranging from 9 to 39,379). In 25 studies, the sample consisted only of healthy
251 adults, and nine studies also included specific populations such as older adults [29]
252 [30] [31], overweight/obese adults [32, 33], as well as patients with rheumatoid
253 arthritis [34], coronary heart disease [35] [36], and chronic obstructive pulmonary
254 disease [37].

255 Thirty-one studies documented who completed the PAQ. In 16 studies, respondents
256 filled out the questionnaire by themselves, and in 15 studies the PA level was
257 collected from an interviewer reading a questionnaire.

258 Questionnaires asked for the amount of PA (n=6), the number of days per week with
259 a sufficient amount of PA (n=4), general exercise participation (n= 3), self-reported
260 activity compared with peers (n=2), or for respondents to choose a categorical
261 descriptor of PA levels ranging from “inactive” to “very active” (n=16).

Seven of the included studies investigated the measurement properties of the Single Item Physical Activity Measure (SIPAM) [27, 30, 38-42], five were related to the Brief Physical Activity Assessment Tool (BPAAT) [37, 43-46], and two to the Stanford Leisure-Time Activity Categorical Item (L-Cat) [32, 33]. The other 28 PAQs were assessed in a single study.

Risk of bias assessment

The average quality score of included studies was 11.75, with a range between 7 and 17 (out of 18). Thirteen studies received 14 or more points, which can be considered a high methodological quality. In most of the studies, the aims and objectives, the process of the measurement properties investigation, the statistical analysis and the results were sufficiently described, and the study design was appropriate. However, the included papers reported poorly on whether the examiners had enough experience with the tool and whether they were blinded to participant characteristics, previous findings, or other observed findings. Only four studies justified their choice of sample size. Additionally, many studies failed to consistently report reasons for dropout and characteristics of non-responders. Quality scores of the individual studies are reported in S1 Table.

Questionnaires' measurement properties

A summary of the reliability, validity, and diagnostic test accuracy data found in the identified studies is presented in S2 Table. Studies commonly use a number of different statistical analyses to define absolute (agreement between the two measurement tools) or relative (the degree to which the two measurement tools rank individuals in the same order) validity and reliability. These types of statistical

analysis include correlations (Pearson's; Spearman's; interclass), regression, kappa statistics, and area under the receiver operating curve (AUC-ROC).

Eleven studies measured the reliability of brief PAQs. All of them used a test-retest procedure to measure the consistency of the PAQs. Statistical methods and test-retest intervals varied widely between studies. Overall, studies showed moderate to good reliability levels of the PAQs.

The validity of PAQs was assessed in 33 studies. As a "gold standard" for validation, 13 studies used other PAQs, 15 studies validated brief PAQs against accelerometers or pedometers, and 11 studies compared results of the brief PAQs to other objective measurements, such as BMI [28, 32, 47, 48], VO_2 max [47, 49], or doubly labeled water [34]. Validity coefficients were in general considerably lower than reliability coefficients. The majority of results showed weak validity of brief PAQs against objective measurements and weak to moderate validity against other PAQs.

Length and complexity of the questionnaires

The texts of all questionnaires and information about their length and readability level are presented in S3 Table. It should be noted that, while some PAQs were created and conducted in other languages, length and readability were calculated for the English versions.

Most of the assessed PAQs (n=26) can be read by in silence by the respondent or aloud by the interviewer in less than one minute. However, some questionnaires presented as "brief" by the authors of the identified studies contain a large amount of text and require more than one minute for reading: the Nordic Physical Activity Questionnaire (NPAQ) short, the Total Activity Measure (TAM), the Self-report Scale

311 to Assess Habitual Physical Activity, the Stanford Leisure-Time Activity Categorical
312 Item, and the Stanford Brief Activity Survey.

313 The calculation of the readability levels with the Flesch-Kincaid formula showed that
314 only 17 out of 31 brief PAQs had readability levels of “easy to read” (n=9) or “plain
315 English,” (n=8) and can be easily understood by the majority of the population (see
316 Figure 2). Other questionnaires have long sentences and/or many complicated words
317 (three syllables and more), making them difficult to read. Seven questionnaires
318 scored as “difficult to read” and require college-level education; it is likely that people
319 with lower education levels will find them hard to understand.

320

321 Figure 2. Readability levels for brief PA questionnaires (English language versions).

322

323 *Comparison of brief PAQs*

324 Data on both validity and reliability were only available for nine of the 31 assessed
325 PAQs. These brief PAQs were chosen for an in-depth comparison. Their summarized
326 measurement properties and linguistic characteristics are presented in Table 1.

327

328 Table 1. Comparison of the brief physical activity questionnaires by reliability, validity,
329 length and readability.

Tool	Reliability	Validity		Word count	Readability
		Against objective measurement	Against other PAQ		
Single Item Physical Activity Measure (SIPAM)	Good (0,63-0,82)	Poor to Good (0,22-0,81)	Poor to Good (0,33-0,81)	63 – original 111 – parents version	Fairly difficult to read Parents version - Very difficult to read

Brief Physical Activity Assessment Tool (BPAAT)	Moderate to good (0,53-0,72)	Poor to moderate (0,18-0,43)	Moderate to good (0,45-0,64)	106	Difficult to read
Three-Question Assessment	Moderate to good (0,56-0,63)	Poor (0,24-0,31)	Moderate (0,39-0,43)	140	Fairly difficult to read
Nordic physical activity questionnaire (NPAQ) - short (open ended) *	Good (0,82)	Poor (0,33)	NM	237	Difficult to read
Total Activity Measure (Version 2)	Good (0,82)	Poor (0,36-0,38)	NM	241	Fairly difficult to read
Japan Collaborative Cohort (JACC) Questionnaire*	Moderate (0.39 - 0.56)	NM	Moderate (0.43-0.60)	90	Plain English
Relative PA Question (Compared to peers)	Moderate (0,56)	Poor to moderate (0,28-0,57)	NM	25	Easy to read
The Seven-level Single-Question Scale *	Good (0,7)	Poor (0,33)	NM	144	Fairly difficult to read
Absolute PA Question	Good (0,7)	Poor (0,1-0,33)	NM	34	Fairly difficult to read
A single-item 5-point rating of usual PA (Usual PA Scale)	Good (0,68-0,88)	NM	Good (0,66)	147	Plain English
The Stanford Leisure-Time Activity Categorical Item (L-CAT)	Good (0,64-0,8)	Poor (0,36-0,38)	NM	331	Plain English

* = the study used not the English version of the questionnaire, PAQ=physical activity questionnaire

The Single Item Physical Activity Measure (SIPAM) uses a single question to assess the number of days per week on which 30 minutes or more of PA are performed (excluding housework and work-related PA). For this PAQ, the largest number of studies on validity and reliability were found. The SIPAM showed good reliability levels in all studies, however, results on validity varied considerably between studies – from poor to good – against both objective measurements and other subjective PAQs. Zwolinsky et al. (2015) also measured SIPAM's ability to identify people that meet or do not meet WHO PA recommendations. The tool showed a low diagnostic capacity compared to the IPAQ. The agreement between the SIPAM and the IPAQ was low for identification of participants who meet WHO PA recommendations

342 (kappa= 0.13, 95% CI 0.12 to 0.14) and moderate for the classification of inactive
343 participants (kappa= 0.45, 95% CI 0.43 to 0.47) [42].

344 The Brief Physical Activity Assessment Tool (BPAAT) consists of two questions, one
345 regarding the frequency and duration of vigorous PA and the other regarding
346 moderate PA and walking performed in an individual's usual week. By combining the
347 results of both questions (scores can range from 0 to 8), the subject can be classified
348 as insufficiently (0–3 score) or sufficiently active (≥4 score). Results of reviewed
349 studies showed that the BPAAT has moderate to good levels of reliability and validity
350 in comparison with other PAQs, however comparison with accelerometers identified
351 only poor to moderate validity.

352 Smith et al. (2005) evaluated the Three-Question Assessment variant of the BPAAT
353 which has separate questions about moderate PA and walking. The results of the
354 study did not find a considerable difference in validity and reliability between the two-
355 and three-question versions. One study also reported that more physicians preferred
356 the two-question version (BPAAT) as it is shorter and therefore easier to use (Smith,
357 2005).

358 The Nordic Physical Activity Questionnaire-short (NPAQ-short) includes one question
359 about moderate to vigorous PA and a second question about vigorous PA. Danquah
360 et al. (2018) compared open-ended and closed-ended questions with the open-
361 ended version achieving better results. The open-ended version showed good
362 reliability but performed similarly to other questionnaires, showing poor validity when
363 compared against objective PA measures. The analyses showed that the
364 questionnaire was one of the longest and was rated "difficult to read". The agreement
365 with accelerometer data in identification of people that meet or do not meet the WHO
366 PA recommendations was low (kappa=0.42).

The Total Activity Measures (TAM) includes three open-ended questions about strenuous, moderate, and mild PA. The revised second version, TAM2, asks about the total time spent at each activity level over a 7-day period. The TAM2 showed good reliability when validated against objectively-measured PA. The word count was high at 241, and readability was rated as “fairly difficult”.

The Japan Collaborative Cohort (JACC) Questionnaire has three items. Two questions focus on leisure-time PA, i.e., time per week engaging in sport or PA (with options ranging from “little” to “at least 5 hours”), and frequency of engagement in sport over the past year (options from “seldom” to “at least twice a week”). The third question asks about daily walking patterns (options from “little” to “more than 1 hour”). The questionnaire showed moderate reliability after one year and moderate validity when compared against a more in-depth interview by a trained researcher. The questions are 90 words in total and were rated as “plain English.”

The Relative PA Question is the shortest of the compared PAQs and has an easy readability level. It allows respondents to compare their level of PA with other people of the same age and to choose from five categories ranging from “much more active” to “much less active.” The questionnaire has a moderate reliability level and showed poor validity against objective measures.

The Absolute PA Question asked participants to choose what best describes their activity level from three options: (1) vigorously active for at least 30 minutes, three times per week; (2) moderately active at least three times per week; or (3) seldom active, preferring sedentary activities. The questionnaire is relatively short, but as most of the other brief PAQs, it showed high reliability and poor validity against objective measures.

The Seven-Level Single-Question Scale (SR-PA L7) requires respondents to assign themselves to a level ranging from “I do not move more than is necessary in my daily routines/chores” and “I participate in competitive sports and maintain my fitness through regular training.” These seven items are supposed to categorize respondents as maintaining low, medium, or high PA. The SR-PA L7 showed poor validity when compared to objectively measured PA but good reliability. The scale is moderately long at 144 words and was rated as “fairly difficult to read.”

A single-item 5-point rating of usual PA (Usual PA Scale) includes descriptions of three PA levels: highly active, moderately active, and inactive. Respondents need to read a description of each category and identify their usual PA level from the list. The questionnaire’s readability level was ranked as “plain English,” and it showed good reliability and validity levels. However, validation was done only against another PAQ which is not well-known and requires further research.

The Stanford Leisure-Time Activity Categorical Item (L-CAT) is a single-item questionnaire that consists of six descriptive PA categories ranging from inactive to very active. Each category consists of one or two statements describing common activity patterns over the past month, differing in frequency, intensity, duration, and types of activity. The categories are described in “plain English,” but this made the questionnaire longer than the other questionnaires. The L-CAT showed high reliability but poor validity against pedometer and accelerometer validation.

Discussion

This review assessed the validity, reliability, and readability of brief PAQs for adults. To our knowledge, it is the first review focused specifically on brief PAQs and also the first to assess their readability. The sheer number of brief PAQs we identified

(n=31) indicates a high research interest in such instruments. However, it also indicates a lack of harmonization when it comes to PA assessment using brief questionnaires.

Overall, most PAQs were reported to have poor or moderate validity. They showed higher validity levels against other self-reported tools than against objective PA measures. Although reliability is an important measurement property, it was assessed only in 11 studies. About half of all PAQs showed a moderate to good level of reliability. These results are in line with other reviews of PAQs [18, 19, 50, 51].

A significant difficulty in conducting this review was that the studies used different methods for validation, varying time-intervals between repeated measurements, and different statistical methods to analyze data. Complete data on validity and reliability were only available for nine PAQs, and only those could be compared in greater detail. All of these factors prevented a more detailed comparison of results across PAQs. In general, the methodological quality of the included studies was modest. The most common flaws were comparably small sample sizes, a lack of sample size justification, and a poor description of the validity and reliability assessment process. Additionally, most studies utilized convenience samples, making it impossible to assess if measurement properties would differ between adults with different levels of socioeconomic status and/or educational attainment.

Also, the included PAQs used different concepts of measuring PA, further complicating a direct comparison between instruments. For example, the SIPAM measures on how many days per week respondents perform 30 minutes or more of PA; the Absolute PAQ requires respondents to choose from several descriptions of different PA levels; and the Relative PA question asks them to compare their level of PA with peers. Some PAQs, such as the TAM, aim to assess the total volume of

441 moderate-to-vigorous PA. Others focus on particular PA domains. And yet others
442 assess only leisure-time PA. This variety can be partly explained by efforts to keep
443 PAQs short at the price of excluding some PA dimensions (type, duration, intensity,
444 volume). However, it could also be interpreted as a lack of common understanding
445 about which dimensions of PA should be assessed with brief PAQs. It should also be
446 taken into account that none of the reviewed brief PAQs allowed for the
447 measurement of whether a person fulfills the WHO physical activity
448 recommendations [52].

449 All of the included studies were conducted in highly developed nations, and 23 of
450 them took place in English-speaking countries. This most certainly biased the results,
451 since terminology related to PA differs between languages, as do levels of literacy
452 [53]. Ultimately, to assess reliability and validity of PAQs, many more studies should
453 be conducted in developing nations to obtain a more accurate assessment of their
454 suitability for international surveillance systems. Certainly, in order to stimulate such
455 research, funding opportunities need to be made available to research teams from
456 such nations.

457 The results of this study point out that many PAQs have low readability levels. This is
458 particularly disturbing when considering that most of these PAQs were tested in
459 developed nations with comparatively high levels of literacy. Potentially, poor
460 readability is related to the modest measurement properties that PAQs commonly
461 come with. Questionnaires that are more readable or might even have been co-
462 developed [54] with the intended population groups are likely to yield better
463 measurement properties. This would also strengthen the case of integrating them
464 into PA surveillance systems. However, at this point in time, there is a dearth of
465 research relating PAQs and their measurement properties to readability. Advancing

knowledge in this field would also benefit longer PA questionnaires (such as IPAQ and GPAQ) that are widely utilized in surveillance and potentially score low on readability as well.

This review comes with certain limitations. The search was limited to scientific databases, and only studies published in peer reviewed scientific journals were included. This can lead to a publication bias, as all other types of publications and gray literature were excluded. The varying measurement methods and conditions complicated the comparison of findings from different studies and limited data analysis to a narrative description of differences between tools. This led to more subjective results and greater difficulty in identifying the tools with the best measurement properties. It should be also taken into account that the Flesch-Kincaid readability formula was created for longer texts and not adapted for short questionnaires. Consequently, the readability levels of PAQs presented here should be interpreted with caution.

Conclusion

This systematic review sheds light on the validity, reliability, and readability of short PAQs. Results indicate that additional research is needed, notably regarding their reliability and validity, their readability, and their applicability to non-English speaking and/or developing countries.

Recent years have seen a shift from self-report questionnaires towards objective measures in PA surveillance. This shift has been partially motivated by the persistently modest measurement properties of PAQs, as well as technological advancements in the quality and affordability of accelerometers and other PA measurement tools. However, objective PA assessment comes with its own set of

limitations [55], and self-report and objective measures have been described as measuring entirely different parameters. Right now, it seems unclear what the future for PA surveillance might hold and if brief PAQs will continue to play a role.

In this regard, the key merit of short PAQs is that they require substantially less time to complete than the established PA surveillance questionnaires (GPAQ, IPAQ, and EHIS-PAQ). This makes them highly appealing for use in surveillance systems. In order for them to be included in existing systems, more studies would need to investigate how they compare to longer PAQs. Potentially, if their validity and reliability are found to be at a comparable level, we might see surveillance systems gradually shift from longer towards briefer PAQs.

List of abbreviations:

BPAAT: Brief Physical Activity Assessment Tool

EHIS-PAQ: European Health Interview Survey Physical Activity Questionnaire

GPAQ: Global Physical Activity Questionnaire

IPAQ: International Physical Activity Questionnaire

JACC: Japan Collaborative Cohort Questionnaire

L-CAT: The Stanford Leisure-Time Activity Categorical Item

NPAQ-short: The Nordic Physical Activity Questionnaire-short

PA: physical activity

PAQ: physical activity questionnaire

SIPAM: Single Item Physical Activity Measure

SR-PA L7: The Seven-level Single-Question Scale for Self-Reported Leisure Time Physical Activity

TAM: Total Activity Measure

517 **Declarations**

518 Ethics approval and consent to participate: Not applicable.

519 Availability of data and materials: Data sharing is not applicable to this article as no
520 datasets were generated or analysed during the current study.

521 Competing interests: The authors declare that they have no competing interests.

522 Funding: This research was conducted as part of projects funded by the German
523 Federal Ministry of Health (ZMI5-2522WHO001, ZMI5-2523WHO001). The ministry
524 was neither involved in writing this manuscript nor in the decision to submit the article
525 for publication. Furthermore, we acknowledge financial support by Deutsche
526 Forschungsgemeinschaft and Friedrich-Alexander-Universität Erlangen-Nürnberg
527 within the funding programme "Open Access Publication Funding".

528 Authors' contributions: AT, KA-O and IR conceptualized the study. AT conducted the
529 literature search. AT, RA, RGP, and TA performed abstract/title screening, full-text
530 selection, data extraction, and quality assessment. AT with support from KA-O, SM
531 and PG prepared a draft of the manuscript. All authors critically reviewed the draft
532 and approved the final manuscript.

533

534 References:

535 1. McPhee JS, French DP, Jackson D, Nazroo J, Pendleton N, Degens H.
536 Physical activity in older age: perspectives for healthy ageing and frailty.
537 Biogerontology. 2016;17(3):567-80. Epub 20160302. doi: 10.1007/s10522-016-9641-
538 0. PubMed PMID: 26936444; PubMed Central PMCID: PMCPMC4889622.
539 2. World Health Organization. Guidelines on physical activity and sedentary
540 behaviour. Geneva: World Health Organization; 2020.

541 3. Welk G. Physical activity assessments for health-related research: Human
542 Kinetics; 2002.
543 4. Centers for Disease Control and Prevention. Behavioral Risk Factor
544 Surveillance System Survey Questionnaire. Atlanta, Georgia: U.S. Department of
545 Health and Human Services, Centers for Disease Control and Prevention; 2021.
546 5. World Health Organization. WHO STEPS Surveillance Manual: The WHO
547 STEPwise approach to noncommunicable disease risk factor surveillance. 2020.
548 6. Finger JD, Tafforeau J, Gisle L, Oja L, Ziese T, Thelen J, et al. Development
549 of the European Health Interview Survey - Physical Activity Questionnaire (EHIS-
550 PAQ) to monitor physical activity in the European Union. Arch Public Health.
551 2015;73:59. Epub 20151202. doi: 10.1186/s13690-015-0110-z. PubMed PMID:
552 26634120; PubMed Central PMCID: PMCPMC4667448.
553 7. Sallis JF, Saelens BE. Assessment of physical activity by self-report: status,
554 limitations, and future directions. Research quarterly for exercise and sport.
555 2000;71:1-14. doi: https://doi.org/10.1080/02701367.2000.11082780.
556 8. Sember V, Meh K, Sorić M, Starc G, Rocha P, Jurak G. Validity and Reliability
557 of International Physical Activity Questionnaires for Adults across EU Countries:
558 Systematic Review and Meta Analysis. International journal of environmental
559 research and public health. 2020;17(19):7161. Epub 20200930. doi:
560 https://doi.org/10.3390/ijerph17197161. PubMed PMID: 33007880; PubMed Central
561 PMCID: PMCPMC7579664.
562 9. Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire
563 (GPAQ): nine country reliability and validity study. J Phys Act Health. 2009;6(6):790-
564 804. doi: 10.1123/jpah.6.6.790. PubMed PMID: 20101923.
565 10. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et
566 al. International physical activity questionnaire: 12-country reliability and validity.

567 *Medicine and science in sports and exercise*. 2003;35(8):1381–95. doi:
568 10.1249/01.MSS.0000078924.61453.FB.
569 11. Lee PH, J. MD, Lam TH, Stewart SM. Validity of the International Physical
570 Activity Questionnaire Short Form (IPAQ-SF): a systematic review. The international
571 journal of behavioral nutrition and physical activity. 2011;8:115. doi:
572 https://doi.org/10.1186/1479-5868-8-115.
573 12. Baumeister SE, Ricci C, Kohler S, Fischer B, Topfer C, Finger JD, et al.
574 Physical activity surveillance in the European Union: reliability and validity of the
575 European Health Interview Survey-Physical Activity Questionnaire (EHIS-PAQ). The
576 international journal of behavioral nutrition and physical activity. 2016;13:61. Epub
577 20160523. doi: 10.1186/s12966-016-0386-6. PubMed PMID: 27215626; PubMed
578 Central PMCID: PMCPMC4877949.
579 13. Ware J, Jr., M. K, Keller SD. A 12-Item Short-Form Health Survey:
580 construction of scales and preliminary tests of reliability and validity. Medical Care.
581 1996;34(3):220-33. doi: https://doi.org/10.1097/00005650-199603000-00003.
582 14. Rolstad S, Adler J, Rydén A. Response burden and questionnaire length: is
583 shorter better? A review and meta-analysis. Value in health : the journal of the
584 International Society for Pharmacoeconomics and Outcomes Research.
585 2011;14(8):1101-8. doi: https://doi.org/10.1016/j.jval.2011.06.003.
586 15. Grabeel KL, Russomanno J, Oelschlegel S, Tester E, Heidel RE.
587 Computerized versus hand-scored health literacy tools: a comparison of Simple
588 Measure of Gobbledygook (SMOG) and Flesch-Kincaid in printed patient education
589 materials. J Med Libr Assoc. 2018;106(1):38-45. Epub 2018/01/18. doi:
590 10.5195/jmla.2018.262. PubMed PMID: 29339932; PubMed Central PMCID:
591 PMCPMC5764592.

592 16. Altschuler A, Picchi T, Nelson M, Rogers JD, Hart J, Sternfeld B. Physical
593 activity questionnaire comprehension: lessons from cognitive interviews. Med Sci
594 Sports Exerc. 2009;41(2):336-43. doi: 10.1249/MSS.0b013e318186b1b1. PubMed
595 PMID: 19127192; PubMed Central PMCID: PMCPMC2694667.
596 17. Wallace LS, Lennon ES. American Academy of Family Physicians patient
597 education materials: Can patients read them? Family Medicine. 2004;36(8):571-4.
598 Epub 2004/09/03. PubMed PMID: 15343418.
599 18. van Poppel MN, Chinapaw MJ, Mokkink LB, van Mechelen W, Terwee CB.
600 Physical activity questionnaires for adults: a systematic review of measurement
601 properties. Sports Medicine (Auckland, N Z). 2010;40(7):565-600. doi:
602 https://doi.org/10.2165/11531930-000000000-00000.
603 19. Helmerhorst HJH, S. B, Warren J, Besson H, Ekelund U. A systematic review
604 of reliability and objective criterion-related validity of physical activity questionnaires.
605 The international journal of behavioral nutrition and physical activity. 2012;9:103. doi:
606 10.1186/1479-5868-9-103.
607 20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et
608 al. The PRISMA 2020 statement: an updated guideline for reporting systematic
609 reviews. BMJ. 2021;372:n71. doi: 10.1136/bmj.n71.
610 21. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et
611 al. COSMIN guideline for systematic reviews of patient-reported outcome measures.
612 Quality of Life Research : An International Journal of Quality of Life Aspects of
613 Treatment, Care, and Rehabilitation. 2018;27(5):1147-57. Epub 2018/02/13. doi:
614 https://doi.org/10.1007/s11136-018-1798-3. PubMed PMID: 29435801; PubMed
615 Central PMCID: PMCPMC5891568.

616 22. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical
617 appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open.
618 2016;6(12):e011458. doi: 10.1136/bmjopen-2016-011458.
619 23. Williams MA, McCarthy CJ, Chorti A, Cooke MW, Gates S. A systematic
620 review of reliability and validity studies of methods for measuring active and passive
621 cervical range of motion. Journal of Manipulative and Physiological Therapeutics.
622 2010;33(2):138-55. Epub 2010/02/23. doi: 10.1016/j.jmpt.2009.12.009. PubMed
623 PMID: 20170780.
624 24. Brysbaert M. How many words do we read per minute? A review and meta-
625 analysis of reading rate. Journal of Memory and Language. 2019;109. doi:
626 https://doi.org/10.1016/j.jml.2019.104047.
627 25. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula
628 differences with written health information materials: application, results, and
629 recommendations. Res Social Adm Pharm. 2013;9(5):503-16. Epub 2012/07/28. doi:
630 10.1016/j.sapharm.2012.05.009. PubMed PMID: 22835706.
631 26. charactercalculator.com. Flesch Reading Ease 2023. Available from:
632 https://charactercalculator.com/flesch-reading-ease/.
633 27. Bauman AE, Richards JA. Understanding of the Single-Item Physical Activity
634 Question for Population Surveillance. Journal of Physical Activity and Health.
635 2022;19(10):681-6. doi: 10.1123/jpah.2022-0369.
636 28. Schechtman KB, Barzilai B, Rost K, B. FJE. Measuring physical activity with a
637 single question. American journal of public health. 1991;81(6):771-3.
638 29. Gill DP, Jones GR, Zou G, Speechley M. Using a single question to assess
639 physical activity in older adults: A reliability and validity study. BMC Medical
640 Research Methodology. 2012;12(20). doi: 10.1186/1471-2288-12-2022373159.

641 30. Macdonald HM, Nettlefold L, Bauman A, Sims-Gould J, McKay HA. Pragmatic
642 Evaluation of Older Adults' Physical Activity in Scale-Up Studies: Is the Single-Item
643 Measure a Reasonable Option? Journal of Aging and Physical Activity.
644 2022;30(1):25-32. doi: 10.1123/japa.2020-0412.
645 31. Portegijs E, Sipilä S, Viljanen A, Rantakokko M, Rantanen T. Validity of a
646 single question to assess habitual physical activity of community-dwelling older
647 people. Scandinavian Journal of Medicine & Science in Sports. 2017;27(11):1423-30.
648 doi: 10.1111/sms.12782.
649 32. Kiernan M, Schoffman DE, Lee K, Brown SD, Fair JM, Perri MG, et al. The
650 stanford leisure-time activity categorical item (L-Cat): A single categorical item
651 sensitive to physical activity changes in overweight/obese women. International
652 Journal of Obesity. 2013;37(12):1597-602. doi: 10.1038/ijo.2013.36.
653 33. Ross KM, Leahey TM, Kiernan M. Validation of the Stanford Leisure-Time
654 Activity Categorical Item (L-Cat) using armband activity monitor data. *Obesity*
655 *science & practice*. 2018;4(3):276-82. doi: 10.1002/osp4.155.
656 34. Johansson G, Westerterp KR. Assessment of the physical activity level with
657 two questions: Validation with doubly labeled water. International Journal of Obesity.
658 2008;32(6):1031-3. doi: 10.1038/ijo.2008.42.
659 35. Blomqvist A, Bäck M, Klompstra L, Strömberg A, Jaarsma T. Utility of single-
660 item questions to assess physical inactivity in patients with chronic heart failure. ESC
661 Heart Fail. 2020;7(4):1467-76. doi: 10.1002/ehf2.12709.
662 36. Orrell A, Doherty P, Miles J, Lewin R. Development and validation of a very
663 brief questionnaire measure of physical activity in adults with coronary heart disease.
664 European Journal of Cardiovascular Prevention and Rehabilitation. 2007;14(5):615-
665 23. doi: 10.1097/HJR.0b013e3280ecfd56.

666 37. Cruz J, Jácome C, Oliveira A, Paixão C, Rebelo P, Flora S, et al. Construct
667 validity of the brief physical activity assessment tool for clinical use in COPD. The
668 Clinical Respiratory Journal. 2021;15(5):530-9. doi: 10.1111/crj.13333.
669 38. Hamilton K, White KM, Cuddihy T. Using a single-item physical activity
670 measure to describe and validate parents' physical activity patterns. Research
671 Quarterly for Exercise and Sport. 2012;83(2):340-5. doi:
672 10.1080/02701367.2012.10599865.
673 39. Milton K, Bull FC, Bauman A. Reliability and validity testing of a single-item
674 physical activity measure. British Journal of Sports Medicine. 2011;45(3):203-8. doi:
675 10.1136/bjsm.2009.068395.
676 40. Milton K, Clemes S, Bull F. Can a single question provide an accurate
677 measure of physical activity? British Journal of Sports Medicine. 2013;47(1):44-8.
678 doi: 10.1136/bjsports-2011-090899.
679 41. O'Halloran P, Kingsley M, Nicholson M, Staley K, Randle E, Wright A, et al.
680 Responsiveness of the single item measure to detect change in physical activity.
681 PLoS ONE. 2020;15(6). doi: 10.1371/journal.pone.0234420.
682 42. Zwolinsky S, McKenna J, Pringle A, Widdop P, Griffiths C. Physical activity
683 assessment for public health: Efficacious use of the single-item measure. Public
684 Health. 2015;129(12):1630-6. doi: 10.1016/j.puhe.2015.07.015.
685 43. Marshall AL, Smith BJ, Bauman AE, Kaur S. Reliability and validity of a brief
686 physical activity assessment for use by family doctors. British Journal of Sports
687 Medicine. 2005;39(5):294-7. doi: 10.1136/bjsm.2004.013771.
688 44. Puig Ribera A, Peña Chimenis Ò, Romaguera Bosch M, Duran Bellido E,
689 Heras Tebar A, Solà Gonfaus M, et al. How to identify physical inactivity in Primary
690 Care: Validation of the Catalan and Spanish versions of 2 short questionnaires.
691 Atencion Primaria. 2012;44(8):485-93. doi: 10.1016/j.aprim.2012.01.005.

692 45. Puig-Ribera A, Martín-Cantera C, Puigdomenech E, Real J, Romaguera M,
693 Magdalena-Belio JF, et al. Screening physical activity in family practice: Validity of
694 the Spanish version of a brief physical activity questionnaire. PLoS ONE. 2015;10(9).
695 doi: 10.1371/journal.pone.0136870.
696 46. Smith BJ, Marshall AL, Huang N. Screening for physical activity in family
697 practice: Evaluation of two brief assessment tools. American journal of preventive
698 medicine. 2005;29(4):256-64. doi: 10.1016/j.amepre.2005.07.005.
699 47. Gionet NJ, Godin G. Self-Reported Exercise Behavior of Employees - A
700 Validity Study. Journal of Occupational and Environmental Medicine.
701 1989;31(12):969-73. doi: 10.1097/00043764-198912000-00007.
702 48. Li S, Carlson E, Holm K. Validation of a single-item measure of usual physical
703 activity. Perceptual and Motor Skills. 2000;91(2):593-602. doi:
704 10.2466/pms.2000.91.2.593.
705 49. Graff-Iversen S, Anderssen SA, Holme IM, Jenum AK, Raastad T. Two short
706 questionnaires on leisure-time physical activity compared with serum lipids,
707 anthropometric measurements and aerobic power in a suburban population from
708 Oslo, Norway. European journal of epidemiology. 2008;23(3):167-74. doi:
709 10.1007/s10654-007-9214-2.
710 50. Forsén L, Loland NW, Vuillemin A, Chinapaw MJ, van Poppel MN, Mokkink
711 LB, et al. Self-administered physical activity questionnaires for the elderly: a
712 systematic review of measurement properties. Sports Medicine. 2010;40(7):601-23.
713 doi: 10.2165/11531350-000000000-00000. PubMed PMID: 20545382.
714 51. Smith TO, McKenna MC, Salter C, Hardeman W, Richardson K, Hillsdon M, et
715 al. A systematic review of the physical activity assessment tools used in primary
716 care. Family Practice. 2017;34(4):384-91. doi:
717 https://doi.org/10.1093/fampra/cmx011.

718 52. Organization WH. Guidelines on physical activity and sedentary behaviour.
719 Geneva: World Health Organization; 2020.
720 53. Levin-Zamir D, Leung AYM, Dodson S, Rowlands G. Health Literacy in
721 Selected Populations: Individuals, Families, and Communities from the International
722 and Cultural Perspective. Studies in Health Technology and Informatics.
723 2017;240:392-414. PubMed PMID: 28972530.
724 54. Marsilio M, Fusco F, Gheduzzi E, Guglielmetti C. Co-Production Performance
725 Evaluation in Healthcare. A Systematic Review of Methods, Tools and Metrics.
726 International Journal of Environmental Research and Public Health. 2021;18(7).
727 Epub 20210324. doi: 10.3390/ijerph18073336. PubMed PMID: 33804862; PubMed
728 Central PMCID: PMCPMC8037812.
729 55. Hankinson A. Association of activity and chronic disease risk factors: utility
730 and limitations of objectively measured physical activity data. Journal of the American
731 Dietetic Association. 2008;108(6):945-7. doi: 10.1016/j.jada.2008.03.018. PubMed
732 PMID: 18502224.
733 56. Ball TJ, Joy EA, Goh TL, Hannon JC, Gren LH, Shaw JM. Validity of two brief
734 primary care physical activity questionnaires with accelerometry in clinic staff.
735 Primary Health Care Research & Development. 2015;16(1):100-8. doi: DOI:
736 10.1017/S1463423613000479.
737 57. Cruz J, Jácome C, Oliveira A, Paixão C, Rebelo P, Flora S, et al. Construct
738 validity of the brief physical activity assessment tool for clinical use in COPD. Clin
739 Respir J. 2021;15(5):530-9. doi: 10.1111/crj.13333.
740 58. Danquah IH, Petersen CB, Skov SS, Tolstrup JS. Validation of the NPAQ-
741 short - A brief questionnaire to monitor physical activity and compliance with the
742 WHO recommendations. BMC Public Health. 2018;18(1). doi: 10.1186/s12889-018-
743 5538-y.

59. Hyvärinen M, Sipilä S, Kulmala J, Hakonen H, Tammelin TH, Kujala UM, et al. Validity and reliability of a single question for leisure-time physical activity assessment in middle-aged women. Journal of Aging and Physical Activity. 2020;28(2):231-41. doi: 10.1123/JAPA.2019-0093.

60. Iwai N, Hisamichi S, Hayakawa N, Inaba Y, Nagaoka T, Sugimori H, et al. Validity and reliability of single-item questions about physical activity. Journal of Epidemiology. 2001;11(5):211-8. doi: 10.2188/jea.11.211.

61. Jackson AW, Morrow JR, Jr., Bowles HR, Fitzgerald SJ, Blair SN. Construct validity evidence for single-response items to estimate physical activity levels in large sample studies. Research Quarterly for Exercise and Sport. 2007;78(2):24-31. doi: 10.1080/02701367.2007.10599400.

62. Moreno-Llamas A, García-Mayor J, De la Cruz-Sánchez E. Concurrent and convergent validity of a single, brief question for physical activity assessment. International Journal of Environmental Research and Public Health. 2020;17(6):1989. doi: 10.3390/ijerph17061989.

63. Rose SB, Elley CR, Lawton BA, Dowell AC. A single question reliably identifies physically inactive women in primary care. The New Zealand Medical Journal. 2008;121(1268):U2897.

64. Taylor-Piliae RE, Norton LC, Haskell WL, Mahbouda MH, Fair JM, Iribarren C, et al. Validation of a new brief physical activity survey among men and women aged 60-69 years. American Journal of Epidemiology. 2006;164(6):598-606. doi: 10.1093/aje/kwj248.

65. Webster S, Khan A, Nitz JC. A brief questionnaire is able to measure population physical activity levels accurately: A comparative validation study. Journal of Clinical Gerontology and Geriatrics. 2011;2(3):83-7. doi: 10.1016/j.jcgg.2011.06.003.

66. Weiss TW, Slater CH, Green LW, Kennedy VC, Albright DL, Wun CC. The validity of single-item, self-assessment questions as measures of adult physical activity. Journal of Clinical Epidemiology. 1990;43(11):1123-9. doi: 10.1016/0895-4356(90)90013-F.

67. Yore MM, Bowles HR, Ainsworth BE, Macera CA, Kohl HW, III. Single versus multiple item questions on occupational physical activity. Journal of Physical Activity and Health. 2006;3(1):102-11. doi: 10.1123/jpah.3.1.102.

68. Bauman AE, Richards JA. Understanding of the Single-Item Physical Activity Question for Population Surveillance. Journal of Physical Activity and Health. 2022. doi: 10.1123/jpah.2022-0369.

Supporting information

S1 Table. Studies overview.

S2 Table. Summary of measurement properties.

S3 Table. Length and readability levels of short physical activity questionnaires.

S1 File. Blank for studies quality assessment.

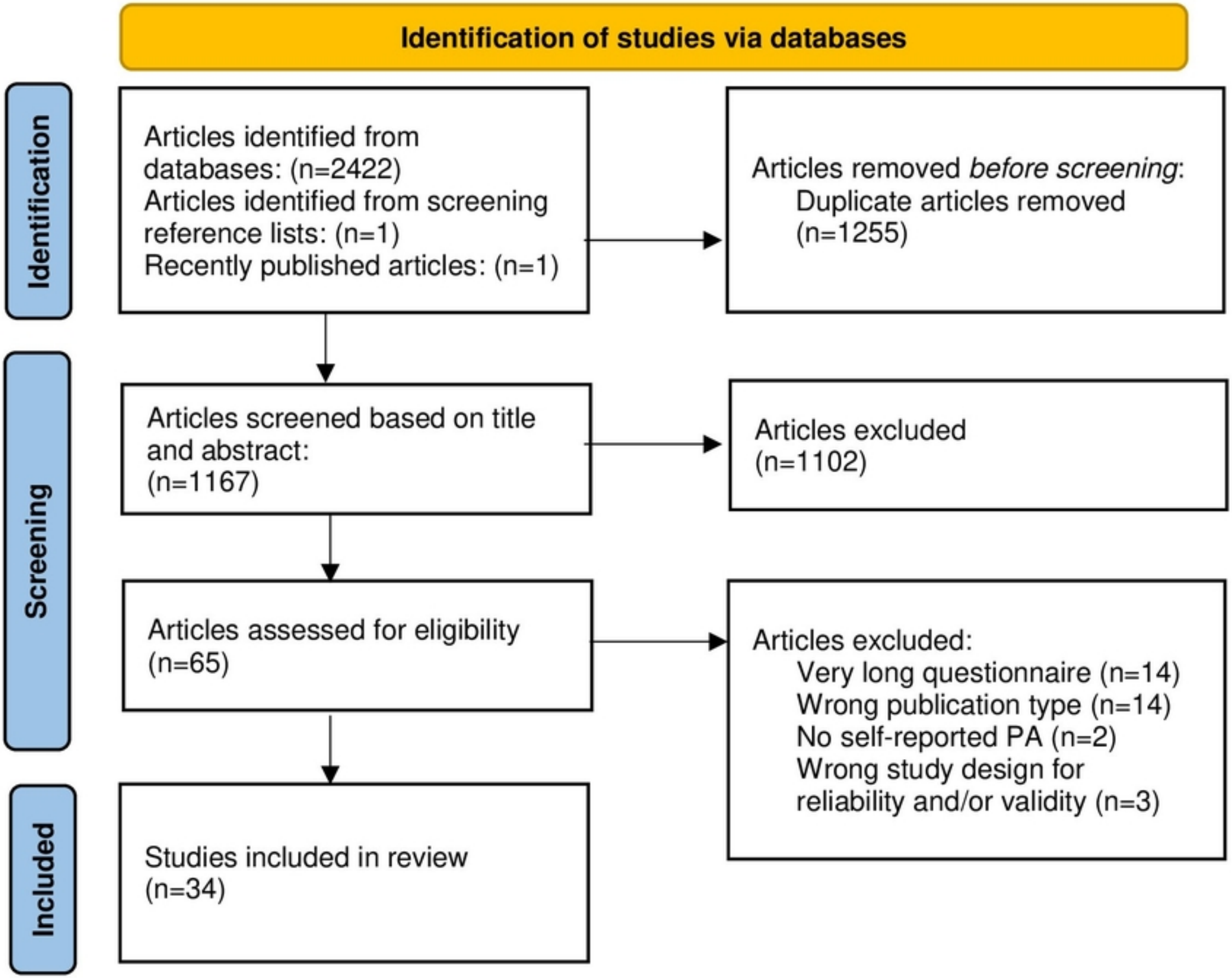
Identification of studies via databases
Identification
Articles identified from databases: (n=2422)
Articles identified from screening reference lists: (n=1)
Recently published articles: (n=1)
Articles removed *before screening*:
Duplicate articles removed (n=1255)
Screening
Articles screened based on title and abstract: (n=1167)
Articles excluded (n=1102)
Articles assessed for eligibility (n=65)
Articles excluded:
Very long questionnaire (n=14)
Wrong publication type (n=14)
No self-reported PA (n=2)
Wrong study design for reliability and/or validity (n=3)
Included
Studies included in review (n=34)

Figure 1

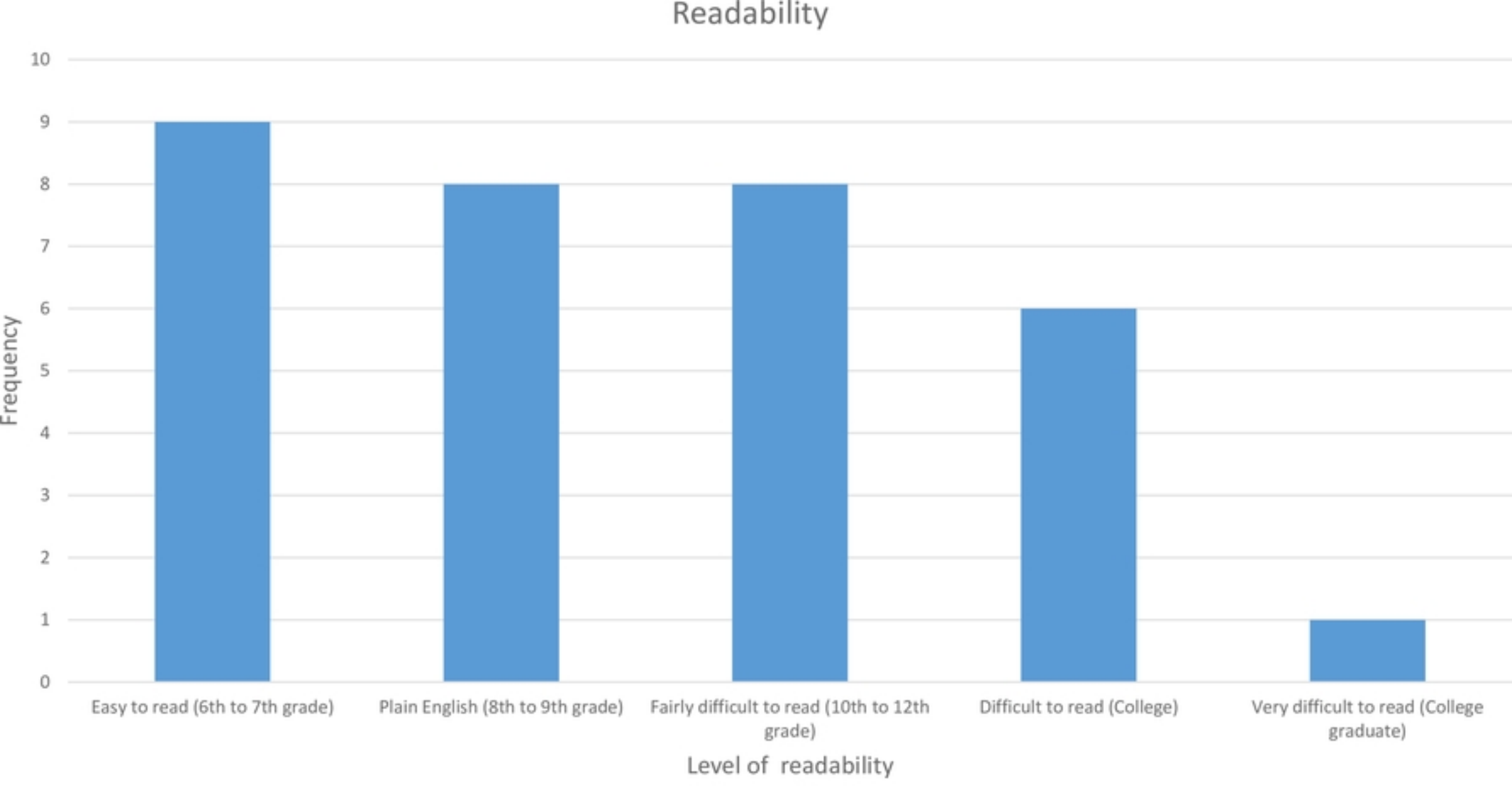
Readability
10
9
8
7
6
5
4
3
2
1
0
Frequency
Easy to read (6th to 7th grade)
Plain English (8th to 9th grade)
Fairly difficult to read (10th to 12th grade)
Difficult to read (College)
Very difficult to read (College graduate)
Level of readability

Figure 2