Author's response to reviews

Title: How well do physical activity questions perform? A European cognitive testing study

Authors:

Jonas D Finger (FingerJ@rki.de)
Lydia Gisle (Lydia.Gisle@wiv-isp.be)
Helene Mimilidis (helene.mimilidis@gmail.com)
Claudia Santos-Hoevener (Santos-HoevenerC@rki.de)
Elin K Kruusmaa (elin-kulliki.kruusmaa@tai.ee)
Ardo Matsi (ardo.matsi@tai.ee)
Leila Oja (leila.oja@tai.ee)
Meera Balarajan (meera.balarajan@cantab.net)
Michele Gray (michandadam@yahoo.co.uk)
Anna Lena Kratz (KratzA@rki.de)
Cornelia Lange (LangeC@rki.de)

Version: 2
Date: 31 October 2015

Author's response to reviews: see over
Revision of the manuscript: “How well do physical activity questions perform? A European cognitive testing study”

Dear Editor and Reviewers,

We thank you for the detailed and useful comments on our manuscript. In the following we respond to your remarks point by point and describe the changes we have made.

Reviewer 1:
Aurelie Van Hoye
Reviewer's report:
Major compulsory revisions
Introduction
1) P. 2, Ch. 2: A sentence could be added on the different forms of questionnaire and journal to capture PA level and their specificity. Moreover, precision about the choices of the two selected measures are needed, why did you evaluated specifically these measures and not any other?

We agree, we have added information about different forms of questionnaires (generic versus domain-specific questionnaires) in the methods section on page 7, lines 6-7. We also added a justification, why we did evaluate those measures and not others (see page 7, lines 9-12). In preparation of the European Health Interview Survey (EHIS) wave 2 it was discussed which dimensions of physical activity should be covered by the survey questionnaire total physical activity (generic instrument) or only leisure time physical activity (domain-specific instrument), as many other international health surveys comparable to EHIS have either focused only on health-enhancing leisure-time PA or on total PA. The cognitive testing study aimed at testing advantages and disadvantages of both options. The NHIS-PAQ and IPAQ-SF were selected because they were used in large-scale population studies before and their validity and reliability had been evaluated.
2) P. 2, Ch.2, l.7: *in the past decade, literature has also focus specifically on moderate to vigorous PA. Could the authors also comment on this?*

Yes, we agree, the questionnaires which had been used in the last decades have usually distinguished between moderate- and vigorous-intensity physical activities. We have added this information on page 4, line 13. This distinction created problems in the past because respondents did not understand the rational of this distinction and had problems to classify their activities into such intensity categories.

3) P.3, Ch.1, l.1: *some PAQs provide adaptations for specific populations, is it the case for this one? Could you also consider the importance of testing these PAQs on different age range?*

The IPAQ-SF is proposed to be used for the age range 18 to 65 years of age and the NHIS-PAQ is used for populations 18 years and older. The questionnaires were tested in a population 15 years and older which will be the age range covered by the EHIS. We have looked at different age strata when analyzing the data. The main finding was that the questionnaires in the age strata 15 to 19 years performed in a comparable way than for the age range 20 to 59 years. However, in the age group 60 + years retired persons reported difficulties to define leisure-time when answering the NHIS-PAQ. We now have added results on how the questionnaires worked in different age ranges in the last section of the results section, which now is called ‘country and age differences’ on page 12, lines 17-20 and we added more examples in the Table 3 which demonstrated the difficulties retired people had to define their leisure time.

Method

4) *Why did the authors choose a questionnaire that measure general PA and one specific on leisure time activity?*

See also answer on comment 1. The instrument eligibility criteria were set by the EHIS Core Group. The PA expert meeting recommended the IPAQ-SF and the EHIS- Core Group decided that a specific leisure-time PA questionnaire should also be tested which is valid and reliable and which has been used in large-scale population surveys before. A systematic search for PA survey questions was conducted using the ‘HIS/HES data base’ (https://hishes.wiv-isp.be/index.php?hishes=home). The HIS/HES data base presents an inventory of national and multi-country health surveys implemented in EU Member States as well as EFTA countries, EU Candidate Countries and USA, Canada and Australia. The NHIS-PAQ and IPAQ-SF were selected because they were used in large-scale population studies before and their validity and reliability were evaluated. We have now added a brief justification for this decision in the method section on page 7, lines 9-12.

5) *In the NHIS-PAQ, some questions are dedicated to adults, considering 18 y.o. and more, did the authors take this limit in consideration in its evaluation process?*
See also answer on comment 3. As the EHIS will focus on 15+, one aim of the testing was to evaluate whether the NHIS-PAQ (and IPAQ-SF) works as intended in this age group. The results showed that adolescents 15 to 19 years had not more problems than respondents in other age groups. However this result should be interpreted with caution as there were only 7 participants in this age group in the sample. We now have added results how the questionnaires worked in different age ranges on page 12, lines 17-20.

6) *They are different method of calculation for the IPAQ-SF, which one was utilized and on which sub-scores were the most difficulties located? Moreover, did the authors considered the different categories proposed to classify the individuals and if these categories are appropriate or not in regard to the cognitive evaluation process? Furthermore, did the authors considered the recoding process for high active individuals and periods of less than 10 minutes?*

To our understanding checking the analyzing guidelines of the instruments is beyond the scope of a cognitive testing study where the main focus is on the perspective of the individual when answering the questionnaire. In principle, in a CT study the respondent do not necessary have to provide true personal answers on the questions, they can even give fictive answers because the object of investigation is not the answer on the question as such but the cognitive processes involved to arrive at an answer. Thus, we have not considered such issues because this study was not a validation study and different calculation methods and how different sub scores performed in terms of validity were not tested. The main focus of the study was on cognitive processes of the respondents when answering the questions (the comprehension, simplicity, though process and certainty with the answers). However, we found different levels of difficulties with different sub-question as mentioned in the results and discussion sections.

7) *How did the authors considered the do not know answers? Did they try to identify why participants provided these answers?*

We did assess the thought process of respondents when answering the questions and revealed potential problems that could result in ‘do not know’ answers and item non-response. It was assessed whether respondents hesitated when answering questions or whether they were not able to answer specific question for varying reasons. This information was documented in the analyzing sheets and the underlying reasons for not being able to answer questions or having problems to answer questions were categorized by the themes comprehension, recalling activities, classifying activities and calculating frequencies and durations (see results section).

8) *P. 3, Ch. 1, l. 4: When did the meeting took place? Who was present? And how were the decision made?*
The meeting took place in January 2011. Researchers from the NIHD, Tallinn, Estonia, IPH, Brussels, Belgium and RKI, Berlin, Germany were present. Researchers from the Nat Cen London, UK participated via video conference. Michel Gray gave an input presentation about the methodology of cognitive testing and all participants decided together on the standard operation procedures (SOPs) via a consensus statement. We now clarify the date and the participants of the meeting on page 5 lines 10-12.

9) P.3, Ch 1, l. 5: When comparing PAQs internationally, there must be a strong emphasis on the translation process, more information are needed here about the method of translation, back translation and testing of the different tools, as it could be determining in the cognitive testing.

We agree that this is an important issue. The translation protocol followed the Eurostat guidelines: Step 1: Development of conceptual translation cards for the instruments. Conceptual translation cards for both instruments were provided by RKI in English. Conceptual translation cards served as the basis for translating the instruments from the source language (English) into the three European target languages (French, Estonian and German). Step 2: Initial or forward translation of all instruments from source to target languages. The translations of instruments were conducted by two independent translators in each country only for those instruments for which no valid version was available in the target language. Step 3: Panel adjudication part I. After translating two independent versions for each instrument, the two translators discussed the differences between their drafts, agreed on consensus solutions for the differences and drafted one preliminary translation product for each instrument. Step 4: Submission of the initial translation product for revision/approval. The preliminary translation products were submitted to a third person, preferable to an HIS core group member or MS partner fluent in the languages. The reviewer checked the translation product on whether the underlying concepts were complied with using the review questionnaire provided by Robin and Jagger 2003 (Report to Eurostat on European Health Status Module, p. 13ff). Step 5: Panel adjudication part II. In the last step, all three persons involved in the translation process (the two translators and the reviewer) discussed the reviewer comments on the proposals and decided on a final translation product for each instrument. The translation protocol and the translation process are documented in the final project report in the chapter ‘internationalization through translation’. We added a brief paragraph on the translation process in the method section on page 7 lines 13-17.

10) P.3, Ch 1, l. 6: How many different interviewers did realized the work? How were they trained?

In the UK three in Estonia six, in Belgium two and in Germany three interviewers realized the interviews. In the coordination meeting common interviewer training materials were developed and the interviewer training was carried out in a standardized way. We added this information in the method section on page 6 lines 10-13.

11) P.3, Ch 1, l. 6: Could the probe sheets be added as supplementary files?
Yes, we added the physical activity probe sheet as a Supplemental File 1. We refer to the file in the methods section on page 5, lines 14, 15.

Study sample
12) P3, Ch1, l.1: the different recruited individuals seemed to live in urban environment? How could this affect the results?

In the UK and in Estonia individuals from rural areas were included in the study as well. In general, the results might be influenced in that the physical activity behavior might be affected by environmental factors such as build environment, geo factors and working conditions which may be not the same in rural and urban areas.

13) P3, Ch1, l.3: recommendations for PA are different at the selected ages, as well as cognitive process regarding vocabulary acquisition, calculation; could the authors provide more information about this point and about expected results in regard to age range?

Yes, indeed the PA recommendation are different in different age groups: adolescents 15 to 17 years of age should perform moderate-to-vigorous intensity aerobic PA 60 minutes per day and at least 3 times per week muscle-strengthening PA, adults 18 to 64 years and those 65+ both should perform at least 150 min moderate-intensity aerobic PA or 75 min vigorous-intensity PA and at least 2 times muscle-strengthening PA; those 65 years and above should additionally perform PAs to enhance balance and prevent falls 3 times per week. As the main two areas of the recommendations on aerobic PA and muscle-strengthening PA are relevant for all age groups and just the dose is different the vocabulary used is the same for all age groups. As we have already pointed out in our answers on the comments 3 and 5 this study did not focus on the calculation procedure of PA recommendation compliance in different age groups. However, we now have added a paragraph in which we discuss age differences in regards to the CT results on page 13, lines 11-15. For example respondents aged 60 years and older had more difficulties in the comprehension and answering of the questions than those younger than 60. It could be that activity limitations (hearing and seeing conditions) and cognitive impairment, which increase with increasing age, play a role, but also the regularity of physical activity behavior – retired respondents reported more often than working-age respondents that their activities are irregularly and thus difficult to recall.

14) P3, Ch1, l. 5: A table summarizing the participants in the different country, including age range, gender and country would help the reader to better understand the population selected.

We added a table with characteristics of the respondents (see Table 2). We refer to the Table 2 in the results section, first sentence.

15) P4, Ch1, l.2: What was the incentive to participate to the study? Incentives could modify social desirability towards the task, could the authors consider this point?
The procedure of incentives was handled differently in the countries: In Germany, Belgium, and in the UK incentives were given after completion of the interview, whereas in Estonia there was no incentive. We added this information in the methods section (page 6, lines 20-21). It is difficult to say whether the incentive influenced the results of the CT study. Again it has to be noted that the participants were informed that the main focus of the study is on the cognitive process when answering the questionnaire and not on their answers on the questionnaires as such. However, we probed whether they think that social desirability may play a role when answering the PAQs (see Results section on ‘sensitivity and social desirability bias’, page 11). It could also be that social desirability may have influenced the CT results in terms that respondents may falsly indicate that they find the questions easy instead of difficult to answer since having good cognition skills might be perceived as a socially desirable characteristic [R (female 41, DE), IPAQ-SF: R had problems to understand the question, asked for response categories. She said, the question was okay to answer, but while answering the questionnaire she really had problems to understand it]. This could indicate that some respondents have in truth more problems with the questionnaire than they admitted during the CT interview. We added this example to the Table 3. However, it’s hard to say whether the incentive does influence social desirability in the context of a CT study.

16) P4, Ch2: Could the authors provide example of questions asked to the participants?

See answer on comment 11. We have now added the probe sheet in the Supplemental File 1.

17) P4, Ch2: Could the authors provide information about the length of the different interviews and the time taken to answer to the questionnaire and if some randomization between answering to one or the other PAQs was used?

Only the total interview time was assessed. However, the CT interview also included probes on mental health and alcohol consumption questions. The duration of the PA part was not assessed separately nor was the interview time for IPAQ-SF and NHIS assessed. The PA section was about one third of the total interview: In Belgium the CT interview total duration ranged between 1:00 to 2:30 hours and the average time was 1:20. In Germany the duration ranged between 55 minutes and 2:40 hours and the average time was 1:25 hours. In the UK the duration ranged between 1:10 and 2:40 hours and the average time was 1:45 hours. In Estonia the interview duration was between 1:00 and 2:15 hours (info on average is missing). All participants answered both PAQs and the NHIS-PAQ was always asked first and then the IPAQ-SF (see probe sheet, Supplemental File 1).

18) P4, Ch2: How was social desirability analyzed?
See Probe sheet in the Supplemental File 1. Respondents were ask whether they ‘think that some people might not give a truthful answer to this question’ and give reasons why this could be the case.

19) P4, Ch3, l.4: The link provided for the NHIS-PAQ lead to a general website, and does not give access to the questionnaire, which is not easy to find, could the authors provide the direct link?

We have replaced the link.

20) P5, Ch2, l.1: Who did translate the transcripts? How?

The transcripts as such were not translated; however, the interviewers did translate the relevant information when entering the results on the themes and probes into the data analysis sheet.

21) P5, Ch2, l.2: What was the content of the standardized analysis sheet? Could it be provided as supplementary material?

The template of the data analyses sheet now is provided in the Supplemental File 2. We refer to it in the first paragraph in the Methods section.

22) P5, Ch2, l.4: What does mean “group discussion”, who was involved in these discussions? Was this information considered as data and transcribed? How were the data analyzed? What does observed pattern means? How did the authors considered individual differences and profile in the data analysis? How was the country difference analyzed in regard to cultural specificities?

After the end of field work a group discussion took place in which the study coordinators and the interviewers participated. The aim of the group discussion was to discuss the interviewer’ experiences and observations and to summarize and classify the major problems the respondents had with the questionnaires. The group discussions were not recorded and transcribed but minutes were taken. With observed patterns we mean frequently observed interviewer’ observations or answers, we have now clarified this sentence in the manuscript. The respondents were grouped regarding the characteristics sex, age, activity level and occupation/education (only in the UK) when analyzing the data. The country differences were analyzed by a ‘compare and contrast exercise’ on whether specific problems reported were observed in only one country or many countries.

Results
23) In general, as the questions of the PAQs are not presented, it is not always easy to follow authors results and discussion. Maybe some precision about the content of the questionnaire and the difference between the two chosen PAQs for specific points (such as example provided) could increase the understanding and facilitate author’s argumentation.
The Supplemental File 1 (probe sheet) does include the full questionnaires. We now inform the reader about this on page 7, line 6-7.

24) P5, Ch3, l.1: There are no information in the method section about the evaluation of high skills and structured daily life? How was this evaluated and could the authors provide more information about participant’s characteristics before presenting such results?

The descriptions ‘high-skilled’ and ‘having a structured life’ were derived from the respondent’s information on their employment status and their professional position. For instance in the UK study ‘occupational type’ was assessed in the categories a) ‘semi routine and routine occupation’ (n = 4), b) ‘intermediate occupation’ (n=2), c) ‘lower supervisory and technical occupation’ (n = 1), d) ‘managerial and professional occupation’ (n= 7) and d) ‘not classifiable’ (n= 1), however, the professional position was not assessed in a standardized way in all countries but it was often mentioned by the respondent when answering the probes. Also the employment status was not systematically assessed but often mentioned as a reason why respondents had difficulties with the PAQs. For instance retired and unemployed persons stated that it is difficult for them to define leisure time when answering the NHIS-PAQ because they are unemployed or retired. In contrast, persons who reported that it was easy for them to recall leisure-time or work-related physical activities and to fill in the PAQ’s indicated that it is easy for them because they work regularly or have a regular exercise schedule or daily life. To provide an overall impression of the key findings at the beginning of the results section, those concepts were used to sum up and classify the main reasons which were mentioned by the respondents to explain why the PAQs were easy or hard for them to answer. We decided to remove those terms from the manuscript and now be more precise instead - see changes in the results section (page 8, Results section, second sentence).

We do now present a table with respondent’s characteristics (age, sex, country) in the method section.

25) P5, Ch3, l.2: What’s mean older respondents? How was structured life evaluated?

See answer on comment no. 24 on “structured life”. Probing results showed that in particular older persons had problems when answering the PAQs, for instance they had difficulties to read the long introductions of the IPAQ-SF, they had to read the questions twice before understanding the essence of the question and they also had difficulties to recall activities (see also examples in the table 2). Such problems were frequently observed in the age group 60+ years of age. We have now clarified this in the text (see page 8, Results section, second sentence).

26) P5, Ch3, l.5: How many participants preferred on PAQ to the other? The sentence here is not precise enough, as well as the way to calculate it is not clear in the method section.
Respondents were asked which of the two PAQs they prefer and which one is easier to answer (we mention this now in the method section where we also refer to the probe sheet in the Supplemental File 1, page 7, line 6). Overall 26 respondents indicated that that the NHIS-PAQ was easier to answer, 19 opted for the IPAQ-SF and 20 indicated that both were the same. 15 respondents preferred the NHIS-PAQ, 24 preferred the IPAQ-SF and 20 had no preference. Frequently reported reasons why respondents indicated that the NHIS-PAQ is easier to answer were that they don’t do any LTPA or that it measures exactly what they do, which is LTPA. Reasons why respondents indicated that IPAQ-SF is harder to answer than NHIS-PAQ were that it does require more calculations because all activities have to be summed up to report total PA. An important reason why more respondents preferred the IPAQ-SF instead of the NHIS-PAQ even though it is more difficult to answer was that they felt that it more accurately measures PA because they could include all activities they do and not only LTPA.

We now present the frequencies for the general probes on simplicity and preference comparing NHIS-PAQ and IPAQ-SF on page 8, lines 13-17).

27) Table 2: Five major problems with the PAQs are presented, and the table 2 includes only two themes, that are not the same than the five presented. Could the authors revise it to have at least one example for each major theme?

We agree with this suggestion and rearranged the Table 3 now providing at least one example for each theme.

28) Moreover, it would be easier for the reader to insert these examples in the text, rather than in a disconnected table, to facilitate the reading.

We think that leaving the examples in the table which supplements the results from the text is a better solution than including all examples in the text, this makes the results section more concise and better readable.

29) P6, Ch3, l.1: Was does many participants means? Did they have a specific profile?

Due to the usually small samples sizes in cognitive testing (qualitative) studies, it is not recommended to quantify the problems reported because it cannot be quantified how representative the problems are for the general survey population. Certain problematic features of the questionnaire for the survey population may remain undetected in the cognitive testing study or some detected problematic features in the sample interviewed may be overestimated in their relevance for the survey population. Although it can be controversially discussed, we avoided to report frequencies and percentages (‘quasi statistics’) for the problems reported and used ‘some’, ‘many’ and ‘most’ respondents instead. This is a common approach used in qualitative studies to
describe how often problems were reported (see Qualitative Inquiry July 2010 vol. 16 no. 6 475-482, ‘Using numbers in qualitative research’).

30) P6, Ch3, l.6: Could the authors provide some examples of misunderstanding of Moderate or vigorous PA?
Some respondents included activities which are stressful and mentally demanding but do not involve extra physical effort. More examples are: -R (female 26, UK): Interviewer observation: R felt that she contradicted herself when talking about ‘vigorous’ in NHIS-PAQ to which she gave different answers in IPAQ-SF. The way the descriptions were presented, made her think of the same term in 2 different ways. She does not get into a heavy sweat [NHIS-PAQ] but does do activities that need ‘hard physical effort’ as described in IPAQ-SF. R did not understand ‘moderate,’ as in both Sets, they also had different definitions. For ‘moderate activities’ she answered ‘yes’ in NHIS-PAQ and ‘no’ in IPAQ-SF. - R (female 33, EE) is confused about what to consider and what not: “In some questions it is necessary to take walking into account and in others not. Is sportive walking included to walking or not? All depends on the intensity, after all”. We included those examples in Table 3.

31) P6, Ch3, l.8: What does some respondents means? Which questions of the IPAQ-SF were confusing? All of them?
See answer on comment 29 on using statistics in qualitative studies. The long introductions and long question wordings of the all IPAQ-SF questions were difficult in general. Many respondents had to read the question more than once to grasp the information.

32) P7, Ch2, l.5: This is not a problem for me that participants consider some activities as having different intensities, as individual differences in engagement and energy expenditure could explain these differences. Could the authors provide more concrete examples of activities? And precise what some respondents means?
It was a problem that respondents did not know in which intensity category they should include their activities, examples: - R (female 33, EE) is confused about what to consider and what not: “In some questions it is necessary to take walking into account and in others not. Is sportive walking included to walking or not? All depends on the intensity, after all”. - R (female 62, UK) included a vigorous walk in Q1. This made it difficult for her to answer Q3; without the walks she had to say that she did no moderate activity. When she came to Q5 she thought she answered wrong and was confused; her vigorous walk potentially could fit in either section. - R (female 73, DE) talks about her gymnastics, thought about sweating, then answers Q8. R describes again her gymnastics and answers Q10. R describes again her gymnastics that are also strengthening, she
seems annoyed that questions are very similar, and answers Q12. The examples can be found in Table 3.

33) P7, Ch2, l.6: What does mean “not being sure of the answers provided”? How the authors did analyzed this point in the interview? Were they not sure when answering or afterwards when talking with the interviewer?

Certainty with the answer in CT studies can be examined directly or indirectly. The indirect form is that you assess the though process to arrive at an answer and it often comes out in this process whether the respondent is sure or not sure with the answer provided. Probes to assess the thought process are for example to ask how they decided what activities to include or not to include or how they calculated frequencies or durations of activities and also information on simplicity often gives hints whether the respondents are certain with their answers or not. The direct form is to directly ask how certain or uncertain respondents are that their answer is correct. Both forms were used in this study, see also probe sheet in the Supplemental File 1.

34) P8, Ch1, l.7-10: Could the authors provide some examples about the activities not included and about the activities not mentioned? Is there a regular pattern? For all participants?

- R (male 17, DE), IPAQ-SF: R seems a little worried about the difference between 'last seven days' and 'usually'; he was on holidays for skiing and is therefore quite confused about what to answer.

- R (female 59, DE), IPAQ-SF: She only thought about work-related activities, so the duration of vigorous PA might be too short; IPAQ-SF moderate activities: Her answer deals with cycling in leisure time, during the interview she said, she is going to work by bike every day; so here her answer might be an underestimation.

We included those examples into the Table 3. It is difficult to evaluate if there is a regular pattern/for all participants.

35) P8, Ch2, l.1-7: How did the authors measured that cognitive effort and calculation was more important from one PAQ to the other? Moreover, why was the time hard to calculate, due to the lack of precision in the recall, the low level in mathematics and the numerous number to sum?

We did not measure cognitive effort. The purpose of CT is to identify problems Rs have with the PAQs when answering them. Many Rs reported that it was more difficult to do the calculation for the IPAQ than for the NHIS. The cognitive effort was higher because the IPAQ-SF to requests to combine multiple activities to answer a single question that asks respondents about average time spent in multiple activities or across many days. The NHIS-PAQ only focuses on LTPA and less activities needs to be combined to answer a single question. This information was for example extracted from the answers (reasons) respondents provided to justify their answers on the General probes (1 and 2), see probe sheet in the Supplemental File 1.
36) P8, ch2, l.4: What does quicker to answer means (five minutes less? Is it really a big difference?)? What is the interest of comparing the speed of filling in with such different questionnaires, one measuring general PA and the other only leisure time?

The reason to assess hesitation when respondents answering PAQs in a cognitive testing context is that this information can be used to learn something about the “simplicity” of the questionnaire which often also influences the “certainty with the answer”. Reading a question more than one and hesitation when answering a question are one of the most important ‘interviewer observations’ in CT which are documented by the interviewer. The exact time is usually not assessed only if the answer was given quickly or with hesitation.

37) P9, ch1, l.9: Could the authors provide example of misclassification of vigorous PA instead of moderate PA?

See examples in the Table 3. - R (female 62, UK) included a vigorous walk in Q1. This made it difficult for her to answer Q3; without the walks she had to say that she did no moderate activity. When she came to Q5 she thought she answered wrong and was confused; her vigorous walk potentially could fit in either section. - R (female 73, DE) talks about her gymnastics, thought about sweating, then answers Q8. R describes again her gymnastics and answers Q10. R describes again her gymnastics that are also strengthening, she seems annoyed that questions are very similar, and answers Q12. - R (female 60, BE); Q1-12: “Everything is relative what we can call vigorous. For me vacuuming is vigorous, I’ve done it 4 times this week, for others it is of course moderate”.

38) P9, Ch2, l.1: The first sentence is not precise enough? If the pattern is the same in more than one country, does that mean that there are no cultural differences? Could the authors provide examples about the difficulties of translation?

We replace the word ‘cultural differences’ with ‘country differences’ to be more precise. The main problem during the translation process that occurred in Germany as well as in France was to find an appropriate translation for moderate and vigorous intensity of activities.

39) P9, ch2, l.7: How were the cultural differences analyzed?

See also answer on comment 39. It was evaluated whether certain problems that were frequently reported by the respondents differed across counties or not.

Discussion
40) P10, Ch1, l.3: Again, participants’ characteristics need to be described earlier to discuss this point. It could be compared to previous results of other studies also.

See also answer on comment 24. We have removed the characterization ‘structured life’ and ‘high-skilled’ from the discussion section and are now more precise. The results are compared to those of Altschuler et al. and Berrigan et al. in the discussion section.

41) P10, Ch1, l.5: The identified problems are similar, but also different than in previous studies, could the authors consider this point, to see if this is common to the self-reported of PA or PAQs or population specific?

We do not fully understand this comment. Do you mean if the differences and similarities with other studies can be due to different modes of data collection, different PAQ’s or settings were the studies were conducted? The study of Berigan et al. which observed finding in line to what we observed used the same mode of data collection (face-to-face interviews) and the same PAQ and a comparable sample. Thus, we think that the discussed similarities and differences are flawed by methodological differences between the studies.

42) P11, Ch1, l.1: What does household activities means?

Household activities are physical activities which are performed when doing household cores such as vacuuming, cleaning, ironing, washing dishes etc. sometimes also referred as to domestic activities which usually have a light to moderate intensity. Usually four domains of PA are distinguished where physical activity in the day course can take place: leisure-time PA, work-related PA, household activities and transport-related PA.

43) P11, Ch1, l.11: They are two different ideas that need to be discussed in the last sentence: (1) the difficulty of divide PA into categories and (2) the multiple reports of some activities. Could these findings be compared to the existing literature?

Yes, during the validation of the IPAQ-SF those problems were observed in many studies. Therefore many attempts were undertaken to solve the problem. But respondents still think that it is not intuitive to categories their activities into such intensity categories. Also Berrigan et al. and Altschuler et al. observed similar issues.

44) P11, Ch2, l.10: as the NHIS-PAQ only measure leisure time PA, this is normal that it underestimate work or other type of PA...Was the questionnaire intended to measure other type of PA? If it is needed, could the authors argue why?

This is correct, but still in many HIS studies only leisure time PA is assessed. One aim of this study was to assess advantages and disadvantages of this strategy (see also answer on comment 1 and 4). This study has demonstrates that this may not be a good strategy (to focus on LTPA only)
because some respondents will then include work-related PA in the leisure-time PA section or
they are frustrated that they look inactive even though they are active at work and their
motivation to participate in the study may be compromised.

45) P12, Ch2, l.1: There is no interest of comparing sample size of a qualitative versus a
quantitative analysis, rather an interest in criticizing the quality of the sample selection, ensuring
all different profiles have been included in the present study.

Those are the major limitations of qualitative studies therefore we think that it is reasonable to
discuss such issues in the limitation section.

Minor compulsory revisions

Abstract

46) l.3: the PAQs “work”: the English could be improved here

We have changed the wording we replaced the ‘worked’ with ‘performed’.

47) l.8: an acronym is used for PA without having referred to the complete wording before.

We have corrected this.

Method

48) What does EHIS means? When was the wave 1?

EHIS means European health Interview Survey the acronym was introduced in the Background
section (page 4, lines 5-6), EHIS wave 1 was conducted between 2006 and 2009; we have
included this information in the Methods section, first sentence.

49) The grant agreement number has to be placed in the acknowledgement.

We moved the grant agreement number to the acknowledgement section.

Reviewer 2:
Alexandre Mouton
Reviewer’s report:
Major Compulsory Revisions

50) - In your Results section, you refer to some characteristics of the respondents such high-
skilled, those who had less structured daily lifes, their employment status. Or, you do not refer to
a socio-demographic questioning in the Methods section. This information should be added in
this section for a better comprehension of your methodology. Moreover, it is not clear whether
you consider somebody as high- or low-skilled. This classification process should also be
presented in the Methods section.
See also answers on comments 3 and 5. We have decided to remove the characterization ‘structured life’ and ‘high-skilled’ from the manuscript and describe now more precisely what we mean. ‘High-skilled’ was defined by the UK study group using information on ‘occupational type’ which was assessed in the categories a) ‘semi routine and routine occupation’ (n = 4), b) ‘intermediate occupation’ (n=2), c) ‘lower supervisory and technical occupation’ (n = 1), d) ‘managerial and professional occupation’ (n= 7) and d) ‘not classifiable’ (n= 1).

Minor Essential Revisions
51) - In Table 2, you should specify to which PAQ you are referring to for each case example. A skilled-reader will be able to guess which PAQ is presented but it is not clear enough for other readers.

We now specify which example refers to which PAQ in Table 3.

Discretionary Revisions
52) - In the study sample section (page 6), it would be useful to specify which incentive was provided (financial?) to the participants.

See also answer on comment no. 15. The incentive was financial, we included this information on page 6, line 20.

53) - In the data analysis section (page 7), I would suggest the inclusion of an inter-rater reliability analysis for the data reduction process in order to strengthen the methodological quality of your study.

The study coordinators checked the information entered into the data analysis sheets randomly. We added this information in the method section.

54) - In the Results section (pages 7-11), it could be interesting to present the prevalence of the difficulties encountered by respondents (number of citations per problem presented). In fact, it is sometimes difficult to make the distinction between minor and major problems using only "some respondents" or "many respondents" statements. In this way, the reader would better understand the choice you have made for your suggestions of adaptations to improve PAQs and make them more suitable for the respondents.

See also answer on comment number 29. Although it can be controversially discussed, we avoided to report frequencies and percentages (‘quasi statistics’) for the problems reported and used ‘some’, ‘many’ and ‘most’ respondents instead. This is a common approach used in qualitative studies. However, we now have included some quantitative information in the results section on easiness and preferences comparing the NHIS-PAQ and IPAQ-SF at the beginning of the results section.
Once again, we would like to thank the reviewers for their comments, which have helped us to improve the manuscript considerably. We hope the revised manuscript can now be accepted for publication.

Sincerely yours,

The corresponding author on behalf of all authors