Reviewer's report

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

Version: 1 Date: 7 July 2015

Reviewer: Aurelie Van Hoye

Reviewer's report:

General comment

This publication aims at testing the cognitive response of participants from four different countries to two PA questionnaires (NHIS-PAQ and IPAQ-SF). This study is interesting, as many studies test questionnaire validity in regard to objective PA, without collecting participants’ point of view. The background resumes well the state of the art, excepted specifically for the two selected questionnaires and the rationale of choosing these instead of other instruments. The method section is lacking of precision about how the process was deployed and data analysed, such as translation process, training process of interviewer, interrater reliability and collective meeting description. The results presented are interesting, but examples are missing, as well as precision of number/characteristics of respondent, to precise author’s argumentation. The discussion could be deepened in adding more comparison to the existing literature; especially in regard to the specificity of cognitive interview results (are the results found specific to the instrument selected or a general pattern of PA self-reported measurement). The limitations and conclusions are clear and the take-home message is important in regard to future study.

Major compulsory revisions

Introduction

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?

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?

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?

Method

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

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
They are different methods of calculation for the IPAQ-SF, which one was utilized and on which sub-scores were the most difficulties located? Moreover, did the authors consider 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?

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

When did the meeting took place? Who was present? And how were the decision made?

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.

How many different interviewers did realized the work? How were they trained?

Could the probe sheets be added as supplementary files?

The different recruited individuals seemed to live in urban environment? How could this affect the results?

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?

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.

What was the incentive to participate to the study? Incentives could modify social desirability towards the task, could the authors consider this point?

Could the authors provide example of questions asked to the participants?

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?

How was social desirability analyzed?

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?

Who did translate the transcripts? How?
be provided as supplementary material?

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?

Results

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.

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?

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

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.

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? 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.

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

P6, Ch3, l.6: Could the authors provide some examples of misunderstanding of Moderate or vigorous PA?

P6, Ch3, l.8: What does some respondents means? Which questions of the IPAQ-SF were confusing? All of them?

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?

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?

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?
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?

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?

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

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?

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

Discussion

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.

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?

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

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?

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?

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.

Minor compulsory revisions

Abstract

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

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

Method

What does EHIS means? When was the wave 1?

The grant agreement number has to be placed in the acknowledgement.
**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I do not have any competing interests