

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Adaptation and validation of the Physical Activity Questionnaire for Adolescents (PAQ-A) among Polish adolescents – cross-sectional study
<b>AUTHORS</b>	Wyszyńska, Justyna; Małosz, Piotr; Podgórska-Bednarz, Justyna; Herbert, Jarosław; Przednowek, Krzysztof; Baran, Joanna; Dereń, Katarzyna; Mazur, Artur

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Matteo Christian Sattler University of Graz, Austria
<b>REVIEW RETURNED</b>	09-May-2019

<b>GENERAL COMMENTS</b>	<p>This work is valuable since PA is often assessed via self-reports. Current evidence shows that most questionnaires have unsatisfying validity. It is not only necessary to develop and improve PA questionnaires in order to ensure satisfying measurement quality but also to adapt questionnaires in different languages. This will allow comparisons between studies and countries. Because of this, the presented work represents an important contribution to the field. However, the manuscript needs substantial revision (e.g., in the writing, reporting and analysis).</p> <p>Introduction: General comments:</p> <ul style="list-style-type: none"><li>- The Introduction gives a nice and broad overview with many examples. However, it would be beneficial to the manuscript if the introduction is modified to increase readability (please check the grammar as well). For example, by providing health benefits of PA for both physical and mental health or by providing actual values behind important statements such as prevalence or incidence (second paragraph). The authors may also consider deleting sentences with redundant information. To my opinion, some of the sentences in the second paragraph do not really link to each other.</li><li>- It would be beneficial for the manuscript to provide a compact overview about the measurement properties of existing non-Polish PA questionnaires. What were the results for the general population and/or adolescents (e.g., from systematic reviews)? This will allow the reader to compare the presented results of the Polish version. The results of previous validation studies may also be used to argue that existing questionnaires often do not show good measurement qualities.</li><li>- The manuscript would improve if the authors mention any hypotheses about important agreements between the questionnaire and the accelerometer. What is the minimum correlation indicating good validity/reliability.</li></ul>
-------------------------	---

Specific comments:

- First paragraph: Please provide a reference for the definition of PA and you may also provide examples for health promotion through PA (e.g., improvements in mental health or social wellbeing).
- Second paragraph: Please cite the mentioned PA guidelines, either of the WHO or the recently updated guidelines for the Americans (2018).
- Second paragraph: "Developmental age population": Does this term indicate young adults, children or adolescents? Since you refer to the percentage of inactivity in the following sentences, it may also be helpful to state the exact number for the Polish adolescent population. This will allow the comparison to the 80% mentioned before.
- Incidence of obesity: Please mention any number to make these sentences more clearly. For example, a sentence such as: "Within a decade, the number of overweight children has doubled" is not helpful since no baseline number is provided.
- Third paragraph: You may carefully use the term "objective" assessment of PA. For example, accelerometers seldom meet this criterion and should be called device-based measures of PA instead (Because of the interaction between participant and instrument during the wear period).
- Third paragraph: The authors may consider to provide more substantial information about the problems and advantages of different PA measurement tools and why researchers should rely on questionnaires, why not on interviews or PA diaries?
- Last paragraph: To my opinion, reliability includes the term repeatable. Please avoid repeating identical wording (e.g., p3., line 54 vs. 60).

Methods

- First paragraph: It would help the reader to follow the manuscript when the study design/setting (including an overview of the procedure) is already mentioned in the first paragraph followed by participants. You may also consider to mention the exact design of the study (see strobe guidelines).
- P. 4: "removal of the accelerometer at any time during the study period (n=23)" you may explain this better. Does this mean the week was not valid or did you really exclude participants because of any existing non-wear periods?
- Why did you exclude participants because of incorrect anthropometry: Please clarify to me?
- What do you mean with exclusion because of incorrect participant identification?
- Please introduce and describe the following measures used in the results section already in the method section: MVPA, steps/day, energy expenditure. I could not find them in the method section. How did you calculate these measures?
- P. 5, line 18: "but not during holidays"?
- P. 5, line 21: obtains instead of evaluate?
- P. 5, line 33: last question was not...
- P. 6, line 29: It may not be necessary to mention the exact model number again.
- P. 6, line 40-42: Please check the grammar of this sentence.
- P. 6, line 44: Did the participants got a diary to fill in any non-wear periods? Please mention if and how you adjust for differences in non-wear periods in the accelerometer

	<p>measurements. This should not be neglected. In addition, please present descriptive information about the amount of non-wear also in the results section.</p> <ul style="list-style-type: none"> <li>- Please explain to me how you dealt with sleeping time since the protocol was 24 h. Was there an additional diary or algorithm in the software? For example, sleeping time should be excluded from sedentary time.</li> <li>- P. 7, line 7: Since many different formulas of ICCs exist, it would be helpful to mention the ICC version used. This would increase comparability and transparency. Please consider also if you need different versions of the ICC for a single item versus the overall mean score of activity.</li> <li>- P. 7, line 9: Cronbach Alpha is an index of inter-item correlation and should only be used within constructs consisting of reflective indicators see (doi: 10.2165/11531370-000000000-00000 OR doi: 10.1007/s40279-018-0961-x) for more information. For example, in contrast to psychological constructs, a specific item in a PA questionnaire does not necessarily has to correlate with the other items. The only important influence on the inter-item correlation is (similar to accelerometer measurements) that a day is limited to 24 h. Therefore, results for Cronbach Alpha may be omitted.</li> <li>- P. 7, line 18: In validation studies, the magnitude (as an effect size) of the correlation is more important than the p value. For example, some authors have recommended that positive validity is indicated by correlations <math>\geq 0.5</math>. This may be considered when interpreting the results.</li> </ul> <p><b>Results</b></p> <p>General comments:</p> <ul style="list-style-type: none"> <li>- You may consider omitting sedentary time from the whole manuscript (e.g., method section) since there seems to be no corresponding score in the PAQ.</li> <li>- Since you had many dropouts it seems to be important to assess differences between included and excluded participants. For example, regarding age, sex, BMI, PA level. Any upcoming differences from this analysis should also be discussed later.</li> </ul> <p>Specific comments:</p> <ul style="list-style-type: none"> <li>- See also previous comments.</li> <li>- For example, p. 9., line 3, a significant association does not indicate good validity. The coefficients may still be very small, only the sample size was large. Therefore, you may also consider omitting statistical significance in this case. Also, the three asterisks in the tables may mislead the reader to the interpretation of statistical significance rather than the magnitude of agreement.</li> <li>- P. 9, Table 4: Please explain to me how you calculated the measure steps per day and why you used min/h for MVPA. Please also explain how you calculated these measures, including energy expenditure (see also previous comments).</li> </ul> <p><b>Discussion</b></p> <p>General comments:</p> <ul style="list-style-type: none"> <li>- The readability of the discussion is good but can be improved by showing the interpretation in different/separated paragraphs. Also, it can be helpful to compare the validity of the PAQ to other PA questionnaires or even diaries/records. Since the results of the questionnaire seem quite promising.</li> <li>- Please check the grammar.</li> </ul> <p>Specific comments:</p>
--	---

	<p>- See comments above regarding the magnitude of the correlation. For example, p. 9 line 44, the explanation of weak/strong/very strong correlation coefficients is missing in the methods or for example in a priori hypotheses (what you would expect).</p> <p>- P 9., line 47: the total score is not a measurement tool. Same sentence: Please write consistently adolescents throughout the manuscript.</p> <p>- P. 9, line 54: Please mention any results of other studies. Also, you may not focus too much on individual items in the discussion, since the main score is the total score of the questionnaire.</p> <p>- P. 10, line 48: "the most promising tool regarding PA self-reports". This statement needs to be supported by substantial literature comparison. What were the previous results for diaries/records/recalls?</p> <p>- Limitations: Please consider the relatively high dropout of participants due to accelerometer non-wear periods and the missing of a sleeping/nonwear diary as potential limitations. You may also consider including some strengths of the study.</p> <p>- Limitations, last sentence: I would be careful with this sentence since this is not true. Please explain to me otherwise. Do you mean if you change the cut points of moderate or vigorous PA but not the one for sedentary?</p> <p>Abstract:  - See all previous comments.  - Conclusions: Please check the grammar of the sentence and consider using adolescents instead of children.</p>
--	---

<b>REVIEWER</b>	Gunay Yildizer Eskisehir Technical University-Turkey
<b>REVIEW RETURNED</b>	01-Jun-2019

<b>GENERAL COMMENTS</b>	<b>REVIEW REPORT</b>
	<ol style="list-style-type: none"> <li>1. The title and abstract includes the study's design with a commonly used terms</li> <li>2. Authors indicated that they utilized mixed method research design to check the validity and reliability of the PAQ-A in Polish culture, they also should state that which mixed method research design they have utilized. Please refer  Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., &amp; Hanson, W. E. (2003). <i>Advanced mixed methods research designs. Handbook of mixed methods in social and behavioral research</i>, 209, 240.</li> <li>3. Authours indicated that the sample of the research consisted of adolescents aged 14-19 in the abstract but 14-20 in the methodology. The age range should be clearly stated in the manuscript.</li> <li>4. Authors explained the importance, need and scientific background of adapting PAQ-A into Polish culture. However, there are other questionnaires for measuring PA level among adolescents, such as IPAQ-A and Adolescent Physical Activity Recall Questionnaire. Explaining the reason for choosing PAQ-A rather than any other</li> </ol>

	<p>questionnaire would contribute to the rationale for the investigation.</p> <ol style="list-style-type: none"> <li>5. The objective of the study clearly stated.</li> <li>6. As previously stated, what type of mixed method research design was used in this study should be explained with the reasons of choosing this design for readers to comprehend the rationale behind the scientific approach for such objective.</li> <li>7. As previously stated, the authors declared that the age ranged between 14-20, on the other hand, it was 14-19 in the abstract. If they distributed consent forms, questionnaires, and accelerometers, and somehow only collected data from adolescents aged 14-19, replacing 20 with 19 would be more beneficial for readers to refrain from any misunderstanding.</li> <li>8. Although there is a very limited number of participants in this study, the age range is relatively wide (14-19, adolescents with 6 different age groups). Authors should state frequency distribution with respect to age, as they did for the weight status in Table1. It also should be balanced with respect to gender variable as well.</li> <li>9. The sampling protocol is not clear, there is one random sampling stage for schools but no clear explanations of how students were chosen in those schools. Were there any states such as class?</li> <li>10. The initial measurement protocol, cultural adaptation of the PAQ-A and its' features were well explained in the methodology.</li> <li>11. Authors should explain why did they choose to place accelerometer on waste rather than the wrist. Please see Fairclough et al., (2016) for a relevant explanation.</li> </ol> <p>Fairclough, S., Noonan, R., Rowlands, A., Van Hees, V., Knowles, Z., &amp; Boddy, L. (2016). Wear compliance and activity in children wearing wrist and hip mounted accelerometers. <i>Medicine &amp; Science in Sports &amp; Exercise</i>, 48(2), 245-253.</p> <ol style="list-style-type: none"> <li>12. Recent data indicates short epochs are more convenient for detecting physical activity behavior. Authors should state why did they choose 60-sec epochs. Is this a limitation engendered by the accelerometer? Please explain and see Sanders et al. (2014) and Fröberg et al. (2017), these articles might be helpful for an explication of the protocol.</li> </ol> <p>Fröberg, A., Berg, C., Larsson, C., Boldemann, C., Raustorp, A. 2017. "Combinations of Epoch Durations and Cut-Points to Estimate Sedentary Time and Physical Activity among Adolescents", <i>Measurement in Physical Education and Exercise Science</i>, 21(3), 154-160.</p> <p>Sanders, T., Cliff, D. P., Lonsdale, C. 2014. "Measuring Adolescent Boys' Physical Activity: Bout Length And The Influence Of Accelerometer Epoch Length", <i>PLoS One</i>, 9(3).</p>
--	---

	<p>13. Authors also should clearly explain why did they choose the cut-off points of Butte et al. What is the rationale of choosing this protocol, which is for pre-schoolers, not Evenson et al. (2008).</p> <p>There are serious problems related to cut off points. Authors indicated the following sentence under the headline Strength and limitation of this study</p> <p style="padding-left: 40px;">“In our study, we used cut-points by Butte et al., other cut-points would have yielded different results”</p> <p>Although they repeat the same sentence in the limitations, they also indicated the following statement without any reference</p> <p style="padding-left: 40px;">“Nonetheless, the relationships and differences with total PA will continue to be the same because this variable must not be highly dependent on cutoff values.”</p> <p>In most cases, the amount of physical activity measured will highly depend on the cut points used, mainly because they use different counts to define moderate PA. To illustrate, using 2000 counts per minute for sure won't give the same results when using 2600 counts for MVPA. Secondly, it does not solely depend on cutoff points but also epochs from which the cut points were obtained.</p> <p>14. Please see the following references for an explanation, and possible correction</p> <p>Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., &amp; McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. <i>J Sports Sci</i>, 26(14), 1557-1565.</p> <p>Trost, S. G., Loprinzi, P. D., Moore, R., &amp; Pfeiffer, K. A. (2011). Comparison of accelerometer cut points for predicting activity intensity in youth. <i>Med Sci Sports Exerc</i>, 43(7), 1360-1368.</p> <p>Trost et al (2011) justify why are the Evenson better than the others, even although they initially derived from older children. Please read this article for improving your manuscript.</p> <p>15. The protocol of collecting data should be comprehensively explained under the title of “Procedure”. If data collected in the classroom setting. This kind of measurements has some methodological problems such as independence of observation. When adolescents don't comprehend one item in the questionnaire, they generally tend to copy answers of their peers. How researchers are ensured about the independence of the observation. Please clearly state procedure.</p> <p>16. Statistical procedure and results are well explained for the used “cut off” points.</p>
--	--

	<p>17. Characteristics were given and the reasons for non-participation were explained in methodology.</p> <p>18. Authors summarized the key results with reference to study objectives.</p> <p>19. Authors should consider improving the discussion part. The results of accelerometer data were well discussed, but ICC and Cronbach's coefficient data was not well discussed.</p> <p>20. The last paragraph of the discussion clearly states why authors actually chose to adapt PAQ-A into Polish culture. Moving this paragraph to the introduction and restricting it accordingly would improve the rationale of the manuscript. Moreover, finishing the discussion with information related to PAQ-C causes a loss of emphasis on PAQ-A.</p> <p>21. Once again, authors state that 14-19 years old adolescents are in the Polish school system under the title of limitation. Then how did you invited adolescents aged 14-20 (possible fail the class, even if so please just indicate 14-19 years of age to refrain from misunderstanding).</p> <p>22. Authors indicated that they used cut-points extensively used in pediatric research but there are no references of that extensive literature.</p>
--	---

## VERSION 1 – AUTHOR RESPONSE

### REVIEWER: 1

Introduction:

General comments:

- The Introduction gives a nice and broad overview with many examples. However, it would be beneficial to the manuscript if the introduction is modified to increase readability (please check the grammar as well). For example, by providing health benefits of PA for both physical and mental health or by providing actual values behind important statements such as prevalence or incidence (second paragraph). The authors may also consider deleting sentences with redundant information. To my opinion, some of the sentences in the second paragraph do not really link to each other.

- It would be beneficial for the manuscript to provide a compact overview about the measurement properties of existing non-Polish PA questionnaires. What were the results for the general population and/or adolescents (e.g., from systematic reviews)? This will allow the reader to compare the presented results of the Polish version. The results of previous validation studies may also be used to argue that existing questionnaires often do not show good measurement qualities.

- The manuscript would improve if the authors mention any hypotheses about important agreements between the questionnaire and the accelerometer. What is the minimum correlation indicating good validity/reliability.

The article has been corrected by a native English speaker. The Introduction was almost entirely redrafted taking into account the suggestions of the reviewer.

Specific comments:

- First paragraph: Please provide a reference for the definition of PA and you may also provide examples for health promotion through PA (e.g., improvements in mental health or social wellbeing).
- Second paragraph: Please cite the mentioned PA guidelines, either of the WHO or the recently updated guidelines for the Americans (2018).
- Second paragraph: "Developmental age population": Does this term indicate young adults, children or adolescents? Since you refer to the percentage of inactivity in the following sentences, it may also be helpful to state the exact number for the Polish adolescent population. This will allow the comparison to the 80% mentioned before.
- Incidence of obesity: Please mention any number to make these sentences more clearly. For example, a sentence such as: "Within a decade, the number of overweight children has doubled" is not helpful since no baseline number is provided.
- Third paragraph: You may carefully use the term "objective" assessment of PA. For example, accelerometers seldom meet this criterion and should be called device-based measures of PA instead (Because of the interaction between participant and instrument during the wear period).
- Third paragraph: The authors may consider to provide more substantial information about the problems and advantages of different PA measurement tools and why researchers should rely on questionnaires, why not on interviews or PA diaries?
- Last paragraph: To my opinion, reliability includes the term repeatable. Please avoid repeating identical wording (e.g., p3., line 54 vs. 60).

The introduction has been significantly revised, taking into account the suggestions of the reviewer.

Methods

- First paragraph: It would help the reader to follow the manuscript when the study design/setting (including an overview of the procedure) is already mentioned in the first paragraph followed by participants. You may also consider to mention the exact design of the study (see strobe guidelines).

Corrected as suggested by the reviewer.

- P. 4: "removal of the accelerometer at any time during the study period (n=23)" you may explain this better. Does this mean the week was not valid or did you really exclude participants because of any existing non-wear periods?

The text has been corrected so that it does not raise any doubts. We excluded 23 participants due to the week was not valid.

- Why did you exclude participants because of incorrect anthropometry: Please clarify to me?



One participant from the study group had an incorrectly recorded body height (16 cm), and therefore we did not qualify this participant, considering that in the study group are qualified only participants with full data.

- What do you mean with exclusion because of incorrect participant identification?

The authors failed to identify 2 participants of the study, due to errors in the coding of personal data.

- Please introduce and describe the following measures used in the results section already in the method section: MVPA, steps/day, energy expenditure. I could not find them in the method section. How did you calculate these measures?

In connection with the reviewers' comments, the authors have significantly rewritten the article. The methods section has been supplemented with information related to the calculation of MVPA and daily number of steps. After analyzing the available literature on a similar themes, we decided to delete the data about energy expenditure.

- P. 5, line 18: "but not during holidays"?

- P. 5, line 21: obtains instead of evaluate?

- P. 5, line 33: last question was not...

- P. 6, line 29: It may not be necessary to mention the exact model number again.

- P. 6, line 40-42: Please check the grammar of this sentence.

Thank you for the above remarks, they have been included.

- P. 6, line 44: Did the participants got a diary to fill in any non-wear periods? Please mention if and how you adjust for differences in non-wear periods in the accelerometer measurements. This should not be neglected. In addition, please present descriptive information about the amount of non-wear also in the results section.

The participants did not get a diary to fill in any non-wear periods. The participants were instructed to wear the accelerometer for seven consecutive days, 24 hours a day, excluding activities in water. If the wear-time was shorter than the minimum specified in the Methods section, such participants was excluded from the study. For the aim of this paper, we only considered two parameters: MVPA and daily step count. The values of these parameters are not disrupted when the non-wear time is not recorded in the diary. Additionally, in the present study, we did not calculate the sedentary time, which could be incorrectly calculated in a situation where non-wear time is not registered or the sleep time is not excluded from analysis.

- Please explain to me how you dealt with sleeping time since the protocol was 24 h. Was there an additional diary or algorithm in the software? For example, sleeping time should be excluded from sedentary time.

We recommended the participants not to take off the accelerometer during the night in order to prevent forgetting to put it on in the morning next day. Sedentary time was not the subject of our studies, thus information related to this parameter have not been used by us.

- P. 7, line 7: Since many different formulas of ICCs exist, it would be helpful to mention the ICC version used. This would increase comparability and transparency. Please consider also if you need different versions of the ICC for a single item versus the overall mean score of activity.

In our paper we used one way ICC calculated using R Software with irr package. In the paper the following reference was added: Bartko, J.J. (1966). The intraclass correlation coefficient as a measure of reliability. Psychological Reports, 19, 3-11.

- P. 7, line 9: Cronbach Alpha is an index of inter-item correlation and should only be used within constructs consisting of reflective indicators see (doi: 10.2165/11531370-000000000-00000 OR doi: 10.1007/s40279-018-0961-x) for more information. For example, in contrast to psychological constructs, a specific item in a PA questionnaire does not necessarily has to correlate with the other items. The only important influence on the inter-item correlation is (similar to accelerometer measurements) that a day is limited to 24 h. Therefore, results for Cronbach Alpha may be omitted.

Thank you for your comment. After analyzing other studies on reliability and validity of the PAQ-A, we decided not to omit the results for Cronbach Alpha. It allowed us to compare our results with the results of other authors, who in some reports introduced modifications involving the change of questions compared to the original version of the PAQ-A.

- P. 7, line 18: In validation studies, the magnitude (as an effect size) of the correlation is more important than the p value. For example, some authors have recommended that positive validity is indicated by correlations  $\geq 0.5$ . This may be considered when interpreting the results.

Thank you for your valuable comment. In the results, we focused on the magnitude of the effects (correlation coefficients).

## Results

General comments:

- You may consider omitting sedentary time from the whole manuscript (e.g., method section) since there seems to be no corresponding score in the PAQ.

We agree with the reviewer, the data related to sedentary time has been omitted.

- Since you had many dropouts it seems to be important to assess differences between included and excluded participants. For example, regarding age, sex, BMI, PA level. Any upcoming differences from this analysis should also be discussed later.

Unfortunately, we do not have data from all participants who were excluded from the research (e.g. due to refusals in the study, non-qualifications due to recent injuries affecting the PA level) or incomplete

questionnaires, therefore we are not able to present data that the reviewer requires. Given the fact that this is an important issue, it is described in the section Limitation.

Specific comments:

- See also previous comments.

- For example, p. 9., line 3, a significant association does not indicate good validity. The coefficients may still be very small, only the sample size was large. Therefore, you may also consider omitting statistical significance in this case. Also, the three asterisks in the tables may mislead the reader to the interpretation of statistical significance rather than the magnitude of agreement.

The asterisks have been removed from the table. The interpretation was focused on describing the size of the effect.

- P. 9, Table 4: Please explain to me how you calculated the measure steps per day and why you used min/h for MVPA. Please also explain how you calculated these measures, including energy expenditure (see also previous comments).

Based on reviewers' comments, we decided to perform new calculations. Data was collected in 5s epochs. We used the cutoff points according to Evenson to assess MVPA. MVPA in the new version of the article is given in minutes/day. The calculation of the presented parameters is given in the Methods section.

Discussion

General comments:

- The readability of the discussion is good but can be improved by showing the interpretation in different/separated paragraphs. Also, it can be helpful to compare the validity of the PAQ to other PA questionnaires or even diaries/records. Since the results of the questionnaire seem quite promising.

- Please check the grammar.

Specific comments:

- See comments above regarding the magnitude of the correlation. For example, p. 9 line 44, the explanation of weak/strong/very strong correlation coefficients is missing in the methods or for example in a priori hypotheses (what you would expect).

Missing information has been added: "Values less than 0.5 are indicative of poor reliability; between 0.5 and 0.75 indicate moderate reliability; between 0.75 and 0.9 indicate good reliability, and greater than 0.90 indicate excellent reliability" [Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;**15**:155-63] and "...It was performed using Spearman's rank correlation coefficients ( $\rho$ ). Values from 0.1 to 0.2 indicate poor correlation; from 0.3 to 0.5 fair; from 0.6 to 0.7 moderate and  $\geq 0.8$  very strong correlation" [Chan YH. Biostatistics 104: correlational analysis. *Singap Med J* 2003;44:614-19.]

- P 9., line 47: the total score is not a measurement tool. Same sentence: Please write consistently adolescents throughout the manuscript.

- P. 9, line 54: Please mention any results of other studies. Also, you may not focus too much on individual items in the discussion, since the main score is the total score of the questionnaire.

- P. 10, line 48: "the most promising tool regarding PA self-reports". This statement needs to be supported by substantial literature comparison. What were the previous results for diaries/records/recalls?

Thank you for the above comments. The article has been adapted to the reviewer's guidelines.

- Limitations: Please consider the relatively high dropout of participants due to accelerometer non-wear periods and the missing of a sleeping/nonwear diary as potential limitations. You may also consider including some strengths of the study.

- Limitations, last sentence: I would be careful with this sentence since this is not true. Please explain to me otherwise. Do you mean if you change the cut points of moderate or vigorous PA but not the one for sedentary?

Thank you for your comments. We agree with the reviewer that both the change of cut-off points and epoch values are significant in the assessment of the level of physical activity. Therefore, the incorrect sentence from the text has been removed.

Abstract:

- See all previous comments.

- Conclusions: Please check the grammar of the sentence and consider using adolescents instead of children.

Thank you for the above comments. The article has been adapted to the reviewer's guidelines.

## **REVIEWER: 2**

The cutoff points and epochs should be reconsidered again. Please follow the attached file.

Thank you for the above comments. The article has been adapted to the reviewer's guidelines. New calculations have been performed. Data was collected in 5s epochs and we used the cutoff points according to Evenson.

### **FORMATTING AMENDMENTS (if any)**

Required amendments will be listed here; please include these changes in your revised version:

1. The in text citation for "Additional file" is missing in your main text of your main document file. Please amend accordingly.

Thank you for the above comments. The article has been adapted to the reviewer's guidelines.

## 2. Patient and Public Involvement:

We have implemented an additional requirement to all articles to include 'Patient and Public Involvement' statement within the main text of your main document. Please refer below for more information regarding this new instruction:

Authors must include a statement in the methods section of the manuscript under the sub-heading 'Patient and Public Involvement'.

This should provide a brief response to the following questions:

How was the development of the research question and outcome measures informed by patients' priorities, experience, and preferences?

How did you involve patients in the design of this study?

Were patients involved in the recruitment to and conduct of the study?

How will the results be disseminated to study participants?

For randomised controlled trials, was the burden of the intervention assessed by patients themselves?

Patient advisers should also be thanked in the contributorship statement/acknowledgements.

If patients and or public were not involved please state this.

Patients and/or public were not involved in the design or planning of the study, what has been described in the Methods section.

## REVIEW REPORT

1. The title and abstract includes the study's design with a commonly used terms.

2. Authors indicated that they utilized mixed method research design to check the validity and reliability of the PAQ-A in Polish culture, they also should state that which mixed method research design they have utilized. Please refer Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). *Advanced mixed methods research designs. Handbook of mixed methods in social and behavioral research*, 209, 240.

Thank you for your comments. We have added information about the study design (cross-sectional study).

3. Authors indicated that the sample of the research consisted of adolescents aged 14-19 in the abstract but 14-20 in the methodology. The age range should be clearly stated in the manuscript.

Thank you for your comments. Due to the fact that PAQ-A was designed in Canada, for high school students who are currently in the school system (approximately ages 14-20), it does not correspond to

the education system in Poland (in which students at the age of 19 finish high school). Nevertheless, students aged 14-20 were invited to the study, bearing in mind that higher education can be provided to students also at age 20 (for various individual reasons). However, we collected data from adolescents aged 14-19. Therefore, in order not to mislead our readers, we have decided to use the range of 14-19 years in all sections of the paper.

4. Authors explained the importance, need and scientific background of adapting PAQ-A into Polish culture. However, there are other questionnaires for measuring PA level among adolescents, such as IPAQ-A and Adolescent Physical Activity Recall Questionnaire. Explaining the reason for choosing PAQ-A rather than any other questionnaire would contribute to the rationale for the investigation.

The Introduction section has been rewritten. The authors decided on language adaptation and validation of PAQ-A due to the fact that PAQ-A is regarded as one of the most suitable self-report tools for examining PA in adolescents [Biddle SJ, Gorely T, Pearson N, Bull FC. An assessment of self-reported physical activity instruments in young people for population surveillance: Project ALPHA. *Int J Behav Nutr Phys Act.* 2011;8:1.]. Moreover, IPAQ-A in the Polish version already exists. In addition, we received the PAQ-A author's approval for the process of linguistic adaptation and validation of the questionnaire.

5. The objective of the study clearly stated.

6. As previously stated, what type of mixed method research design was used in this study should be explained with the reasons of choosing this design for readers to comprehend the rationale behind the scientific approach for such objective.

Thank you for your comments. The article has been adapted to the reviewer's guidelines. After analyzing the publication on cultural adaptation, reliability and validity of the PAQ-A, we decided to change the type of research design to "cross-sectional study", which is a type of observational study that analyzes data from a population or a representative subset, at a specific point in time.

7. As previously stated, the authors declared that the age ranged between 14-20, on the other hand, it was 14-19 in the abstract. If they distributed consent forms, questionnaires, and accelerometers, and somehow only collected data from adolescents aged 14-19, replacing 20 with 19 would be more beneficial for readers to refrain from any misunderstanding.

Thank you for your comments. The explanation is described in point no 3.

8. Although there is a very limited number of participants in this study, the age range is relatively wide (14-19, adolescents with 6 different age groups). Authors should state frequency distribution with respect to age, as they did for the weight status in Table 1. It also should be balanced with respect to gender variable as well.

In Table 1, we placed the percentage distribution of the age of the examined group.

9. The sampling protocol is not clear, there is one random sampling stage for schools but no clear explanations of how students were chosen in those schools. Were there any states such as class?

Information on the selection of the study group was supplemented.

10. The initial measurement protocol, cultural adaptation of the PAQ-A and its' features were well explained in the methodology.

11. Authors should explain why did they choose to place accelerometer on waste rather than the wrist. Please see Fairclough et al., (2016) for a relevant explanation. Fairclough, S., Noonan, R., Rowlands, A., Van Hees, V., Knowles, Z., & Boddy, L. (2016). Wear compliance and activity in children wearing wrist and hip mounted accelerometers. *Medicine & Science in Sports & Exercise*, 48(2), 245-253.

Fairclough et al. (2016) confirmed that children prefer wrist-worn devices than hip-mounted accelerometers, what is confirmed by longer wear time, when they used wrist-worn devices. Wrist-worn devices may not only result in superior compliance but according to the recent evidence, may also provide better estimates of children's energy expenditure compared with hip-mounted accelerometers (Crouter SE, Flynn JI, Bassett DR Jr. Estimating physical activity in youth using a wrist accelerometer. *Med Sci Sports Exerc*. 2015; 47(5):944–51). However, we chose to place accelerometer on waist, to be able to compare the results of our research with the research of other authors (who have made validation of PAQ-A using accelerometers placed on the waist).

12. Recent data indicates short epochs are more convenient for detecting physical activity behavior. Authors should state why did they choose 60-sec epochs. Is this a limitation engendered by the accelerometer? Please explain and see Sanders et al. (2014) and Fröberg et al. (2017), these articles might be helpful for an explication of the protocol.

Fröberg, A., Berg, C., Larsson, C., Boldemann, C., Raustorp, A. 2017. "Combinations of Epoch Durations and Cut-Points to Estimate Sedentary Time and Physical Activity among Adolescents", *Measurement in Physical Education and Exercise Science*, 21(3), 154-160.

Sanders, T., Cliff, D. P., Lonsdale, C. 2014. "Measuring Adolescent Boys' Physical Activity: Bout Length And The Influence Of Accelerometer Epoch Length", *PLoS One*, 9(3).

13. Authors also should clearly explain why did they choose the cut-off points of Butte et al. What is the rationale of choosing this protocol, which is for pre-schoolers, not Evenson et al. (2008).

There are serious problems related to cut off points. Authors indicated the following sentence under the headline Strength and limitation of this study

"In our study, we used cut-points by Butte et al., other cut-points would have yielded different results"

Although they repeat the same sentence in the limitations, they also indicated the following statement without any reference "Nonetheless, the relationships and differences with total PA will continue to be the same because this variable must not be highly dependent on cutoff values." In most cases, the amount of physical activity measured will highly depend on the cut points used, mainly because they use different counts to define moderate PA. To illustrate, using 2000 counts per minute for sure won't give the same results when using 2600 counts for MVPA. Secondly, it does not solely depend on cutoff points but also epochs from which the cut points were obtained.

14. Please see the following references for an explanation, and possible correction Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *J Sports Sci*, 26(14), 1557-1565.

Trost, S. G., Loprinzi, P. D., Moore, R., & Pfeiffer, K. A. (2011). Comparison of accelerometer cut points for predicting activity intensity in youth. *Med Sci Sports Exerc*, 43(7), 1360-1368.

Trost et al (2011) justify why are the Evenson better than the others, even although they initially derived from older children. Please read this article for improving your manuscript.

Thank you for your valuable comments. We agree with the reviewer and this is a reason why we decided to perform a new calculations. Data was collected in 5s epochs and we used the cut-off points according to Evenson. Results are presented in the article.

15. The protocol of collecting data should be comprehensively explained under the title of "Procedure". If data collected in the classroom setting. This kind of measurements has some methodological problems such as independence of observation. When adolescents don't comprehend one item in the questionnaire, they generally tend to copy answers of their peers. How researchers are ensured about the independence of the observation. Please clearly state procedure.

Study participants were instructed to complete the questionnaires at home. To avoid possible errors in completing the questionnaires, help and supervision of researchers was offered.

16. Statistical procedure and results are well explained for the used "cut off" points.

17. Characteristics were given and the reasons for non-participation were explained in methodology.

18. Authors summarized the key results with reference to study objectives.

19. Authors should consider improving the discussion part. The results of accelerometer data were well discussed, but ICC and Cronbach's coefficient data was not well discussed.

The discussion about the missing comparisons was supplemented.

20. The last paragraph of the discussion clearly states why authors actually chose to adapt PAQ-A into Polish culture. Moving this paragraph to the introduction and restricting it accordingly would improve the rationale of the manuscript. Moreover, finishing the discussion with information related to PAQ-C causes a loss of emphasis on PAQ-A.

This part of the text has been moved for introduction.

21. Once again, authors state that 14-19 years old adolescents are in the Polish school system under the title of limitation. Then how did you invited adolescents aged 14-20 (possible fail the class, even if so please just indicate 14-19 years of age to refrain from misunderstanding.

22. Authors indicated that they used cut-points extensively used in pediatric research but there are no references of that extensive literature.

The above issues have been corrected, and the explanation is given above in point no 3 and no 14.



**VERSION 2 – REVIEW**

<b>REVIEWER</b>	Matteo Christian Sattler Institute of Sport Science, University of Graz, Austria
<b>REVIEW RETURNED</b>	08-Aug-2019

<b>GENERAL COMMENTS</b>	<p>General: The manuscript has now been substantially revised. The introduction and the motivation for this work is decisive. Also, the article focusses on the magnitude of agreement between questionnaire and accelerometer. The results are discussed with respect to recent evidence available in the field. Overall, this study adds valuable evidence regarding the validity and reliability of a polish physical activity questionnaire. However, minor points have to be addressed and clarified.</p> <p>Methods</p> <ol style="list-style-type: none"> <li>1) &gt;= 4012 counts ng per minute. Typing error.</li> <li>2) Accelerometer cut points. Please check again the cut points. 4012 should be the cut point for vigorous PA, not MVPA. 2296 = MVPA. Please check.</li> <li>3) ICC: The transparency would be higher if the authors mention the exact ICC version used in the R package. Was there a one-way or two-way random parameter? (Please see other publications)</li> <li>4) “The internal consistency was reduced after removing single items, which may indicate that no redundant items are included in the questionnaire” (p. 9): I would be careful by using this approach for evaluating redundant items. Please provide a reference of this approach.</li> <li>5) “Removing the first and second items had the lowest values in internal consistency, showing a high contribution to final test score.” Similar, please mention a reference that this approach provides accurate results. For me, it is a bit tricky to use just Cronbach Alpha since all the values are similar. Also, I cannot see the smallest Cronbach Alpha values for the first and second item? They are all above .90. The correlation between the item and the overall scale (item-total), which I think you mean in this case, is not the same as Cronbach Alpha. Please clarify the section in the manuscript.</li> </ol> <p>Results:</p> <ol style="list-style-type: none"> <li>1) There were no results regarding vigorous PA (above 4012 cpm). So, you may not need this information in the method section?</li> </ol> <p>Discussion</p> <ol style="list-style-type: none"> <li>1) Last two paragraphs. You may also consider to discuss the problem when using only a categorical score from a PA questionnaire. Although the results for the PAQ-A are clearly promising, it is difficult to use such a score to address dose-response patterns between PA and health (where information about duration, frequency and intensity is explicitly needed). So it is more about general levels and ranking of participants...</li> <li>2) There is also an updated review for PA questionnaires in children and adolescents: doi: 10.1007/s40279-018-0987-0. This may be useful for your discussion.</li> </ol> <p>Conclusion</p>
-------------------------	---

	<p>1) You may consider “This is the first study addressing” instead of analyses.</p> <p>Limitation</p> <p>1) Participants completed the PAQ-A at home. So, it might be difficult to control when they filled in the questionnaire. The questionnaire asks for PA during the last 7 days. So, for some participants this might be the period when they actually wore the accelerometer, where we usually see higher agreements. For others not. You may consider to mention this issue.</p> <p>2) Since you did not adjust for non-wear. You may provide a reference, that non-wear is more a problem for sedentary time rather than for MVPA.</p> <p>3) Please mention that some authors (also the reviews you cited) did not suggest calculating Cronbach alpha for this purpose. Thus, you can argue that, as you mentioned, you did this due to comparison reasons.</p>
--	--

<b>REVIEWER</b>	günay yıldizer Eskisehir Technical University, Turkey
<b>REVIEW RETURNED</b>	11-Aug-2019

<b>GENERAL COMMENTS</b>	All requested changes and amendments were done by the authors
-------------------------	---

## VERSION 2 – AUTHOR RESPONSE

### Methods

1)  $\geq 4012$  counts ng per minute. Typing error.

The error has been removed.

2) Accelerometer cut points. Please check again the cut points. 4012 should be the cut point for vigorous PA, not MVPA. 2296 = MVPA. Please check.

Thank you for comment. The error has been corrected.

3) ICC: The transparency would be higher if the authors mention the exact ICC version used in the R package. Was there a one-way or two-way random parameter? (Please see other publications).

We used a one-way random-effects. The information was added to *Statistical analysis* section.

4) “The internal consistency was reduced after removing single items, which may indicate that no redundant items are included in the questionnaire” (p. 9): I would be careful by using this approach for evaluating redundant items. Please provide a reference of this approach. 5) “Removing the first and second items had the lowest values in internal consistency, showing a high contribution to final test score.” Similar, please mention a reference that this approach provides accurate results. For me, it is a bit tricky to use just Cronbach Alpha since all the values are similar. Also, I cannot see the smallest Cronbach Alpha values for the first and second item? They are all above .90. The correlation between the item and the overall scale (item-total), which I think you mean in this case, is not the same as Cronbach Alpha. Please clarify the section in the manuscript.

Thank you for your valuable comment. The entire paragraph has been edited.

#### Results:

1) There were no results regarding vigorous PA (above 4012 cpm). So, you may not need this information in the method section?

Information about vigorous PA has been removed from method section.

#### Discussion

1) Last two paragraphs. You may also consider to discuss the problem when using only a categorical score from a PA questionnaire. Although the results for the PAQ-A are clearly promising, it is difficult to use such a score to address dose-response patterns between PA and health (where information about duration, frequency and intensity is explicitly needed). So it is more about general levels and ranking of participants...

2) There is also an updated review for PA questionnaires in children and adolescents: doi: 10.1007/s40279-018-0987-0. This may be useful for your discussion.

The discussion has been completed about the issue indicated by the reviewer. The reference proposed by the reviewer has been included.

#### Conclusion

1) You may consider "This is the first study addressing" instead of analyses.

The conclusion was corrected as suggested by the reviewer.

#### Limitation

1) Participants completed the PAQ-A at home. So, it might be difficult to control when they filled in the questionnaire. The questionnaire asks for PA during the last 7 days. So, for some participants this might be the period when they actually wore the accelerometer, where we usually see higher agreements. For others not. You may consider to mention this issue.

Participants were asked to fill the questionnaire after 7 days of wearing the accelerometer, so that the questionnaire replies refer to the week they wore accelerometry. Researchers paid special attention to this aspect of the study.

2) Since you did not adjust for non-wear. You may provide a reference, that non-wear is more a problem for sedentary time rather than for MVPA.

Thank you for suggestion. We added this information and references: Vanhelst J, Vidal F, Drumez E, et al. Comparison and validation of accelerometer wear time and non-wear time algorithms for assessing physical activity levels in children and adolescents. BMC Med Res Methodol. 2019;19(1):72. Published 2019 Apr 2. doi:10.1186/s12874-019-0712-1  
and

Winkler EA, Gardiner PA, Clark BK, Matthews CE, Owen N, Healy GN. Identifying sedentary time using automated estimates of accelerometer wear time. Br J Sports Med 2012;46:436–42.

3) Please mention that some authors (also the reviews you cited) did not suggest calculating Cronbach alpha for this purpose. Thus, you can argue that, as you mentioned, you did this due to comparison reasons.

Thank you for suggestion. This information has been added to *Limitation*.

### VERSION 3 – REVIEW

<b>REVIEWER</b>	Matteo Christian Sattler Institute of Sport Science, University of Graz, Graz, Austria
<b>REVIEW RETURNED</b>	30-Aug-2019

<b>GENERAL COMMENTS</b>	<p>The manuscript has been improved substantially. All points have been clarified except the following.</p> <p>The authors report that wear time significantly affects sedentary time but not MVPA time (p. 13 discussion) and provide 2 references (Drumez et al., Winkler et al.). However, these studies compared different algorithms and showed that those may more strongly affect sedentary time than PA time. These studies did not assess the influence of differences in wear time on sedentary or PA estimates. Several other studies (e.g., Herrmann et al. 2013; How many hours are enough? Accelerometer wear time may provide bias in daily activity estimates) reported an influence on both sedentary and PA estimates.</p> <p>Therefore, I recommend to further discuss and solve this statistical topic in the proposed publication. The problem was also discussed by Katapally et al. 2014 (Towards Uniform Accelerometry Analysis: A Standardization Methodology to Minimize Measurement Bias Due to Systematic Accelerometer Wear-Time Variation).</p> <p>Several approaches exist to handle differences in non wear. Some authors (e.g., also Herrmann et al.) calculate correlations between non wear time and SED/PA estimates. These should be reported. If there is a systematic relationship, some authors then consider to adjust for non wear time (or normalize data to 12 or 24h; similar to “as proportion of total wear time) in an additional analysis (sensitivity analysis). This may be done in addition to your results. Although I know that most validation studies do not consider influences of different wear times, I would recommend the following:</p> <ul style="list-style-type: none"> <li>- Update the two references provided in the discussion section since they do not substitute the statement</li> <li>- Evaluate and discuss the effect of non wear time</li> </ul>
-------------------------	--

### VERSION 3 – AUTHOR RESPONSE

“The manuscript has been improved substantially. All points have been clarified except the following.

The authors report that wear time significantly affects sedentary time but not MVPA time (p. 13 discussion) and provide 2 references (Drumez et al., Winkler et al.). However, these studies compared different algorithms and showed that those may more strongly affect sedentary time than PA time. These studies did not assess the influence of differences in wear time on sedentary or PA estimates. Several other studies (e.g., Herrmann et al. 2013; How many hours are enough? Accelerometer wear time may provide bias in daily activity estimates) reported an influence on both sedentary and PA estimates. Therefore, I recommend to further discuss and solve this statistical topic in the proposed publication. The problem was also discussed by Katapally et al. 2014 (Towards Uniform Accelerometry Analysis: A Standardization Methodology to Minimize Measurement Bias Due to Systematic Accelerometer Wear-Time Variation)”

Thank you for valuable your comments. We agree with the Reviewer. Therefore we removed the wrong sentence “(...) however other studies showed that wear time significantly affects the assessment of sedentary activity but not MVPA” and inappropriate references from the text.

In the article provided by the Reviewer, Herrmann et al. wrote that „wearing an accelerometer for too few hours per day can result in an underestimation of time spent in MVPA”. In our study, participants wore an accelerometer for 24 hours/day, which means that we used a procedure that allows the most accurate estimation of MVPA and the number of steps.

“Several approaches exist to handle differences in non wear. Some authors (e.g., also Herrmann et al.) calculate correlations between non wear time and SED/PA estimates. These should be reported. If there is a systematic relationship, some authors then consider to adjust for non wear time (or normalize data to 12 or 24h; similar to “as proportion of total wear time) in an additional analysis (sensitivity analysis). This may be done in addition to your results. Although I know that most validation studies do not consider influences of different wear times, I would recommend the following:

- Update the two references provided in the discussion section since they do not substitute the statement
- Evaluate and discuss the effect of non wear time”

Thank you for your suggestion and indication of the articles, the content of which has enabled us to properly rewording part of the Limitation and to prepare the answer for the Reviewer. After carefully reading the article (Herrmann SD, Barreira TV, Kang M, Ainsworth BE. How many hours are enough? Accelerometer wear time may provide bias in daily activity estimates. *J Phys Act Health* 2013;10:742-9), and based on the authors' suggestions, we came to the conclusion that it is unnecessary to carry out additional analyzes suggested by the Reviewer, because participants of our study wore an accelerometer for 24 hours/day.

Herrmann et al wrote: “*The majority of studies do not require participants to wear accelerometers for 24 hours, creating a problem for researchers to try to determine how many hours of wear time per day represents a valid day. Recent studies have reported a variety of criteria for selecting the number of hours per day one should wear an accelerometer to reflect daily free-living physical activity.*<sup>3,6,14,19,21,28</sup> *These studies have hours per day values ranging from as few as 6 h/day<sup>28</sup> and up to 16 h/day<sup>19</sup> to constitute a valid day. Sloopmaker et al<sup>19</sup> used an assumption that people sleep 8 h/day and therefore 16 hours was considered the upper limit allowed. Results from the 2004 accelerometer consensus meeting suggest using the 70/80 rule<sup>3</sup> for required daily wear time.<sup>26</sup> This rule provides a sample specific recommendation based on 70% of the sample having accelerometer data with at least 80% of those having at least the same amount of hours. (...)*

*If so, then further efforts are needed to improve the daily assessment of physical activity. One option is to place a greater emphasis educating the participants on the importance of wearing the device during all waking hours. Sharpe et al.<sup>18</sup> showed it is possible to have high compliance in an accelerometer study by providing detailed instructions to study participants. **Another alternative is to use the accelerometer for 24 hour surveillance.** The area of sleep medicine and sleep research has employed accelerometry with great success.<sup>17</sup> **Wearing accelerometers for 24 hour periods may provide richer data and help explain more complex relationships between sleep quality/quantity and physical activity and provide us with a more accurate measure of minutes of daily activity”.***

In our study, participants wore an accelerometer for 24 hours/day. We observed that a wear time over 80% of study participants was a minimum of 16 hours/day, that means that an accelerometer was worn during all waking hours. The mean non-wear time recorded by the accelerometers was 477 minutes per day (7 hours 57 minutes), which corresponds to the average number of hours of youth sleep. Study participants did not complete the sleeping diary, as demonstrated in the Limitation of the study. However, the length of sleep time of individual participants does not affect the MVPA or the number of daily steps calculation. In summary, it seems unreasonable to conduct an additional analysis, because we used a procedure that allows the most accurate measurement of MVPA and the number of steps/day.

#### VERSION 4 – REVIEW

<b>REVIEWER</b>	Matteo Christian Sattler Institute of Sport Science, University of Graz, Graz, Austria
<b>REVIEW RETURNED</b>	12-Oct-2019
<b>GENERAL COMMENTS</b>	Thank you. All comments have been clarified.