Author's response to reviews

Title: Validating the Physical Activity and Leisure Motivation Scale (PALMS)

Authors:

Keyvan Molanorouzi (keivannorozy@gmail.com)
Selina Khoo (selina@um.edu.my)
Tony Morris (tony.morris@vu.edu.au)

Version: 3
Date: 30 July 2014

Author's response to reviews: see over
Dear Editor,

Re: MS: 1340384153120128 “Validating the Physical Activity and Leisure Motivation Scale (PALMS)”

Thank you for inviting us to submit a rebuttal to the reviewers’ comments on the revised version of this manuscript. We fully understand the conventions associated with the journal publication process. The authors have substantial experience of this process. We accept that the Editor’s decision must be based on the comments made by expert reviewers, along with the “fit” of the paper with the journal. Sometimes that will result in rejection of manuscripts. We appreciate your perception that in this case there is the potential for a case to be made against rejection.

We consider that the “impasse” that has occurred relates to the major revisions proposed by Reviewer 2. We consider that we made appropriate changes to the manuscript in response to this reviewer’s comments on the first submission. We believe that the reviewer’s comments were justified in relation to the first manuscript. Although in his comments on our responses, the reviewer has signaled some recognition of the efforts we have made to address those concerns, we consider that his position on aspects of the four issues he continues to raise reflect professional differences on what constitutes validation, which is an ongoing process, that could be debated for a long time, and a misunderstanding of the role of levels of physical activity that we aim to clarify here.

In this rebuttal letter, we first present information on the context of the development of the PALMS that we consider to be helpful in understanding the decision to move on from validation of the previous measure, the REMM, but to use the REMM in the validation process for the PALMS. Then we address the four “major” issues raised by Reviewer 2 and sustained in his second review despite the amendments and additions we made to the manuscript.

The Recreational Exercise Motivation Scale (REMM) was developed in the course of a PhD some years ago. We consider that it was the first measure of motives for participation in physical activity to be developed through empirical processes and then placed in a theoretical framework. In the development and validation process, the REMM was examined in two studies, one focused on participants in non-competitive physical activities and the
other in a sample of competitive sport performers. During presentation of the REMM at a major international congress, problems associated with its length were discussed and we decided to create a short form that could be more readily applied in real world contexts.

We developed the Physical Activity and Leisure Motivation Scale (PALMS) to reduce the 73-item REMM to a more manageable size. At the same time, we produced a more balanced measure with an equal number of items on each of the eight subscales. We considered that five items per subscale was sufficient to ensure that subscales reflected respondents’ true response to the motive measured by that subscale, producing a 40-item measure. We propose that this process produced a sound structure for the short form and that the next step was to validate the new scale. We argue that it would not make sense to continue to examine the REMM in further studies of reliability and validity, when we propose that the PALMS is the measure that should be employed in future in most research and in all applied work. At the same time, we are confident that the REMM has been validated adequately to be recognized as a viable measure.

Thus, our focus has shifted to validation of the PALMS. In discussion with colleagues, we have established an international program in which researchers in 18 countries have agreed to examine the PALMS in a wide range of different languages, cultures and contexts. We expect multiple journal papers to emerge from each of those projects over the coming years. The first project to be completed involves the validation of the PALMS in its original English language form, among English-speaking physical activity participants in Malaysia. It is reported in the present paper. Thus, this paper represents the foundation publication, which will certainly be cited in many future journal papers, some of which are already in preparation.

This includes two further papers in Malaysia in which the construct validity of the PALMS is examined in large samples of physical activity participants. We considered the possibility of adding those studies to the present paper. We decided that this was not a viable proposition. We propose that, as the foundation paper, the present manuscript must present the theoretical background and with several reliability and validity analyses, it is around the maximum length for a single publication. Further, the two construct validity studies are based on different samples and each requires a substantial rationale to be presented to establish the basis for the different construct validation processes undertaken in each study.

We consider that the present manuscript represents a valuable contribution to the motivation for physical activity literature that is the foundation for substantial research and application that will enter the literature in the near future. The revised manuscript has been strengthened by the insightful comments of the reviewers for which we are grateful. We maintain that in that revision we have addressed the reviewers’ concerns appropriately (see the attached Authors’ response document). We now comment on the specific concerns of Reviewer 2 that do not seem to have recognized our responses to his concerns.
I reviewing the first submission of the manuscript, Reviewer 2 raised four concerns that he suggested required major compulsory revision of the manuscript. In each case, we made responses that included additions to the manuscript designed to address the issues raised, bearing in mind the restrictions imposed by the maximum standard length of the manuscript.

Reviewer # 2 (Pedro F. Saint-Maurice)

Major revisions:

1 – Again, refrain from the use of the term “criterion” validity. Firstly, criterion validity should not be defined based on the best measure available in your study. The term concurrent would be more appropriate. I’m still not convinced on the utility of the REMM nor have the authors fully demonstrated that this tool is valuable and it’s worth creating another version (short form). The authors argue that there is evidence to support the validity of this tool however, they don’t seem to consider the implications of correlated error obtained from validation designs based solely on self-report tools.

Secondly, the REMM was developed to facilitate the understanding of PA participation however, the one “validity” study that has been cited (reference 18) that refers to measures of PA does not include an appropriate assessment of PA nor used an appropriate validation design for this purpose. Please provide additional evidence that scores from this tool (REMM) are related with physical activity levels. I still see this as a major limitation and indicator of early stage development of the tool used as a reference (REMM) in this study.

• Given that the manuscript reporting the validation of the REMM in two Exploratory Factor Analysis studies is currently under revision for resubmission to a journal, we have cited the original report of the validation process involving these two studies in an unpublished PhD, which is accessible electronically in full text from the Victoria University, Melbourne library. We consider that the extensive description of the two validation studies in that freely accessible PhD addresses the appropriateness of examining the eight PALMS subscales against the corresponding subscales of the REMM. Thus, the present version of the manuscript does provide justification for a valid analysis of criterion validity against the best possible criterion, the longer version of the same measure (REMM). In addition, in the revised manuscript (see page 8) we cite use of the REMM in two studies in different countries and different languages that support the factorial invariance of the REMM. We consider that this is adequate background on the REMM because, soon after the cited work on the REMM was conducted, we considered that the 73-item scale was too long for practical use and embarked on the process of producing a short form, the PALMS. This approach has many precedents in the psychometric literature, as does criterion validation using the original, longer measure [such as the short form of the Profile of Mood States (POMS-SF) and the short form of the Flow State Scale-2 (FSS-2)]. The research now cited in the paper does provide adequate evidence of the validation of the REMM and the REMM is the most appropriate criterion with which to assess the convergent criterion validity of the PALMS. We consider that this response also addresses the second and third aspects raised by Reviewer 2 within point 1, because these are contingent upon the first aspect. That is, by demonstrating that material can
be accessed from several sources that thoroughly describe the validation process of the REMM, we contend that we have adequately addressed all three points raised by Reviewer 2 under this issue. The reviewer also raises the concern that the REMM has not been demonstrated to be a valid measure of motives for physical activity because no research has been conducted that related the REMM to level or amount of physical activity. We consider that this depends on researchers’ view of the most appropriate sequence for development of a measure. We contend that the decision to develop a shorter form of the measure of motives for participation in physical activity preempts the demonstration of an association between the longer measure and physical activity levels. We have now conducted a study of the relationship between the new measure the PALMS and physical activity, measured by the IPAQ an internationally recognized measure of amount of physical activity. That large sample study will be submitted very soon for publication, and it broadly supports the construct validity of the PALMS. That process should and does follow the basic validation of the short form in terms of factor analysis, reliability, and criterion validity, which is what we report in the present paper. For the purposes of that process, we maintain that the evidence we cite for the REMM indicates that it is the best criterion available.

2 –I understand that full evaluations of self-reports do not need to be provided in one single paper. However, the tone of the paper and terms used should be aligned with what the study is reporting. Again, the use of “validity” was not used appropriately and most likely, you should limit your discussion and results to indicators of reliability. You should also add these limitations to the discussion section.

- We acknowledge that the psychometric properties of the PALMS do need further examination. We have conducted studies in different countries and languages that demonstrate factorial invariance. Several of these studies have been reported at major international congresses, in particular the World Congress of Sport Psychology, 2013. Manuscripts describing these studies are in preparation for submission to journals. In our Malaysian research, we have completed discriminant function analysis for gender, age and type of physical activity in another large sample that demonstrates construct validity. That study has now been submitted to a journal. We have also examined the relationship between PALMS motive subscales and actual physical activity in a third study with another large sample. We did not consider it appropriate to include reports of those studies here because we consider that each of these issues needs to be presented in considerable detail to present the rationale, methods, analyses, and conclusions of each construct validity study appropriately. Adding them to the present paper would make this paper far too long to be acceptable for publication. We have now flagged in the revised manuscript that the other issues are in the process of being reported in separate papers. We further recognize that there is some variability in the use of the terms “validation” and “validity”. The reviewer has applied one interpretation and we have applied another. The solution proposed by the reviewer seems extreme to the authors. We consider that this
manuscript addresses more than the reliability of the PALMS. As noted in point 1, we provide further evidence to support the contention that correlation of PALMS subscales with the corresponding subscales of the validated REMM constitutes criterion validity. Further many researchers consider confirmatory factor analysis to be a form of validity. Thus, we propose that the term validation, which is commonly used to refer to examination of a combination of aspects of reliability and validity is used appropriately and that it is acceptable to refer to validity, but we have qualified that term to refer to “initial validity” in acknowledgement of the reviewer’s point that further tests of validity, especially construct validity, are necessary and we note that in the Discussion section under limitations.

3 – This information is useful however the author should include the operational definitions used to screen language issues. The authors would still need to provide objective measure of the English level of the participants involved in the study since English is still not their first and possibly, nor their second language. Perhaps the authors can report the English test score obtained from the initial screening.

The reader would still benefit from additional explanation of why an English version was preferred. As the authors agree, different countries require cultural and language adaptations and most of the development work done so far with PALMS was done in a country where English is their first language (Australia). This explanation should be added in the discussion along with a limitation of the study and also with possible implications for the results obtained.

- We have added the following statement in the Procedure section of the present manuscript on page 12 to address this issue:

- “Malaysia is a country in which the national language, Malay, is widely spoken, while several other languages associated with the large ethnic Indian and Chinese populations are also spoken by substantial numbers of people. Nonetheless, because of its British colonial heritage, English language education starts in primary school in Malaysia and a substantial proportion of the population from all ethnic backgrounds speak English well, even if it is not their “native” language. For this study, the PALMS and other measures were administered in English. To ensure that participants’ responses were based on sound understanding of the instructions, the items, and the response format, participants were screened for their capacity to read and comprehend English at a high level. We examined the questionnaire responses made by participants and did not identify indications in those responses that suggested lack of comprehension for the participants included in the analyses cited in this paper. Based on standard questionnaire checking processes, any participants whose responses showed signs of such response patterns were eliminated from the sample before the analyses were conducted. The number eliminated was small.”
We consider that this process and the explanation of it makes it clear that, as is the case for many Malaysians, participants included in this study did understand English well and experienced no problems with the requirements of the study.

The second author is a senior researcher at the University of Malaya. She has conducted many studies and has wired experience of research in the context of Malaysia. She advised that we employ the English language version of the measure based on her extensive use of questionnaires and other research methods in similar contexts.

We consider that the reviewer is far overstating the concern that participants did not adequately understand the questionnaires in this context in which with respect perhaps he does not have substantial experience. Is it possible that he has worked in other cultures where English as a second language is not spoken every day and with fluency and competence as it is by many Malaysians? We contend that the authors working in the Malaysian context are good judges of this. In addition, there was no evidence in our screening of the responses made to all the questionnaires that there was lack of comprehension on the part of the participants. Further the results support our contention showing patterns for all aspects that we examined that are consistent with previous research and with our predictions.

As a mark of respect to the reviewer, we have added the following comment to the limitations section (page 20, lines 16): “Acknowledging our statement in the Procedure section, it is recognized that the use of English language versions of the PALMS, REMM, and MCSDS questionnaires in the present study could be considered a limitation. We are confident from our screening of responses by the participants for all the questionnaires and the findings for all analyses, which are consistent with predictions, that participants included in the final sample clearly understood the content of the questionnaires and responded in a meaningful manner to the items in those measures.”

4 – Again, given the importance of PA in this study, the authors should include some descriptive information on PA levels (e.g., what was the percent of active participants in your sample?) and provide a copy of the self-report used. Why not use a validated PA scale? Can the reader trust the PA scores obtained? Also, please rephrase the statement: “To calculate regular physical activity, only the activities lasting at least 150 minutes of moderate- to vigorous-intensity PA per week were taken into account.” You probably want to indicate that individuals were defined as “physically active” if they accumulated 150 minutes or more of MVPA.

The reviewer’s suggestion is noted. We have added the following at the end of the subsection in which we describe the Demographics Form on page 10: “The items in regular PA were structured to provide separate domain-specific scores for walking, moderate intensity and vigorous intensity activity within each of the leisure time PA. To calculate regular physical activity, only the activities lasting at least 150 minutes
of moderate- to vigorous-intensity PA per week were taken into account. We are looking at PA in a later study.”

- Further in the Measure section on page 10, lines 12. We state: “In the present study participants were instructed to respond to the PALMS and the REMM with reference to their main physical activity. To provide information about what this activity was, we asked participants to state their main physical activity in the Demographic Form. We cite that information in describing the demographics of the sample. No further analysis in this study addresses type of physical activity. In a related study we have examined the relationship between motives and type of physical activity.”

Minor revisions:

1 – Add the implications in your discussion section. This was not done in the revised manuscript. Do you expect that your reliability scores will be affected by level of participation in PA? Most likely, yes.

- This information was added in the discussion on page 20, line 9: “Participants were recruited by direct invitation to people exercising in public environments. They were given an information statement and if they agreed to participate completion of the questionnaires was considered to indicate consent. Thus, this was a convenience, sample in which participants were accessed through local recreation facilities. One implication of this is that there was no systematic control over the gender, age, and regular physical activity patterns of the participants.”

2 – This explanation (perhaps more succinct) is helpful and should be added in the procedures section.

- This information was added in the procedures section on page 12, line 19: “Social desirability scales are all different, so there is no “index score”. The developers of each scale indicate the cut-off point or range of scores that reflect social desirability responding. There are few social desirability scales. The MCSDS is the most widely used measure. One way to use the MCSDS in practice or research is to exclude participants who score above the cut-off point. Another conventional way to apply the MCSDS is to use the scores from a large sample to check whether people respond to other questionnaires in socially desirable ways. This is done by correlating the MCSDS with the target scale. A strong correlation indicates a systematic pattern of responding to the target questionnaire that is consistent with the responses given to the MCSDS. For example, given that high scores on the MCSDS reflect that participants are responding in a socially desirable way, a positive correlation with high scores on the PALMS would indicate that participants are systematically responding in a socially desirable way on the PALMS. Conversely a low correlation between the
MCSDS and the PALMS would indicate that participants are not responding systematically to the PALMS in a socially desirable way.

3 – The author would still need to add confidence intervals to the correlations obtained. I’m familiar with the use of effect sizes and think these should not be used to replace confidence intervals.

- We have added the following comment in the manuscript to explain the use of effect sizes as opposed to confidence intervals in psychological research (Page 13, lines 21): “With reference to the correlations cited in this paper, it is not conventional in psychological research to cite confidence intervals. Instead, psychologists consider effect sizes as indicators of practically meaningful differences. Cohen (1988) indicated that Pearson's Product Moment Correlation Coefficient (r) is itself a measure of effect size. Values of r can be interpreted in the same way as Cohen's d, a common indicator of effect size. Thus, values of .2-.3 are considered to be small effect sizes, those around .5 are medium, and values of .7-.8 are viewed as large effect sizes. We have now interpreted the measures of association in terms of effect size.”
- We have reached an impasse here. It is our firm conviction that confidence intervals are not used in most psychological research of a correlational nature. Thus, we have not added them to this paper.

To reiterate, we acknowledge that the comments made by Reviewer 2 in their review of the first submission were fair. With the exception of further study to demonstrate factorial invariance and further construct validity, all four concerns appear to reflect information that we did not explicitly provide in the first submission. These were noteworthy omissions. To rectify this, in the revised submission, we provided information that addresses those concerns and we comment on it in our responses to the reviewer’s concerns. In our response and in the revised manuscript, we also explained that studies of factorial invariance are being conducted separately and publications are in preparation. Similarly, further evidence of construct validity has been examined in studies of the capacity of the PALMS motives to discriminate between participation in different types of physical activity in ways that can be predicted on the basis of theory and research, and in studies of the relationship between PALMS motives and actual physical activity participation. Again manuscripts are in preparation. In these cases, addition to the present manuscript was considered inappropriate because of the substantial nature of the information involved in these studies. We consider that the present manuscript includes aspects of reliability and validity that are conventionally considered to be necessary and sufficient for description of the initial development and validation of a questionnaire. With respect to factorial invariance and construct validity, we consider that the expectation stated by Reviewer 2 is excessive for one standard length manuscript.

In the present document we have done all we consider reasonable to address the elaborated concerns of reviewer 2. WE consider that we have clarified several of his concerns, but we must disagree with the reviewer with reference to the use of the terms validation and validity.
based on our interpretation of those concepts. We also consider that the reviewer has misunderstood the use of the demographic information on level of physical activity as we explain. Finally, we stand firm in our conviction that in our field of psychology correlations are conventionally cited without confidence intervals.

We trust that you will consider what we are confident is a sound and valuable contribution to the field of motivation for participation in physical activity that has great potential for use in research and practice in the future.

Thank you for your attention.

Yours sincerely,

Keyvan Molanorouzi  Selina Khoo  Tony Morris