**Revisor B:**

Dear Revisor,

We are thankful for your careful reading of our manuscript as well as for all your insightful comments and suggestions. Based on your guidelines we deeply reformulated our manuscript and we believe it is now richer in contents and based on a sounder methodology. **STRUCTURE OF THE MANUSCRIPT**

**•Title: Although the title describes the content of the manuscript it is
quite vague and lacks precision related to the aim and the content of the
study**

The manuscript has been renamed. We believe that the title provides a better insight of the manuscript’s contents and describes its goals more precisely.

 **•Abstract: The abstract can be improved to better reflect the remaining
manuscript. Specifically, the introduction does not drive to the main
objective and results of the study as there are no description about the
consequences of sleep disturbance, but rather the demonstration of the
feasibility and results related to sleep habits and disturbances. Moreover,
the sections results and specially conclusions may be improved.**

The abstract has been reformulated and we believe it now reflects the remaining manuscript more adequately. The introduction and conclusion sections have also been reformulated in order to provide information in a more logical sequence. The “results” section now includes additional results.

**•Introduction: Introduction describes well the interaction of sleep habits
with daytime events and cultural drives but could explain what motivated the
authors to explore sleep habits in those populations and better describe the
option for the Portuguese version of the Children’s Sleep Habits
Questionnaire. As the authors don’t explore daily or other consequences of
sleep disturbances, they should be more specific about sleep and sleep
habits and not so much about the consequences they didn’t explore in this
study, at least in their description.**

We chose the Portuguese version of the children’s sleep habits questionnaire (CSHQ) because both samples were collected on Portuguese-speaking African countries and hence all respondents from both samples speak Portuguese as their main language. These two populations were chosen for two reasons:

* Lack of scientific literature about sleep habits and sleep disturbances respecting to children living in the African continent
* Geographical convenience: we wanted to conduct the surveys ourselves and it happened that in Mozambique and Cape Verde did had the opportunity to do so.

Following your recommendation, we decided to restrain ourselves from describing the consequences of sleep disturbances that were not addressed in the manuscript. A more detailed description of the sleep habits that were studied was provided instead.

**•Methods: The study design and methodology are appropriate to its
objective. Nevertheless, the authors should describe better the method of
sampling of the population in each setting, which can originate implicit
problems when the authors compare results from populations that were not
normalized at the beginning.  This is perhaps the bigger methodological
problem from the study. A suggestion would be that the authors describe the
feasibility of the questionnaire and results from the study but refrain to
compare the populations. Mostly the authors must also describe the
confidence about the application of a Portuguese questionnaire in such
environments and how they can affirm that these populations were capable of
the right understanding and interpretation of the questions.**

In the revised version of the manuscript we provide a formal comparison between the two samples with respect to the variables of interest (table 1). By analyzing the results presented in this table we concluded that the two populations are not homogeneous with respect to the studied variables. Because of that, a separate analysis was conducted for each sample. Every inferential statement that involving comparing one sample to the other was removed from the manuscript.

The Portuguese version of the CSHQ has been adapted (and validated) to be used with Portuguese speaking respondents. Since all respondents from both our samples use Portuguese as their main language, we are confident about their capability of understanding and correctly interpreting the questionnaire. For the aforementioned reasons, we believe that the application of the CSHQ is adequate in the context of our study.

**•Results: Results must describe main differences between the populations
that answered and those that didn’t complete the questionnaires, in each
setting as this may influence the validity of the study. The results are
very interesting, but the authors could improve their presentation making
the description clearer and showing results in accordance with methodology.
There are some gaps between methodology and results. Comparisons between
populations implies that the authors can demonstrate that they are
comparable.**

As described above, in this version of the manuscript one separate analysis is presented for each sample. The “results” section has been reformulated in order for our findings to be presented in a more logical way. Also, this section now includes results from a regression analysis that was not conducted in the previous version of the manuscript but was insightfully suggested by another reviewer.

 **•Discussion: Mostly discussion explains the relevance of the results and
describes limitations. Suggestion goes to start by the most relevant
findings, then showing the applicability of the questionnaire in these
settings and further explain and contextualize the results in each setting
again refraining from comparing between populations except if they can prove
that these populations are comparable.**

Discussion was reformulated considering your suggestions. As already mentioned, the two populations were not compared in this version of the manuscript.

**•Conclusions: The conclusions must be reviewed**

Conclusions were reformulated considering your suggestions and the new results.

 **References: The literature review seems adequate, but References must be
reviewed as some are not presented correctly.
•Does it follow AMP’s style? The main objective of peer-review is to
ensure the accuracy of the manuscript and therefore reference should be
checked. Do the citations actually contain the information described in the
manuscript? Was any recent or relevant article omitted? Is the percentage of
recent references adequate?**

References were carefully reviewed, reformulated and formatted in order to follow AMP’s style. Some typos and inaccuracies were corrected. Some new references were included, and they are all mentioned in the main text.

 **•Tables / Figures: Tables and figures must be rearranged to reflect
changes in the presentation of results, but mostly they are well presented
and clear.**

The tables were reformulated in order to reflect changes in the manuscript. New tables were included in order to present the results from regression analysis. Tables respecting to comparisons between samples were removed except for the table that allowed us to compare that the two samples are not homogeneous with respect to the variables of interest.

 **EXTENSION: The extension of the manuscript could be smaller in the results
section.**

Results section was reformulated and shortened. This section is now more parsimonious, but it also includes new results that were not included in the previews version of the manuscript. This is because new analysis were conducted (like regression analysis). We believe this section is now richer in contents and more informative.

 **PRESENTATION: The manuscript is clearly and logically presented but
improvements already describe should be introduced. There is the need to
slightly improve the English**

The manuscript has been proofread and reviewed by a proficient English user. We believe the English of the manuscript is now improved, as well as its readability.

**Revisor C:**

**In "Sleep habits among children from Mozambique and Cape Verde", the authors**

**aim to describe the sleep habits of two pediatric populations from Africa,**

**namely from Mozambique (MZ) and Cape Verde (CV). The CV sample consisted of**

**206 children and the MZ sample consisted of 445 children. For each sample,**

**several demographic characteristics were studied (children’s gender,**

**children’s age, parents’ education level, number of cohabitants,**

**mother’s nationality, …) The sleep patterns were assessed by the**

**Portuguese version of the Children’s Sleep Habits Questionnaire (CSHQ) and**

**the authors aimed at evaluating the association between the observed scores**

**and the studied demographic variables. The questionnaire has 8 subscales**

**(not totally independent, as two items are common to two subscales) and 33**

**items.**

 **I disagree from the statistical approach followed by the authors. There is**

**a lot of information to be studied and the statistical methodologies have to**

**be carefully chosen as to control type-I error. For each of the 33 items, 8**

**subscales and for the full CSHQ score, the authors performed statistical**

**hypotheses tests, almost desperately…, looking for demographical**

**explanations for the differences in distributions observed for each item. It**

**is so much information, and with so many imprecisions at the statistical**

**analysis, that the message doesn’t go through. I thus reject the**

**present version of the paper and urge the authors to correct the**

**statistical analysis in a less ambitious way, due to the limited sample**

**sizes. I also would like to provide the authors with some statistical**

**analysis guidelines and specific remarks:**

Dear Revisor,

We are thankful for your careful reading of our manuscript as well as for all your insightful comments and suggestions. Based on your guidelines we deeply reformulated our manuscript and we believe it is now richer in contents and based on a sounder methodology. The main changes are:

* We provide formal statistical tests on the difference between the samples with respect to the variables of interest. By interpreting the results of those tests we concluded that the two samples are heterogeneous with respect to the variables of interest. Because of that, a separate analysis was conducted for each sample. Every inferential statement that involving comparing one sample to the other was removed from the manuscript.
* To avoid the multiple comparisons problem, we restricted inference on the items to a very restricted set of items (daytime nap and bedtime television). Those items were chosen for their prevalence and because there are prior motivations to believe they have an impact on sleep quality. No more inferences were conducted at the subscale level.
* In order to have a *ceteris paribus* interpretation of the estimated effect of demographical variables and of the two aforementioned items we conducted regression analysis with the logarithm of the CSHQ scores as the dependent variable and the demographic variables (and the two aforementioned items) as regressors.
* Some typos, imprecisions and ambiguities were corrected. The manuscript has been shortened and reorganized in order to present the contents in a more parsimonious and logical way.

 We now proceed to address your comments one by one:

 **a)   The demographic characteristics of the two samples (Table
1) provide statistical evidence to conclude that the two populations are
different. Why don’t the authors include adequate p-values on table I?
They will show that the two populations are very different with respect to
the studied demographic features (all but sex) and that substantiates the
two independent analysis (MZ vs CV) that they carry over along the paper.**

In the revised version of the manuscript we included the p-values you suggested. The p-values were obtained by conducting statistical tests of equality of proportions for each category of each demographic variable (across the two samples). As you suspected, only in the gender variable did we fail to reject all null hypotheses of equality proportions. This is evidence that two samples respect to different data generating processes and hence, as you suggested, this time we conducted two independent analysis (one for each sample). Except for this table, everything that involved conducting inference on the differences between the two samples was removed from the manuscript. The new demographics table also differs in the number of rows because the classes respecting to some variables were aggregated and/or different cutoff values were used (see next point).

**b)       Do the authors actually need 4 age classes and 3 parents’
education levels? For the latter variable, I can easily think about two
classes only (>12 years and otherwise in MZ; >10 years and otherwise in CV.
The cut-off has to be different as the two samples have very different
populations with this respect). The higher the number of classes you use
(thus, dummies), the lower the number of variables you will be able to use
in the joint/multiple analysis.**

In the revised table we now include the categories you suggested. Age categories were reduced from four to three (2 to 7 years old, 8 to 10 and 11 to 15). Parents’ education categories were reduced from three to two and, as you insightfully suggested, different cutoff values were used for the two samples: In the MZ sample the categories are now “<10 years” and “≥10 years” while for the CV samples the categories are “<12 years” and “≥12 years”. Please not that in the reviewed version of the manuscript it is only in table 1 (described above) that we use the same father and mother education level categories for both samples, as at that point we were still assessing about the homogeneity of the samples. It was only after interpreting the new table 1, with particular focus on p-values column, that we concluded that the two samples were different with respect to the variables of interest, and it was only from that point on that conducted two separate analysis with the new categories based on your suggested cutoffs.

**c)       The full CSHQ score is the easiest part to be studied. To
start with, it would be interesting for the reader to have a table with the
statistical description of the score on each sample (mean and standard
deviation, if the distribution seems to be reasonably symmetric, or median
and range otherwise) within each class of children’s gender, children’s
age, parents’ education level, number of cohabitants, …. That table can
have p-values, one for each variable, but they will only be informative. A
full regression model having the CSHQ score as dependent variables and all the remaining variables of interest as independent variables has then to be performed. At least for the MZ sample, the sample size seems to be enough to circumvent deviations from normality. For the CV sample, a similar analysis would be performed, for comparison reasons.**

In the revised version of the manuscript we included a new table (table 2) with the information you suggest in this point. We decided to present median and range instead of mean and standard deviation, as the distribution of CSHQ scores in both samples seem to be asymmetrical (skewness coefficients: 0.76 for the CV sample and 0.72 for the MZ sample). To have formal evidence of such asymmetry we conducted the bootstrap symmetry test from Miao, Gel, and Gastwirth (2006) for each sample. The null hypothesis of symmetry was rejected in both samples with p-values arbitrarily close to zero. In this table we summarize each sample’s CSHQ distribution by gender, age group, father and mother education levels, number of cohabitants and children cohabitants and nationality of the mother. All these variables are categorical, and the considered categories are as described in “a)” and hence incorporate your suggestions regarding aggregation of groups and cutoff values.

We also took your suggestion and performed regression analysis on both samples. For each sample we used a multiple linear regression model with the CSHQ scores as dependent variable and the demographic variables and other variables of interest (daytime nap and bedtime television) as regressors. In order to be possible to interpret the impact that changes in the regressor’s levels have on the CSHQ scores in a percentual scale, we applied a logarithmic transform to the CSHQ scores (for regression analysis only). All included regressors are in levels. In the MZ regression we obtained an r-squared of 0.15 and we found evidence of heteroscedasticity (Breush-Pagan test) or residual correlation (Durbin-Watson test). The model passed on the f-test of overall significance of slope coefficients with a p-value arbitrarily close to zero. In the CV regression we also obtained an r-squared close to 0.15 but we found evidence of residual correlation. Hence, to be able to conduct inference, we used Newey-West’s robust standard errors.

**d)       For the 8 subscales, I recommend the authors to apply a
MANOVA-like analysis or else correct the significance level of each multiple
comparison (eg, using Bonferroni correction). This will allow for the
simultaneous inspection of the effect of the variables of interest on the 8
subscales, at once.**

Inferences respecting to CSHQ items and subscales indeed required multiple comparisons to be performed. On the other hand, every comparison that was conducted was mentioned on the manuscript. In such a context, Bonferroni-like multiplicity adjustments would be useful to control the family-wise (type I) error rate at desired levels, although at the cost of reducing the statistical power of the tests.

 Nevertheless, we decided to exclude subscale analysis from the reviewed version manuscript because we were advised to shorten it. In what respects to items, very few were considered this time. Additional details on inferences with respect to items on the next point.

 **e)       I would avoid making inferences on the items, unless it was
being done on a very restricted group of items (the most important for the
analysis, only).**

We took your suggestion and only conducted inference on a very restricted set of items: daytime nap and bedtime television. Information regarding these two items is collected on the CSHQ but their scores are not accounted for in the total scale. We chose these two items for their prevalence and because there are prior motivations to believe that they have an impact on sleep quality. To make inference on these two items we included them in regression analysis together with the demographic variables. Note that daytime napping was included as a regressor only in the CV sample, since in the MZ sample the subgroup of children who “usually” took a daytime nap consisted in 25 children only.

 **f)        The authors have determined cut-off values for the CSHQ
score, from ROC curve analysis. They provide sensitivity and specificity
values accordingly and also give the percentage of children with their
parents reporting a sleep problem. These values allow me to compute the
percentage of children scoring greater than the cut-off value, for MZ and
CV. My calculations do not agree with the values presented by the authors. I
thus recommend the authors to review this matter.**

We carefully reviewed our calculations of prevalence (i.e. percentage of children with parent-reported sleep problems), sensitivity, specificity, optimal cutoff value and percentage of children testing positive. No mistakes were found. Since in ROC analysis we used only complete cases with respect to all items (324 observations for MZ and 200 for CV), and the “true” prevalence was computed using all complete cases with respect to the variable “has parent-reported sleep problem” (416 observations for MZ and 206 for CV), is it possible that the discrepancies that you found are due to the different sample sizes that were used to calculate the “true” prevalence and the “test prevalence”?

Nevertheless, we would like to highlight the fact that our results are 100% reproducible and we can provide self-contained R scripts that disclose every step of our calculations and can be run in any computer, returning the results that we present on this manuscript, as long as our dataset is provided as input.

 **g)        The numbers in the abstract describing the MZ sample do not
match those mentioned in the Results section.**

These inaccuracies are now corrected. We are thankful that you have spotted them.

 **h)       It is now clear from the text which correlation coefficient
the authors have used in the their analysis.**

We assume you that you meant “not clear” instead of “now clear”. Every correlation that we computed is a Spearman correlation. We apologize for this unintended ambiguity. **i)         Results from Table 4 have to be interpreted carefully as
the populations have already been shown (at least MZ and CV) to be
significantly different in many aspects. Once again, if the table and all
p-values are absolutely needed, at least correct the type-I error for the
multiple tests (use Bonferroni correction, for instance).**

Given the aforementioned finding that our samples are not homogeneous with respect to the demographic variables, we decided to refrain ourselves from making any inferences with respect to differences in items, durations or times between the two samples. The table that you mentioned here was therefore removed from the manuscript.