Response to Reviewer 1 Comments

The study aims at comparing two scales of general exertion (Borg RPE, Omni RPE) with percent heart rate reserve and one scale of local pain/exertion (Borg CR10) with the EMG mean power frequency of the trapezius muscle to evaluate the validity of these scales in US Spanish harvest workers. By estimating the validity of these scales in during harvest work of non-English speaking immigrants the study addresses a relevant issue.

Major points

1. The authors state that 24 individuals participated in the study. This is a fairly small sample size, especially when investigating three subgroups of different harvesting methods. I am doubtful if it is appropriate to use parametric statistics to analyze the data. I would suggest using non-parametric statistics (Spearman rank correlation, Friedman Test etc.).This would make the results more robust and increase their validity. As it is, the results seem somewhat questionable to me.

Response: Thank you for your suggestion. We agree with you that, with the sample size of 24, the parametric tests such as Pearson’s correlation may not be adequate. However, we transformed the data and conducted Shapiro tests prior to the analysis to ensure that the assumption of normality for parametric test is met. For instance, in the line 276-278, we indicated that “*The calculated % HRR was not normally distributed according to the p-value of 0.013 in the Shapiro-Wilk test for normality. After the % HRR was square-root transformed, the data became normally distributed, i.e., p-value of 0.48 in the normality test.*” And in the line 345-346 for the regression slope of the time variable on the EMG MPF that “*the Shapiro-Wilk test indicated that the parameter could be considered as normally distributed (p-value = 0.059).*”

Also, for the known nonparametric data of subjective scores, i.e. Borg RPE, Omni RPE and Borg CR10, we used nonparametric Kruskal-Wallis tests. We added a paragraph in the statistical analysis section to clarify that “*Moreover, these measures were evaluated for the effect by harvesting method and work period. The effects on the % HRR were tested using ANOVA and the effects on the Borg RPE, Omni RPE and Borg CR10 were tested using the Kruskal-Wallis tests*” in the line 255-257. Also, the post hoc analysis for pairwise comparison after the Kruskal-Wallis test was conducted using nonparametric Dunn’s test package in R as described in the line 263-266 that “*for the nonparametric tests to show the effect of harvesting method and measurement time point on the subjective measures, the ‘dunn.test’ package was used for post hoc pairwise comparisons using rank sums*”.

For the Spearman’s rank correlation that you suggested, we applied it to test for the relationship between the Borg RPE and Omni RPE. The detailed reply is under the point #17.

1. The number of individuals in the three subgroups (harvesting methods) should be stated. Or was the same group studied on three different days while engaged in different harvesting methods?

Response: Individuals in the three subgroups were different. We clarified in the line 88-89 that “*A total of 24 farmworkers participated in this study. The participants were equally divided into three groups, i.e. 8 different farmworkers per group, based on three different harvesting methods…*”

1. Line 110: Please discuss the use and validity of “percent heart rate reserve” in more detail. What does it measure? Is it really a proxy for physical exertion similar to oxygen uptake?

Response: In the line 114-117, we added that *“% HRR is a measure of the heart’s ability to recover to resting state after physical activities. % HRR was validated as a proxy for physical exertion in comparison to the oxygen uptake in laboratory settings for cardio exercises (Guimaraes et al., 2019) as well as during resting and sleeping (Matsuura et al. 2019)*.”

1. Line 115: As far as I understand, at least one assessment (T2) is carried out in the last 10 minutes of a 30 minute break. I am not sure if the mean power of the EMG during rest (rather than during muscular activity) is a valid measure for muscular fatigue.  Please argue this point (see for example Kimura M, Sato H, Ochi M, Hosoya S, Sadoyama T (2007) Electromyogram and perceived fatigue changes in the trapezius muscle during typewriting and recovery. Eur J Appl Physiol 100(1): 89-96.).

Response: The time period T2 was when the subjective questionnaires were asked and was right before the end of the 30-minute break. However, we did not use the mean power frequency (MPF) of the EMG to represent muscle fatigue directly. Instead, we used the regression slope of the time on the MPF, as shown in the equation (iv) line 231, to indicate the shift or difference in MPF before and after the second work shift.

1. Line 143: Please provide the exact phrasing of the Borg CR10 scale you used. Did you asked for exertion or pain?

Response: In the line 185-188, we revised the text to include the exact phrasing that “*Borg CR10 scale, asking the participants questions such as “how tired does your right shoulder feel”, was used as a subjective measure of local discomfort, particularly the tiredness (cansado in Spanish) they felt at each specific body part at the moment.*”

1. Line 146: You mention that the Borg should be calibrated. Did you calibrate the scale you used?

Response: We represented the Borg as an increase or a decrease rather than “normalizing” it with the maximum value. In the line 237-241, we clarified that “*Unlike the previous study recommending the calibration to the maximum value of the scale (Spielholz, 2006), the effort surveys including Borg RPE, Omni RPE and Borg CR10 scales at the specific time point were analyzed in terms of the increase or decrease as compared to the values at the beginning of the work shift (T0) to make the data interpretable*.”

1. Line 190: A study-design section should be included, describing among others the measurement time points.

Response: For the subjective measures, we included the time points for the subjective measures in the line 197-203. For the direct measures, we instrumented the equipment before the work shift and remove them after the shift. In the reply to your comment #8, we added a diagram in attempt to better describe the study design.

1. Liens 190-196: The measurement time point should be described more clearly, ideally also using a figure. This would make it easier for the reader. As it is, it is unclear whether the individuals were working or resting during the HR/EMG assessments. If they were resting, the body position should be indicated. Also, the exact time of the subjective assessments should be indicated.

Response: Thank you for your suggestion. We added the Figure 2 at the line 205-206.

1. Line 207: Please provide a reference for the equations you used.

Response: We added the reference Karvonen et al. 1957 in maximum heart rate equation in the line 217.

1. Line 214: The citation is incorrect.

Response: In the line 224-227, we rephrased the sentence to refer to the previous study, and not the unsupervised learning algorithms.

1. Line 214: Could you explain this algorithm briefly?

Response: In the line 224-227, we explained the algorithm that “*errors and artifacts in EMG data were diagnosed using principal component analysis of several parameters including the percentiles of EMG amplitudes, and mean and median power frequencies, and then removed as described in the previous study*”.

1. Line 281: You state that Omni RPE was higher for ground workers than platform workers. This is not apparent in the figure and may be a result of inappropriate use of parametric statistics.

Response: Thank you for pointing this out. We had used inappropriate Tukey HSD tests after the Kruskal-Wallis test. To correct this mistake, we used the Dunn’s test for multiple comparison using rank sum. The new results are reported in the line 295-296 for Borg RPE and line 302 for Omni RPE.

13.   Line 284: The figure is misplaced.

Response: Thank you. We realigned all the figures once the text was revised.

1. Line 304/307: The order of the figure numbering is incorrect.

Response: Thank you for noticing and informing the details of our mistakes. We revised the figures number as shown now in line 326 and 329.

1. Line 327: “decreased over time” instead of “reduced”

Response: The text in line 351 was revised that “*Muscle fatigue, i.e. the EMG MPF, decreased over time as shown by…*”

1. Line 354: The negative correlation is surprising. What may be the reason for this negative correlation? My suspicion is that the statistics and/or the sample size may not adequate (see point 1).

Response: We believe that the sample size of 24 is adequate, especially after checking the normality assumption. As the results came out different from the intended use of Borg scale, we interpreted that “The negative correlations between Borg RPE and % HRR suggested that the Borg RPE may not be useful as subjective measures for this population” in the line 379-380.

17.   Line 360: I am wondering if it would not also be interesting to report the correlation between the Borg and the Omni scales?

Response: Thank you for your suggestion. We conducted the Spearman’s rank test and indicated them in both the method and results section. In the line 252-254, we added that “*Additionally, between the two subjective measures of overall effort, i.e. Borg RPE and Omni RPE, Spearman’s correlations were used as the non-parametric tests for their relationship.*” Also, in the line 306-306, we found that “*Borg RPE and Omni RPE were found positively correlated with the Spearman’s correlation of 0.618*”.

18.   Lines 429-433: Please rephrase. As it is, it is unclear what you mean.

Response: In the line 456-458, we rephrased that “*part of the systematic bias during the data collection were not able to fully eliminated due to the nature of the fieldwork. That is, firstly, researchers were present and indirectly influencing how the farmworkers behave and, secondly, the quiet sitting period for the resting heart rate measurement was difficult to be ensured.*”