Author's response to reviews

Title: Full title: Learning aptitude, spatial orientation and cognitive flexibility tested in a virtual labyrinth after virtual stress induction

Authors:

marcel delahaye (marcel.delahaye@upkbs.ch)
patrick lemoine (patrick.lemoine@upkbs.ch)
ralph mager (ralph.mager@upkbs.ch)
johannes beck (johannes.beck@upkbs.ch)
gunnar deuring (gunnar.deuring@upkbs.ch)
henning hachtel (henning.hachtel@upkbs.ch)
marc graf (marc.graf@upkbs.ch)
marlon pflueger (marlon.pflueger@upkbs.ch)

Version: 4 Date: 15 May 2015

Author's response to reviews: see over
Dear editors, dear Prof. Capella, dear Prof. Graziano,
Thank you very much for your evaluation.

We put your comments in *italics* and the answers to your comments in **blue letters**.

**Reviewer's report**

*Title:* Full title: Learning aptitude, spatial orientation and cognitive flexibility tested in a virtual labyrinth after virtual stress induction  
*Version:* 2  
*Date:* 6 February 2015  
*Reviewer:* Emanuele Antonio Maria AM Cappella

**Reviewer's report:**

1) *Minor essential revisions.*

I think that this article needs a good english revision. There are in my view some sentences that do not sound natural in english. By way of example, line 17: “Farhadbeigia stated that it has been unsolved, to which degree different levels of stress impair accuracy in decisions”. I think it will be better to change the sentence in something like “F. stated that the degree to which different levels of stress impair accuracy in decisions is currently unclear.”.  
Another example from line 22 “... how researchers induced stress (therefore it is one of the aims of the study to create a standardized way for stress induction)”. I would suggest “...how researchers induced stress. Therefore, our study aims at providing a standardized way for stress induction”.  
I found even some grammatical inaccuracies:  
- line 9 pag 3: “is that both stress induction AND (not as well as) cognitive...”  
- line 20 pag 5: “Collectively” should be changed in “Taken together”.

The entire article has been corrected by a native English speaker who works as a psychiatrist.

2) *Discretionary revisions*  
- **Table 1: Standard Deviations should be reported.**  
  Standard Deviations have been included.  
- **Descriptives: you should describe more thoroughly your participants.**  
  Education/Employment? Ethnicity? You should also report the standard deviation value for Age.

A description of the participants has been included.

- **Limitations: I don’t see this section...**

Research limitations and future research needs have been added.

3) *Major compulsory revisions*  
- The major problem is the reliability of the items measuring stress. This is very important because all your analysis (and consequently your conclusions) are based only on this measure.
Why did you choose three simple questions instead of one already validated instrument? It is of course possible to create a new stress index, but in this case its reliability should be explored and reported. It is therefore of paramount importance to report at least the Cronbach's Alpha to show the consistency of the items used. Another solution would be to repeat the study taking into account reliability with new participants/ a new stress measure.

Validity. We don't know anything about the convergent and discriminant validity of the instrument used to test the hypotheses. As for content validity I personally don't think that your items capture the complexity of the stress construct. I think it is important to specify in the article which model led you in the construction of the items, preferably with some literature references.

We utilized these three Items as they have a high face validity. We are aware that these items only measure individual aspects of stress.

We recognize, in retrospect, that the lack of use of a validated instrument is a weakness of the study. We thought that the combination of physiologic data in combination with the three questions would be sufficient and nearly equivalent to an instrument. We have included the articles that describe the model (Lazarus) which lead to the construction of the three questions we used to evaluate the participants level of stress.

Without taking into account reliability and validity we cannot draw any certain conclusion. In fact, you assume that your version of TSST induced stress (hypothesis 1) only on the basis of the results provided by your items - which as I said are not psychometrically sound. I think that a possible solution in this case would be to add some physiological data (cortisol, heart rate...). On the one hand this would support your first conclusion; on the other hand such addition would allow to evaluate the validity of your stress index. By the way, I found some studies that provide such measures to support their conclusions about the same virtual reality test (a slightly different versions of the Trier Social Stress Test, see Jonsson et al. 2010).

The same line of reasoning applies to hypothesis 2 and 3: how can you test the effect of stress induced by TSST on spatial orientation and cognitive flexibility if you don't show that you are adequately measuring stress with your items (validity) and that they are reasonably consistent (reliability)? To sum it up, I think that there is a serious issue of psychometric soundness, which needs to be addressed.

The authors agree with the reviewer's comment. In the original study, the heart rate had been recorded, but not calculated and reported. In the current revised version, it has been integrated along with other physiologic data.

Prof. Graziano

Reviewer's report:
Discretionary Revisions

Every part of the experiment is clear and detailed but statistical analysis is poor. I suggest to implement “Methods” in particular Statistical Analysis.
A chapter titled “Methods” has been added.

*It's not enough to say which program do you use. In particular, in this case I recommend to specify what type of analyses was used, why do you stop enrollment at 31 volunteers (write something about power analysis). Do you adjust analysis for sex or age?*

We attempted to recruit as many participants as possible during a ten week period. During that time, we had 43 people express interest in participating in the study. The twelve people were excluded either because they did not show up for the study, or they changed their mind and decided not to continue in the study. One left due to low motivation and two stopped due to Simulator Sickness. Because our analysis was explorative, power analysis was not focused upon.