

No clear or consistent evidence that wearing an eye mask leads to meaningful improvement in learning and alertness: A reanalysis of Greco et al. (2023)

Stephen Rhodes*

Last updated: 2023-04-04

Greco et al.^[1], in the “Statement of significance” section, claim that their findings suggest that “[...] wearing an eye mask during sleep is an effective, economical, and noninvasive behavior that could benefit cognitive function and lead to measurable impacts on everyday life.” Here we examine some of the analysis choices made in this article and critically assess this claim. Firstly, Greco et al. should be applauded for clearly stating their analysis model and for publicly sharing their data (<https://osf.io/q4p9v/>), which made this reanalysis possible.

In Experiment 1, paired associates learning (PAL) was assessed on day 6 and 7 of the manipulation (eye mask or control), following 5 nights of habituation. On day 6 participants learned words paired to particular cue words up to a 60% correct criterion. The first test of PAL was then 10 minutes later with the number of correctly recalled words (out of 80) being the outcome. The second PAL assessment occurred on day 7 with no additional learning opportunities in between. The two PAL assessments were split up and analyzed separately via the following mixed effects model (using syntax from the `lme4` package for R):

```
lmer(y ~ eye_mask + (1 | ID) + (1 | year), data = data, REML = FALSE)
```

where `y` is the outcome of interest, `year` was a factor with two levels representing changes in the nature of the control condition from 2018 to 2019, and `eye_mask` refers to whether or not participants had worn an eye mask (1 = yes, 0 = no) in the 5 or 6 nights prior to testing.

Using this model there was a significant difference in the average number of words recalled between mask and control on day 6. However, these conditions did not differ in terms of an ‘absolute consolidation’ score, which was the difference in performance between days 6 and 7 (reported in the supplement).¹

It is not clear why the day 6 and 7 results were separated, as performance on the PAL task could have been assessed via a single model. Indeed a single model analysis is preferable if the difference between immediate and delayed recall is of interest. Splitting the data to assess a ‘mask effect’ obviously inflates the type I error rate. A more appropriate model would have also allowed for an interaction between `year` and `eye_mask` as changing the nature of the control condition may have influenced the *difference* between the control and eye mask conditions, which is encoded in the fixed effect for `eye_mask`. Including `year` as a random effect only allows for overall performance to vary by `year` of study. Further, as `year` only has two levels, it seems inappropriate to try to estimate a variance component for this factor and, as we found when recreating the analysis, there were situations where this caused fit issues in the reported analysis (the ‘singular fit’ warning from `lme4`).

As participants were assessed under both the control and eye mask conditions, assessing the PAL data as a whole allows one to model between participant variability in the size of the eye mask effect (i.e., a random slope term for `eye_mask`). Not accounting for this source of variability in the model is important

*steverho89@gmail.com

¹It is interesting to note that a similar overnight change analysis is presented for the motor skill learning task in the supplement but the main article reports an analysis in which the interaction of condition and day was tested for this outcome.

as the standard errors for the fixed effect of `eye_mask` will be too small, leading to incorrect p-values, and, potentially, incorrect decisions as to whether to reject the null hypothesis of no difference.^[2]

We reanalyzed the PAL data with the following model, which allows for an interaction between condition (`eye_mask`) and year, to account for the different control conditions, and an interaction between condition and day, which codes whether the assessment was on day 6 (immediate) or day 7 (delayed).²

```
lmer(y ~ eye_mask + year + day + eye_mask:year + eye_mask:day +  
      (1 + eye_mask + day | ID),  
      data = pal1, REML = FALSE)
```

In this analysis the coefficient for `eye_mask` is not significantly different from zero ($b = 0.99$ [-0.01, 2.00], $p = 0.053$).³ Further, the interactions between condition and year or day are not significant. There is also no significant eye mask vs control condition main effect in a mixed models ANOVA, which, unlike the above model, includes all two way and the three way interactions ($F(1, 81) = 3.77$, $p = 0.056$).

We also reanalyzed average reaction times from the psychomotor vigilance test (PVT) via a mixed model that allowed for a fixed effect interaction between year and condition, as well as a participant-level random intercept and slope for the difference between eye mask and control. The difference between conditions is significantly different from zero, albeit with greater uncertainty in the magnitude of the difference, $b = -6.239$ [-12.363, -0.115], $p = 0.047$.

For the data from Experiment 2, it is not possible to estimate a random participant slope for eye mask as there was only one score per condition. Greco et al. also note that this experiment was not powered to detect differences in PAL or PVT scores, as the goal was to assess correlations with electrophysiological measures (see supplement to article). Nevertheless, we recreated the results for both tasks but found that the number of unique participant IDs in the data files was lower ($N = 28$) than that reported in the article ($N = 33$ following exclusions).

So what are we to make of the effects of eye masks on cognition? When reassessing the data presented with more appropriate models we find less consistent evidence for cognitive effects of wearing an eye mask. However, this reanalysis should also be interpreted with caution. Information on condition order was not available in the data files and is potentially important given that some participants were omitted from analysis, thereby possibly leading to an imbalanced design. Further, the participant-level data from the PVT are average reaction times from an unspecified number of trials where responses could be missing due to overly short or long response times. Trial level data would allow one to take into account these additional sources of variation/uncertainty.

Even if we were to take these improved estimates at face value it is difficult to imagine how being able to, on average, remember 0 to 2 (the 95% CI around the `eye_mask` effect) extra words out of 80 or being able to respond 0.1 to 12 milliseconds faster after wearing an eye mask could lead to measurable impacts on everyday life.

(1180 words)

Disclosures

None

²We have set `REML=FALSE` to fit the model via maximum likelihood, rather than restricted maximum likelihood (REML). This is to match Greco et al., who used ML to allow for a likelihood ratio test. However, `REML=TRUE` would likely lead to better estimates of fixed effect standard errors.

³`year` and `day` were ‘sum-to-zero’ coded so that the `eye_mask` fixed effect can be directly interpreted as an ‘overall’ eye mask effect.

References

- [1] Greco, V., Bergamo, D., Cuoccio, P., Konkoly, K. R., Muñoz Lombardo, K., & Lewis, P. A. (2023). Wearing an eye mask during overnight sleep improves episodic learning and alertness. *Sleep*, 46(3), zsac305.
- [2] Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255-278.