



# Does running pleasure result from finding it easier than you thought you would?

A recommendation by **Zoltan Dienes**  based on peer reviews by **Jasmin Hutchinson** and 1 anonymous reviewer of the STAGE 1 REPORT:

Damien Brevers, Guillaume Martinent, Irem Tugce Oz, Olivier Desmedt, Bas de Geus (2023) Do error predictions of perceived exertion inform the level of running pleasure? OSF, ver. 4, peer-reviewed and recommended by Peer Community in Registered Reports (TEST). <https://osf.io/y8d9m>

Submitted: 08 September 2023, Recommended: 15 September 2023

## Cite this recommendation as:

Dienes, Z. (2023) Does running pleasure result from finding it easier than you thought you would?. *Peer Community in Registered Reports (TEST)*, . <https://test.peercommunityin.org/RR3/articles/rec?id=453>

Published: 15 September 2023

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

The reward value of a stimulus is based on an error in prediction: Things going better than predicted. Could this learning principle, often tested on short acting stimuli, also apply to a long lasting episode, like going for a run? Could how rewarding a run is be based on the run going better than predicted? Understanding the conditions under which exercise is pleasurable could of course be relevant to tempting people to do more of it! Brevers et al. (2023) will ask people before a daily run to predict the amount of perceived exertion they will experience; then just after the run, to rate the retrospective amount of perceived exertion actually experienced. The difference between the two ratings is the prediction error. Participants will also rate their remembered pleasure in running and the authors will investigate whether running pleasure depends on prediction error. The study plan was refined across four rounds of review, with input from two external reviewers and the recommender, after which it was judged to satisfy the Stage 1 criteria for in-principle acceptance (IPA). **URL to the preregistered Stage 1 protocol:** <https://osf.io/xh724>

**Level of bias control achieved:** Level 6. *No part of the data or evidence that will be used to answer the research question yet exists and no part will be generated until after IPA.* **List of eligible PCI RR-friendly journals:**

- [Communications in Kinesiology](#)
- [In&Vertebrates](#)
- [Peer Community Journal](#)

- [PeerJ](#)
- [Royal Society Open Science](#)
- [Studia Psychologica](#)
- [Swiss Psychology Open](#)

#### **References:**

1. Brevers, D., Martinent, G., Oz, I. T., Desmedt, O. & de Geus, B. (2023). Do error predictions of perceived exertion inform the level of running pleasure? In principle acceptance of Version 5 by Peer Community in Registered Reports. <https://osf.io/xh724>

## **Reviews**

### **Evaluation round #4**

#### **Authors' reply**

- [Download author's reply](#)
- [Download tracked changes file](#)

### **Evaluation round #3**

DOI or URL of the preprint: <https://osf.io/s72bt>  
Version of the preprint: 2

#### **Authors' reply, 02 September 2023**

- [Download author's reply](#)
- [Download tracked changes file](#)

### **Decision by [Zoltan Dienes](#) , posted 07 August 2023, validated 07 August 2023**

#### **Minor Revision**

Thank you for your revisions. I am not sure if you fully updated the version of the manuscript I see in the osf link. You still need to address the point of specifying your IV more clearly: State explicitly in the results section and the design table that the IV will be absolute prediction error. State explicitly in both places that you will use model 2. And for power, you still need to justify a roughly smallest effect of interest. I am not familiar with the pre-existing power programs for lme, and it may be easiest to simulate yourselves: generate data from a H1 model with a fixed raw slope (e.g.) a 1000 times, run the lme, and determine the proportion of significant effects. This way you can estimate error variance from your pilot, fix the raw slope at a just interesting value, and vary N. Maybe hiring a graduate statistics student for a few hours would enable this to get done.

### **Evaluation round #2**

DOI or URL of the preprint: <https://osf.io/s72bt>  
Version of the preprint: 2

## Authors' reply, 06 August 2023

[Download author's reply](#)

## Decision by **Zoltan Dienes** , posted 20 July 2023, validated 21 July 2023

### Revision invited

Thank you for your revisions, which have addressed many of the reviewers' points. One reviewer asks for some further clarifications. I would like to go back to the points I made.

1) Only report those analyses in a Stage 1 that you will base inferences on. That is, if you plan to base inferences only on model 2, do not even report the possibility of model 3 at this stage. If you want to leave open that you will base conclusions on model 3, then indicate precisely how you will decide which model the conclusions will be based on. Incidentally as you plan to keep all variables in the equation, you don't need different steps. Just enter all variables simultaneously. So if you are just going to use model 2, specify only the final equation with all variables in, and then base your conclusions on that. Note that pre-registering a specific analysis now does not stop you exploring other analyses later in Stage 2 when the data are in. But these other analyses will be in a separate non-pre-registered results section, and the abstract will draw conclusions only from the pre-registered analyses, and the discussion will keep pre-registered analyses centre-stage.

2) You could treat your two IVs, absolute and relative prediction error likewise. Choose one now, and keep the other for exploration later. Just bear in mind your main conclusion in the abstract and discussion will be based on the pre-registered analysis. If you keep both in the pre-registration you need to decide how you will base conclusions on different possible patterns of results. The multiple testing problem I mentioned before remains. So if you keep both, use a multiple testing correction, e.g. Bonferroni. If you use Bonferroni, the decision rule can be if either or both slopes are significant, then there is an effect of prediction error on pleasure. You also need to calculate power with respect to the alpha determined by the multiple testing correct, e.g. .025 for Bonferroni. Now more on power. Power is to control the risk of missing out on an effect of interest. Thus, it must be calculated with respect to the smallest effect just of interest. This is of course hard to determine, but you mention the rough heuristic of using the smallest available estimated effect. Based on one pilot, this heuristic no longer captures the spirit of what is needed. I find thinking in terms of raw effect sizes more helpful than standardized ones. You usefully provide raw regression slopes from the pilot. No one can judge whether an  $R^2$  of 0.40 is interesting whereas one of 0.3 is getting boring. Whether a one unit change on your raw scale of absolute prediction error predicts 0.15 units of pleasure on your scale is interesting at least feels vaguely judgeable. (Presumably the slope mentioned next, 0.57, is for relative prediction error, so I think there is a typo here.) I presume you found this interesting, even though it is small, because it motivated you to plan this study. But it does feel on the edge. How about setting an interesting effect just a bit lower than this, say 0.1 pleasure units/abs PE unit, and calculating the power for this? I think this would work best by referring to the raw slopes in the power calculation section. As I mentioned earlier, a 80% CI on the slope, and using the bottom limit, would be a touch more objective (even if 80% is relatively arbitrary), because it is a procedure that could be repeatedly used in many cases.

best

Zoltan Dienes

## Reviewed by anonymous reviewer 1, 12 July 2023

The authors have addressed my previous comments. I have no further comment.

[Download the review](#)

## Evaluation round #1

DOI or URL of the preprint: <https://osf.io/at62v>

Version of the preprint: 1

### Authors' reply, 29 June 2023

[Download author's reply](#)

[Download tracked changes file](#)

### Decision by **Zoltan Dienes** , posted 08 June 2023, validated 08 June 2023

#### Revision

I have gone with the two reviews I have in, following your email, to expedite the process. The reviewers are largely positive, but have a number of points to clarify. I have a couple of points as well:

1) In a Registered Report, one ties down all analytic and inferential flexibility. While in the pilot models 2 and 3 fare best, in the main study this may or may not be true. Be absolutely explicit about the criteria you will use to end up with the one model you will draw inferences from. Make sure anyone reading your planned analyses and having your raw data for the main study would end up making the same decisions. Further, you have two DVs- there is room for inferential flexibility here (what if one DV shows one thing and the other another?), as well as a familywise error rate problem. Given your pilot, I would suggest picking one of the DVs - probably absolute prediction error. Otherwise you need to correct for familywise error and specify your decision rule depending on the different pattern of possible results.

2) The function of calculating power is to control the error rate of missing out on interesting effects. But that means one needs to calculate power with respect to roughly the smallest effect that you do not want to miss out on. The PCI RR guidelines for authors puts it "power analysis should be based on the lowest available or meaningful estimate of the effect size". As one reviewer points out, you actually use an effect size estimate for power larger than that obtained in the pilot. Yet presumably an effect considerably smaller than found in the pilot would still be interesting and one you would not want to miss out on. The issue is discussed in some detail here: <https://doi.org/10.1525/collabra.28202> I would suggest finding the 80% CI for each effect in the pilot that is in your Design Table, and using the bottom limit of the interval in each case as the effect you use for calculating power. But you might have other ideas, after the reading the linked paper, for how to address the point.

Finally, you phrase the issue as predicting pleasure from prediction error, but prediction error is not completely assessed until pleasure is. A causal diagram of what is going on could of course take many forms. No need to do anything about this now, but I presume the discussion will touch on this point - all in good time.

best

Zoltan

[Download the review](#)

### Reviewed by anonymous reviewer 1, 18 May 2023

The authors present a stage 1 manuscript to investigate how error predictions of perceived exertion are related to feelings of pleasure. I commend the authors on their willingness to do a registered report, detailed methods, and their sharing of data and code. I view this as an opportunity to try and help the authors improve their future study (planned to start in October, 2023).

I have a few major comments and some minor comments.

Major comments:

1. In several places in the manuscript, the authors refer to "experienced level of running pleasure" (e.g., abstract, main text). This is an important note because it is one of the primary variables. However, the authors are not measuring experienced pleasure. They are planning on measuring retrospective ratings of pleasure, with measurement taking place after the running session. This would be more appropriately referred to as remembered pleasure than experienced pleasure.

It is also important to clearly identify this as remembered pleasure or remembered affect. This is retrospective, and retrospective evaluations do not perfectly align with moment-to-moment experienced pleasure.

2. In the methods, when "Running pleasure" is introduced as a variable, the authors should describe the model of affect that they are adopting. If they conceptualize pleasure-displeasure as bipolar (which I would suggest, see Russell, 1980), then they should allow for the measurement of displeasure. Currently, they only allow for no pleasure or extreme pleasure, and do not allow runners to report levels of displeasure. I strongly encourage the authors to adopt a bipolar measure that allows for the measurement of displeasure.
3. On page 4, paragraph 1, the authors could also discuss the findings of Rhodes & Kates (2015). <https://doi.org/10.1007/s12160-015-9704-5>
4. Given the importance of affective responses experienced while exercising, why not measure RPE and affective valence during exercise too? Why not also measure anticipated affect and remembered affect as well? Is this not possible, technically?
5. The authors say "Importantly, this can explain why some people find their physical exercise unpleasant...". I think I understand what the authors are trying to convey, but the link between pleasure and perceived exertion (from the prior sentence) itself does not seem to explain the affective rebound. They seem like separate concepts. In other words, people seem to experience displeasure during exercise followed by an increase in pleasure after exercise, but I am not sure that this is explained by the fact that perceived exertion and pleasure seem to be negatively associated.
6. The authors mention the safety of the SET. While true, this sample answered no to every question on the PAR-Q+, and should be able to safely do a maximal test. Therefore I'm not sure safety is a good justification here.

#### Minor Comments

1. The authors mention that they will report the intention-to-treat analysis. While useful, I encourage the authors to also report a per-protocol analysis. Intention-to-treat is great, but both could be maximally informative especially if dropout seems high.
2. There are some instances of RPE referring to rating of physical exertion, but in most cases it is rating of perceived exertion. Please be consistent.

3. In the description of the relative index of RPE prediction error, it seems that the parenthetical suggests the text should read "subtracting the score of retrospective RPE from prospective RPE". Please clarify.
4. In the design table, I encourage the authors to also report the interpretation if their hypotheses are not supported.
5. On page 4, there is an extra period after the Hartman reference.
6. The comma after "A key tenet from the literature on reward processing," can be removed.
7. On page 7, I think "fell" should be "feel".
8. On page 7, there is an extra comma in "In addition, to the weekly coaching sessions".
9. I think "technics" should be "techniques" (also page 7).
10. On page 7, "greater sense of autonomy toward physical exercise, *but* also increased...". I think but can be "and".

Thank you for allowing me to review this project. I think it has promise, and I hope that my comments are helpful. I especially encourage the authors to strongly consider their conceptualization and measurement of affect, and whether they are interested in experienced pleasure, remembered pleasure, or both (I encourage both).