Thank you for the opportunity to review this work. The proposed study addresses an important question and has the potential to contribute to the literature on exercise-related affect. The research question is scientifically justifiable, based upon existing evidence and current literature and the hypothesis is logical and plausible. Overall, this is a strong manuscript, however I do have some concerns and suggestions that I would encourage the authors to consider. These are listed below. Good luck with this research!

**General comments**

1. There are a few grammatical errors throughout – I pointed out a few examples in the specific comments below, but the final paper will need to be proofread / edited carefully
2. **Insufficient detail to enable replication**. The start-to-run program is not clearly described. I understand that these is some flexibility (based on the SET) but there is no description of the program at all. I’m also curious why there is only one group session per week. If these are beginning runners more frequent meetings might help with adherence? Another consideration, if the “free run” days are entirely self-selected (“allowing participants to choose the duration, frequency, and intensity of each run”, p. 8), how will you ensure progression toward the 5-mile goal?
3. Related to the above point, the authors state that “participants will be able to choose how they want to follow the proposed training program (e.g., to strictly follow it, to use it as a basis for training, or choose not to follow it)” (pg. 8). I am not sure that this is the best approach. There is no standardization of the intervention in this case. Specific to the variable of interest, forecasted RPE, this could be problematic. it is possible that if one’s forecasted RPE is high that they might decide to not follow the program and do something easier. Or they may make an adjustment during the workout that affects the prediction error. Page 7 indicates that runners can decide to “walk when they feel the need to do so” – I understand this from a safety and adherence perspective, and for the other reasons given in the paper (autonomy, enjoyment), but I am concerned that the difference between forecasted and retrospective RPE could be affected by this. For example, let’s say today, according to the app, I am supposed to run for 4 miles; my forecasted RPE is 7. I start running and decide I don’t feel like doing 4 miles today so I only do 3 miles. Or I complete 4 miles but I walk for half the distance. Now, my retrospective RPE is 5. Does this mean I had a prediction error of 2? Or does it reflect that I decided not to follow the recommended workout?
4. Some other methodological questions I have are… Is the program held on the same indoor track as the jog test, or somewhere else? In the weekly group sessions, how many people will run at a time? Is there a group warm-up and/or cool down? What is the role of the coaches, do they run with the participants? Are participants allowed to listen to music while they run? This will likely influence RPE.
5. Has the Formyfit app been validated? What is the algorithm used by the app to design the running program? Does the app take anything else into consideration other than the results of the submax VO2 jog test? I tried to check the link provided ([www.formyfit.com](http://www.formyfit.com)) but got an error message indicating “access is forbidden”. However, I checked the Google app store, and (assuming I located the correct app) there are some details that are not reflected in the method – for example “each training session is accompanied by voice coaching to guide and motivate you”. This seems like an important point to discuss.

**Specific Comments**

Abstract, Line 9: Delete “on” before “the level of pleasure”

Abstract, Line 17: Replace “advance” with “indicate”, or you could use “advance the idea that”

Abstract, Line 18: Delete “better” before “identifying”

Page 3, Line 3: Sentence beginning “’It allows to effectively prepare…” is unclear.

Page 3, Lines 4-5: I would change this to “including health behaviors”

Page 3, Lines 5 - 6: Bit of a jump here to describing this process as memory based. Needs an intervening sentence to explain.

Page 3, Line 8: If this is ‘extensively studied’ is there a review article or meta-analysis that could be cited here instead of a single study?

Page 3, Line 16: The phrasing “this current gap of knowledge stems from the adoption of rating of perceived exertion…” doesn’t really make sense, need to rephrase.

Page 3, Paragraph 2: Just a note to be very clear with language here. The terms effort and exertion are often used interchangeably when they refer to different things (see e.g., Hutchinson 2021, <https://doi.org/10.51224/B1013>, Smirmaul 2012, <https://doi.org/10.1136/bjsm.2010.071407> or Swart et al. <https://doi.org/10.1136/bjsports-2011-090337>. In other words, asking “what intensity of effort do you feel now” is not the same as asking for a rating of perceived exertion. In addition (and adding to the confusion), Foster’s Session RPE measure uses the verbal prompt ‘How was your workout?’.

Page 3 Lines 21-25: This is true, but the vast majority of this literature comes from well-trained athletes – is there anything from an exercise population?

Page 4: Good section about the link between RPE and pleasure. Where you discuss prospective RPE (Lines 24 & 25), I wonder if this could this be linked to forecasted pleasure as well? See e.g., Hutchinson et al. 2023 <https://doi.org/10.1123/jsep.2022-0243>.

The described mismatch between anticipated and experienced effort is also in line with research on forecasting error (see Ruby et al., 2011, <https://psycnet.apa.org/doi/10.1037/a0021859>). It might be useful to briefly make this point here?

Page 6, Section 2.2: How will participants be recruited?

Page 6, Section 2.2: In the interest of equity, would you consider using the additional questions from the PAR-Q+ as a secondary screening, or requesting a physician referral rather then excluding participants completely?

Page 6, Section 2.2: If it is feasible, a VO2 max test would be preferable to the submaximal estimation – this might enable you to identify HR associated with ventilatory threshold which is theoretically important in this area of study.

Page 7 Line 2: A VO2 max test is considered safe in this population. Also, I don’t see the George et al. citation in the references.

Page 7 Line 9: Referring to “coaches” – what specific training will they have? PE graduate students are not the same thing as running coaches, so I assume some training would be needed.

Page 7, Line 13: Replace “fell” with “feel”

Page 7, Lines 13-14: Referring to the “general advice about running techniques, nutrition and sport injury prevention”, how will this information be standardized?

Page 7, Line 14: Replace “technics” with “techniques”

Page 7 Line 20: Specify what RPE literature you are referring to here. If it is the work cited at the end of the next sentence, I would not describe this as RPE literature.

Page 8 Line 10: Need more detail on this app.

Page 9, Last Line: Referring to “relative difference will inform on the magnitude…” – Doesn’t absolute difference also give this? I’m not sure I understand this?

Page 10, Section 2.4.2: Why not use the Feeling Scale? The proposed measure is positively oriented (i.e., 6 of the 7 options indicate some degree of experienced pleasure).

Page 10, Section 2.5:1: Needs detail – how are speed and distance recorded? With GPS? If GPS, how will indoor runs or treadmill runs be recorded?

Page 10, Section 2.5:2: Regarding group impact. This detail addresses one of my earlier questions, however, will this also be accounted for in the group sessions? The text suggests it is only for the free sessions.

Page 10, Section 2.5:2: Regarding Habits. How will “degree of habits” be quantified?

Also, if I understand, degree of habits refers to familiarity with the route. In this case should the modulation not be the other way round? Whether the degree of habit modulates the impact of RPE prediction error on running pleasure, rather than, “whether the impact of RPE prediction error on running pleasure modulates the degree of habits linked to the running program” (p. 10).

Page 11, Figure 1B: Referring to “retrospective RPE” – This measure is not the same as Borg’s CR-10 scale. For example, the CR-10 scale has and an ‘absolute maximum’ and decimal numbers (e.g., 0.5 and 1.5). Also, I don’t see the verbal anchors for either scale? (Bi and Bii)

Page 12: Will the model be adjusted for the described covariates (running distance, running speed, group impact, habits)? I did not see this described.

Page 14: Did you gather any feasibility information from the pilot study? What was the retention rate? Useability of the app, etc.?

Page 14, Line 11: Replace “linear mixed model” with “LMM”.

Page 14, Line 17: I am a bit confused by the reference to the variance in individual empowerment?

Table 1 and 2. Were the indicated covariates included in these models?

Page 15: I’m unable to comment extensively on these analyses as this is outside of my area of expertise.

Page 16, section 2.8: If I understand correctly, the pilot data indicated effect sizes of .42 and .41? Then why not use .40 (rather than .50) for the estimation?

The estimation suggests that 16 participants will need to run a minimum of 5 sessions? If so, it might be clearer to write this as “…16 participants along five measurement points (i.e., five running sessions) is required …”. At first, I thought it meant five measurement points within each run.

Design Table: Perhaps it was cut off, but I do not see the last column (“theory that could be shown wrong by the outcomes”) in the table.