

Author's response: Dear Reviewers and Recommender, please accept our sincerest thanks for all your invested time and useful feedback.

Both reviewers mentioned the challenge in following multiple URLs for different pilot studies. I want to add that the appendices seem to be numbered differently in the manuscript and in different components on OSF, which may add to this confusion. Furthermore, the URL for Appendix 3 is linked to a component on OSF rather than the appendix itself, and the URL for Appendix 6 is linked to Appendix 5. Please make sure that the Appendices are numbered consistently, and correct URLs are provided to each appendix throughout the manuscript. You may also combine all appendices into one document and share that on OSF too, as suggested by Reviewer 1.

Author's response: We apologize for the lack of clarity in listing the appendices. We have combined all appendices into one document, available at the same OSF link (<https://osf.io/fxhjd>), which should make searching and viewing attachments easier and faster.

For Pilot 1, "The five most important variables in MOBA games (League of Legends (LoL) and DotA 2) were strong will, attention, speed of decision-making, good teammates, resilience, and self-confidence and in FPS games (Counter-Strike: Global Offensive (CSGO), Tom Clancy's Rainbow Six: Siege, and Overwatch) the five most important were attention, speed of decision-making, good teammates, resilience, self-confidence, and persistence." Six rather than five variables are listed here. For MOBA games, "self-confidence" is ranked number 7 in Table 1 of Appendix 2. For FPS games, "attention" is ranked number 9 in Table 2 of Appendix 2.

Author's response: We apologize for this mistake. Both sentences have been shortened to list exactly 5 factors and these now correspond to those listed in the appendix.

2A. Whether the data are able to test the authors' proposed hypotheses (or answer the proposed research question) by passing the approved outcome-neutral criteria, such as absence of floor and ceiling effects or success of positive controls or other quality checks.

The description of outcome-neutral control results could be made clearer. If I understood it correctly, participants first provided their highest rank (in the past 12 months, and ever) at the beginning of the survey. At the end of the survey, they were asked their highest rank ever again, using icons instead of text for LoL, and with the order reversed for CSGO and Fortnite. The check then shows there is a high correlation between the two 'highest rank ever' responses. I think such a more detailed description will make it easier to understand what the correlations in outcome-neutral control exactly mean.

Author's response: It was exactly as you have described. Thank you for your suggestion. We have updated the description of correlations in the following form: "*Participants provided their ranking at the beginning of the survey (measurement 1 of DV) and also at the end of the survey (measurement 2 of DV). In the second measurement of the DV, icons were used instead of text for LoL, and reversed order of ranks for CSGO and Fortnite. High correlations between measurement 1 and measurement 2 (CSGO, $r = .99$ and LoL, $r = .96$) supported their reliability.*"

Please check whether multicollinearity occurred in the regressions. Providing zero-order correlations between IVs and DVs would be informative - as already proposed during the review of the Stage 1 version. These results can be presented in the exploratory analyses

section, to make it clear that they are not pre-registered.

Author's response: An expanded answer to the multicollinearity problem is provided by Dr. Bonny's comment. Spoiler: we found no evidence for multicollinearity.

There is some concern over the validity and reliability of the practice questionnaire. It would be useful to provide information on the psychometric properties of the practice questionnaire (also for the future use of the questionnaire), as already proposed during the review of the Stage 1 version. Again, these results can be presented in the exploratory analyses section. This is something you may also wish to discuss in the Discussion section of your Stage 2 manuscript.

Author's response: Responses to Dr. Bonny's concerns regarding the validity of our data (not the practice measure) are stated below.

As to the evidence of validity of the practice measure, we have calculated several correlations between practice and constructs to which practice should be theoretically connected within the nomological network (see Dr. Bonny's comment below). Moreover, the validity evidence of our practice measure is primarily based on content validity - on the use of subject matter experts - SMEs (top players) both during the initial stage of creating the items and when checking their appropriateness. We have also provided estimates of reliability - the McDonald Omega total coefficient, which works similar to factor analysis. The omega coefficient is an estimate of the general factor saturation of a measure. Omega coefficients calculated from all practice and deliberate practice items would be more than sufficient: Practice (CSGO, .80; LoL, .85), deliberate practice (CSGO, .76; LoL, .79). We add, however, that such psychometric validations are not optimal since both practice constructs do not conceptually align well with the reflective measurement approach - there should be no common cause and indicators do not have to correlate with each other (e.g., someone practicing alone does not necessarily also practice with a coach). We consider this situation better assessed by a formative approach to measurement where individual indicators build a construct - representing different aspects of the construct based on theoretical underpinnings. From this point of view, most of the traditional statistical procedures used to assess validity cannot be used and therefore we have opted for convergent validity evidence.

I want to emphasize that these results will be exploratory in nature and should not change the main conclusions. Reviewer 2 has suggested more discussion of naive practice, which you may do in the Discussion (but not in the introduction, to avoid post-hoc changing the study aim). Related, I recommend not changing the discussion to focus more on naive practice and less on deliberate practice, otherwise there may be risk of over-emphasizing positive results at the expense of negative results, which we want to avoid. Achieving a good balance of the results (both positive and negative) in the Discussion section is key here.

Author's response: We agree and have tried to find optimal balance.

2B. Whether the introduction, rationale and stated hypotheses (where applicable) are the same as the approved Stage 1 submission.

Yes.

2C. Whether the authors adhered precisely to the registered study procedures.

There are two deviations from the registered analyses, namely (1) the operationalization of decision-making and (2) the new exclusion criterion (practice time > 168 hours per week). Although both are well-justified, it is nevertheless crucial to precisely adhere to the registered

procedure. For deviation (1), I think using the total number of correct trials and the percentage of correct trials should give the same results, assuming the total number of trials is the same across participants. For registered results, please use the total number of correct trials, as originally planned. You may add a note that this was a mistake during Stage 1, and that switching to the percentage of correct trials does not change the results (if that is indeed the case).

Similarly for deviation (2), please report the results with these participants included in the manuscript, as originally planned. You can then also report the unregistered results in which this post hoc exclusion criterion is adopted, but this should be very transparently flagged as unregistered.

Author's response: The change in operationalization of decision-making (from the total number to percentage) concerned only exploratory analyses. The pre-registered confirmatory operationalization (the percentage of successful stops in nogo trials) has not been changed. As you have pointed out, this did not affect the results but made the data processing easier as the results were provided in this format as part of the Psytoolkit output.

We have modified the results to correspond to the preregistered plan and therefore excluded the new exclusion criterion. We have transparently described analyses with this exclusion as exploratory and included their full results within the supplementary file dedicated to exploratory findings.

2D. Where applicable, whether any unregistered exploratory analyses are justified, methodologically sound, and informative.

The exploratory analyses are in line with what has already been proposed in Stage 1 RR. Providing a link to the exploratory results on Fortnite seems okay to me.

2E. Whether the authors' conclusions are justified given the evidence.

The discussion and conclusions are justified based on the current results.

Other comments:

The Results in the Abstract could be made clearer. For example, "in both esports, we found evidence for deliberate practice not having a meaningful effect on performance. On the other hand, the results confirmed younger age predicting better performance in both games." Could you add effect sizes to these results, and clarify what you mean by "a meaningful effect on performance"?

Table 3: Please explain what ω_{total} means in the table note.

Author's response: Both pieces of information were added to the abstract, including the explanation of the Omega total coefficient.

Reviews

Reviewed by Maciej Behnke, 18 Oct 2023 13:19

The authors have done an excellent job conducting the study and preparing the Stage 2 manuscript. I enjoyed reading it and did not find any weaknesses. The only thing I would

reconsider is the structure of the supplementary materials. I think it would be easier to navigate in one document rather than switching between appendices.

Author's response: We apologize for the lack of clarity in listing the appendices. We have combined all appendices into one document, available at the same OSF link (<https://osf.io/fxhjd>), which should make searching and viewing attachments easier and faster.

Reviewed by Justin Bonny, 07 Nov 2023 13:49

Overall

I appreciate that the authors have a wide-reaching set of experiments that this manuscript draws upon. However, it has become difficult to keep track of which results / methods were motivated by which pilot experiment and how they all tie into the current manuscript. I urge the authors to present all of the relevant findings and prior work, succinctly described, in the main body of the manuscript. Having to follow URLs to each pilot experiment to try and understand what is happening in the manuscript has been challenging; I may have missed some of key information in those attached repositories when writing this review.

Author's response: We have tried to present all the necessary but at the same time not as much detailed information about Pilot studies in the Introduction. However, after the IPA we are not allowed to change the structure of the Introduction. We agree that the solution of several self-standing appendices was unfortunate and led to potential confusion - since both reviewers have drawn attention to this aspect. We have replaced all individual OSF links with a single link leading to a combined document with all appendices. We believe this edit will help increase the clarity of the study rationale.

Naïve Practice

The authors need to define, discuss, and place naïve practice in juxtaposition with deliberate practice within the skill development theoretical framework more thoroughly. Much of the manuscript hinges on how these two concepts relate to each other and (may) be differentially related to skilled performance, both short- and long- term. The authors do discuss deliberate practice on pg. 4, but state, "We return to these conceptual differences later." but do not do so before the hypotheses are presented. The authors need to more concretely place naïve practice within the theoretical framework of expertise development because practice is crucial to their hypotheses. Furthermore, in discussing the pilot experiment where their measure was evaluated, the authors' discussion of "naïve practice" needs greater clarification. The items in the pilot study include physical conditioning as well as playing esports without the intent of improving skills. These seem quite different, which is alluded by the authors indicating that the study used for the manuscript dropped the physical conditioning items (but yet they are still included in the table). Altogether it is hard to tell what exactly "naïve practice" refers to in the manuscript and how it relates to existing literature in skill development research.

In addition, the authors need to better frame their results with naïve practice within the existing literature. They provide a table of results from prior esports research regarding the association between (presumably) naïve practice and skilled performance in the introduction. But they do little to refer to these results in the discussion. The authors should provide a greater description as to how their results align or deviate from these prior studies.

Author's response: As noted earlier, and as the recommender states, we should not

change the Stage 1 text but we agree that further clarity on the relation between naive and deliberate practice in the discussion definitely improves the manuscript. We have thus added further discussion on this relation also with reference to additional earlier literature.

Do the Results Disconfirm Deliberate Practice Theory in esports?

The authors argue in the discussion that their results do not support deliberate practice theory: "Based on the present study, deliberate practice is not a meaningful predictor of long-term success in esports" (pg. 19); "This study adds falsifying evidence for the applicability of deliberate practice theory to esports" (pg. 20). But was their study really providing evidence of this? I am hard pressed to think of a research article that argued that deliberate practice has no association with skilled performance. Most of the recent ones that have questioned deliberate practice have framed their hypothesis about deliberate practice having a weaker association with skilled performance, but not zero association (e.g., Macnamara et al., 2016, Hambrick et al. 2020).

Author's response: This is a very good comment thanks to which we can further explain the difference between weaker, meaningful, and zero association.

First, we stress that in the recent literature it is made explicit that "the claim that practice largely accounts for differences in performance across people (i.e. inter-individual differences) even at elite levels of performance is controversial" (Macnamara & Maitra 2019, citing Gobet et al 2014, Hambrick et al 2014, Macnamara et al 2014). In this context, it is hardly surprising that we do not find meaningful effect sizes.

Nowadays, in the (post-)replication crisis period, scientists are encouraged to move beyond significant / non-significant dichotomy when discussing their results. Instead of a vague assessment of effect sizes based on Cohen's recommendations (weak, medium, strong), researchers should adopt a strategy to assess practical significance, i.e., the practical meaning of their results. This is what we did in our study, specifically in "Appendix 5: The rationale for selecting SESOI $r = .3$ (CSGO) and SESOI $r = .2$ (LoL)". We have used the unit of "one rank" as the smallest *visible indicator of change* in one's long-term esports success and defined our SESOI by the *standardized effect that would lead to a change of rank* in real life. We regarded a change of (just) one rank as already large enough to be of practical significance. Before the preregistration, we have also discussed much smaller effect sizes (e.g., $r = .1$), but in each scenario (game title, variable), we realized that these effect sizes lacked practical significance - a disproportionately large amount of effort would be required to achieve a change just by one rank.

We would like to stress that instead of "no association" we have preregistered and subsequently found "no meaningful association". By this term, we note that a relationship between the variables may exist, but from our perspective, it is not practically meaningful. After a literature search, we have been able to identify 2 longitudinal studies (8 and 52 weeks long) carried out on this topic and both did not find a significant relationship between performance in CSGO (measured either as a Kill/Death ratio or tournament success) and practice (Pluss et al., 2021; 2022). At the same time, our data also point to the existence of a relationship between deliberate practice and ranking (zero order correlation was approx. .35). This zero-order correlation is roughly in line with what previous meta-analyses have found (meta-analytic estimates for different fields were .38, .40, .45, .56, .61).

In our study we have followed the recommendations of cited Hambrick et al. (2020), specifically the one to assess a wide range of potentially relevant causal factors on a sample

of a wide range of performance, and most important, to "estimate the relative contributions of the factors to the prediction of individual differences in expertise".

Thus, from a strong zero-order relationship, after controlling for other factors that explained the variance of esports performance better, remained only a not practically meaningful (but not zero) relationship.

We agree with the reviewer, that, from a common sense perspective (and also our own prior beliefs), there should be a meaningful association. However, the data speaks otherwise and more research is needed.

I would encourage the authors to elaborate further on the alternative hypotheses they present, namely, that the measures of deliberate practice were low in construct validity, that the participants were esports players but not experts. The authors do echo these concerns, such as defining deliberate practice when implemented in a research study (e.g., Hambrick et al., 2020) and that professional esports players were unlikely to have been included in their sample. However, the authors have the dataset at hand to at least start investigating these alternative hypotheses using exploratory / post hoc analyses. For example, if the questionnaire was indeed measuring something about deliberate practice, the authors should use their dataset to provide some tangible statistical evidence that this was indeed the case. Without making additional use of their datasets, the manuscript, as it stands, is inconclusive about the relation between deliberate practice and skilled performance in esports.

Author's response: We have added further limitations to highlight especially the issue with construct validity. Regarding the statistical evidence for deliberate practice, due to the construct validity issues in the original literature by Ericsson et al. (e.g., some definitions involve only teacher-created training and others additionally include self-created training), it would be problematic to claim whether the construct we have measured 'accurately' represents the 'real' deliberate practice construct (because Ericsson et al have proposed multiple 'real' constructs). We thus would prefer accepting the current, preregistered measure as such (which includes both self- and teacher -driven forms of practice) and leave further criticism and development for future studies.

PCI RR Criteria

2A. Whether the data are able to test the authors' proposed hypotheses (or answer the proposed research question) by passing the approved outcome-neutral criteria, such as absence of floor and ceiling effects or success of positive controls or other quality checks.

I understand that the pilot experiments were used to estimate the effect sizes for each game title, but it still seems unbalanced to have different significance test criteria for LoL and CSGO. This makes it harder to observe significant effects with CSGO, compared to LoL. Conceptually, this assumes that the effect of practice on long-term performance is contingent on esports title, but that assumption was framed as a motivation for the present study.

I would argue the bigger concern here with the regressions is multicollinearity. Naïve practice and deliberate practice are likely to be strongly correlated (I would be concerned if these two measures were not correlated, given what the authors presented about deliberate practice theory); this may also be the case for intelligence, reaction time, and attention. If the robust regressions are sufficient to address multicollinearity, then this should be mentioned; if not, this needs to be addressed. But I would suggest that the authors at the least provide zero-order correlations between all predictors and DVs for the reader.

Author's response: The robust regression we used is sufficient to handle non-normality and outliers but not multicollinearity. As you have suggested, we have included a correlation matrix with zero-order correlations within the Results section. Here we also provide a collinearity diagnostic (VIF coefficients) and correlograms (for easier orientation) that do not point to a multicollinearity problem.

Both practice constructs and cognitive constructs among themselves showed moderate correlations. From our point of view, this is in no way surprising, considering that each of the above-mentioned constructs is conceptually different (even naive practice and deliberate practice refer to activities of a different nature, as is defined by the original developers of the constructs) but at the same time they are to some extent related (both practice constructs refer to the practice, intelligence, attention, and reaction time are aspects of cognitive functions).

If we observe strong correlations (e.g., $r \geq 0.7$ and between two constructs such as naive and deliberate practice) this would raise questions about the validity of considering them as two separate constructs and whether they truly represent distinct aspects of a particular phenomenon.

VIF as measures of multicollinearity (CSGO)

Model1

Low Correlation

Term	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
scale(practice)	2.18	[1.79, 2.77]	1.48	0.46	[0.36, 0.56]
scale(deliberate_practice)	2.24	[1.83, 2.85]	1.50	0.45	[0.35, 0.54]
scale(intelligence_total)	1.09	[1.01, 1.58]	1.04	0.92	[0.63, 0.99]
scale(persistence_total)	1.08	[1.01, 1.64]	1.04	0.93	[0.61, 0.99]
scale(age)	1.12	[1.03, 1.52]	1.06	0.89	[0.66, 0.97]

Model2

Low Correlation

Term	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
scale(decision_making_speed)	1.03	[1.00, 5.38]	1.01	0.97	[0.19, 1.00]
scale(attention)	1.29	[1.13, 1.63]	1.13	0.78	[0.62, 0.88]
scale(reaction_time)	1.23	[1.09, 1.56]	1.11	0.81	[0.64, 0.91]
scale(team_work_total)	1.28	[1.12, 1.61]	1.13	0.78	[0.62, 0.89]
scale(intelligence_total)	1.08	[1.01, 1.60]	1.04	0.92	[0.62, 0.99]
scale(persistence_total)	1.31	[1.14, 1.64]	1.14	0.77	[0.61, 0.87]

VIF as measures of multicollinearity (LoL)

Model1

Low Correlation

Term	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
scale(practice)	1.86	[1.63, 2.18]	1.37	0.54	[0.46, 0.61]
scale(deliberate_practice)	1.91	[1.67, 2.23]	1.38	0.52	[0.45, 0.60]
scale(intelligence_total)	1.12	[1.04, 1.33]	1.06	0.89	[0.75, 0.96]
scale(persistence_total)	1.08	[1.02, 1.34]	1.04	0.93	[0.75, 0.98]
scale(age)	1.09	[1.02, 1.33]	1.04	0.92	[0.75, 0.98]

Model2

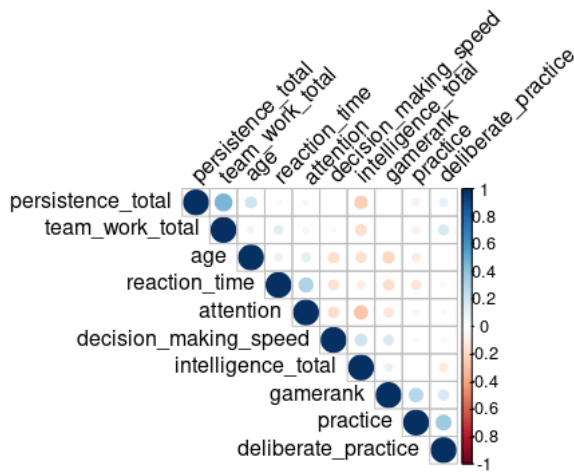
Low Correlation

	Term	VIF	VIF 95% CI	Increased SE	Tolerance	Tolerance 95% CI
	scale(decision_making_speed)	1.07	[1.02, 1.34]	1.04	0.93	[0.74, 0.98]
	scale(attention)	1.25	[1.14, 1.45]	1.12	0.80	[0.69, 0.88]
	scale(reaction_time)	1.18	[1.09, 1.38]	1.09	0.85	[0.72, 0.92]
	scale(team_work_total)	1.30	[1.17, 1.50]	1.14	0.77	[0.67, 0.85]
	scale(intelligence_total)	1.17	[1.08, 1.37]	1.08	0.86	[0.73, 0.93]
	scale(persistence_total)	1.31	[1.18, 1.52]	1.14	0.76	[0.66, 0.84]

Correlogram (CSGO)



Correlogram (LoL)



The outcome-neutral control is not sufficient as a quality check of the dataset. There needs to be more evidence presented in the main manuscript that the deliberate and naïve practice measures are valid and reliable for assessing practice using the datasets presented in the manuscript. There should be some analyses that can present evidence that they worked as intended in the present study. The authors should consider including additional checks to test assumptions that are based on prior literature and conceptual similarity such as: correlation between career length and age, reaction time and intelligence, etc. These types of analyses would provide further evidence that the dataset was valid and adequate for the present manuscript.

Author's response: We have ensured the validity of our data by applying several procedures, precisely following the preregistered plan: 1) Based on the bogus item, instructed response item, and Mahalanobis distance statistic, we have removed careless participants, 2) by applying outcome-neutral control, we have ensured that our responses are not randomly generated or significantly biased. It would be extremely unlikely for participants to provide the same response to both O-N control items if not paying attention to the survey and providing biased or untrue responses. The Back button and the possibility to correct responses were removed from the survey, the first O-N control item was presented at the very beginning and the second O-N control item as the last and in a different form (reversed order or in the form of images).

We have tried to provide more evidence that our data are valid by correlating some variables that should be theoretically at least to some extent related (career length and game rank = participants who play longer should have better rank; items that ask about similar things - naive practice item 1 = Physical practice and physical training) or to check for the existence of differences where they should exist (e.g., younger participants should have better attention and reaction time than older participants, although we expect no dramatic difference given the limited age range). See the results below:

career length and game rank - 0.31 (CSGO), 0.41 (LoL)

game rank and game rank highest ever - 0.91 (CSGO), 0.90 (LoL)

naive practice (Item 1 - Physical practice (gym, running, etc.)?) and physical training (How

much time per day in minutes on average do you do physical training (e.g., running, yoga, cross-fit?) - 0.25 (CSGO), 0.27 (LoL)

gaming disorder - deliberate practice (item 3) - 0.01 (CSGO), 0.11 (LoL)

gaming disorder - deliberate practice (item 4) - 0.15 (CSGO), 0.14 (LoL)

gaming disorder - naive practice (item 4) - 0.17 (CSGO), 0.14 (LoL)

gaming disorder - naive practice (item 5) - 0.14 (CSGO), 0.06 (LoL)

gaming disorder - ADHD - 0.30 (CSGO), 0.41 (LoL)

(Based on the recent meta-analysis, doi: [10.1016/j.addbeh.2022.107590](https://doi.org/10.1016/j.addbeh.2022.107590) gaming disorder should correlate with gaming time of about $r = .22 - .38$ and gaming disorder with ADHD of about $r = .26$). Please note that practice items and gaming time items are distinct.

difference between 33% youngest and 33% oldest in attention

Cohen's $d = 0.19$ ($p = 0.30$) (CSGO); 0.15 ($p = 0.24$) (LoL)

difference between 33% youngest and 33% oldest in reaction time

Cohen's $d = 0.49$ ($p = 0.01$) (CSGO); 0.27 ($p = 0.03$) (LoL)

difference between 25% youngest and 25% oldest in attention

Cohen's $d = 0.35$ ($p = 0.07$) (CSGO); 0.19 ($p = 0.17$) (LoL)

difference between 25% youngest and 25% oldest in reaction time

Cohen's $d = 0.58$ ($p = 0.01$) (CSGO); 0.33 ($p = 0.01$) (LoL)

2B. Whether the introduction, rationale and stated hypotheses (where applicable) are the same as the approved Stage 1 submission.

These seem to be consistent.

2C. Whether the authors adhered precisely to the registered study procedures.

There seem to have been more pilot studies conducted between the last and this state of review. These may have been necessary, but need to be better integrated into the manuscript.

Author's response: After the In Principle Acceptance, no additional Pilot studies were conducted or added in the manuscript.

I was not able to find the R script used to run the statistical analyses. These should be made available for closer review, or at least more clearly linked in the analysis section.

Author's response: We are sorry for the problems with locating the R file. To make it easier to find (originally stored in the same OSF folder called Data with data files), we have created a new component called Analysis and dropped the R file there.

2D. Where applicable, whether any unregistered exploratory analyses are justified, methodologically sound, and informative.

I do not understand why reaction time and percent error are used as independent correlates for the attention and speed of decision-making measures. These need to be motivated further or removed.

Author's response: We agree that the reasoning for these operationalizations was lacking at Stage 1; we have added reasoning with the hope that recommenders allow us to add information to this part.

Fortnite is alluded to in the methods, but then not discussed in the manuscript. Yes, there is a link to another repository, but if the results are not sufficient for placement in the manuscript, then they should be removed altogether. Again, Fortnite is another esports title and another opportunity to explore the hypotheses, but the authors need to be more purposeful: either include it in the main manuscript with the disclaimer it was an exploratory title for analysis or remove it.

Author's response: For the sake of a better flow of the article, we have decided to transfer Fortnite results into the self-standing supplement, which is easily accessible from the Results section. We found it counterproductive and confusing to mix and later also discuss exploratory results that have weak evidential value with confirmatory ones. Since the Fortnite results were from the beginning planned as exploratory (we had additional resources to collect exploratory data from another game), we considered a transparent transfer of Fortnite results into the supplementary as appropriate. At the same time, after the IPA we are not allowed to change the Introduction/Methods section anymore; thus we would kindly prefer to keep Fortnite results in the supplement.

Some of the exploratory analyses do not seem justified and raise more questions about the dataset. For example, what does “ping” have to do with testing deliberate practice theory? The authors need to consider which variables they have in their dataset are most relevant to the goals of the manuscript.

Author's response: Ping indeed has probably only little to do with deliberate practice theory. However, our study is not solely focused on testing deliberate practice theory and we have tried to communicate this within the Research question right in the introduction: *“In the present study, our goal is to test if the deliberate practice theory of performance development applies to esports, and how other psychological, demographic, and environmental components might also contribute to long-term esports success.”* It is by including other variables (*“psychological, demographic, and environmental components”*) that we have tried to find a more holistic approach to explain esports performance differences. Furthermore, we have clearly described each of the exploratory analyses as exploratory and, as the nature of the matter implies, exploratory analyses may not always be supported by existing theory.

2E. Whether the authors' conclusions are justified given the evidence.

Considering the concerns raised earlier, it is unclear if the conclusions are supported. From what I can gather, the authors are suggesting that deliberate practice is not as important to the long-term skilled performance of esports players compared to naïve practice. To me, the authors overemphasize the deliberate practice piece and undersell the importance of the naïve practice piece. There are too many unknowns regarding the measure of deliberate practice, whether the sample contained any esports experts, and if statistical issues (e.g., multicollinearity) were present. The authors should consider focusing more on the role of naïve practice in the discussion and, depending on revisiting the analyses, how much to discuss deliberate practice.