

Dear reviewers,

We are appreciative of the continued feedback that has allowed us to further revise this piece of proposed research. To Dr. Bertsch's comment, we do still see merit in collecting confidence ratings, but have opted to remove the confidence ratings questions to focus on the primary aims of the study. We appreciate your endorsement of our methodological choices in the last draft provided. In cases where we inadequately addressed initial concerns or comments, we are grateful for the continued feedback in this revision. On the following page, we have included the main comments provided by Dr. Loaiza for this draft, with a small explanation as to how we have addressed the issue in the revised manuscript. Small comments about wordings have also been changed and should be clearly reflected in the Method section. Given that we are now using associated word pairs derived from Nelson et al.'s (2004) free association norms, we have slightly amended the word pairs to be word pair associations instead of strictly synonyms, which allowed us to more appropriately stay within a moderate range of forward strength relatedness. This small change is reflected in the Method section and will also be reflected in the instructions and examples given to participants.

We look forward to hearing from you in due course.

All the best,

Michaela Ritchie

Comments	Addressed	Manuscript Reference
<p>Fully within-subjects design: The authors did not address why they plan to use a mixed design. Besides affording more power (see next), the use of the Nelson norms as stimuli gives plenty of options so that the cell size could increase, in case that was at all a barrier to a fully within-subjects design. I do not insist on this, but it seems like it would address likely issues of power.</p>	<p>Thank you for suggesting we use Nelson norms in selecting word pairs, as it has been immensely helpful. We have shifted to a fully-within subjects design and have re-specified our power analysis to reflect a sample size required for this design. Using a fully within subjects design, we were also reasonably able to increase our desired power from 0.90 to 0.95</p>	<p>See Method (p.12) See Table (p.29).</p>
<p>2. Power: Thanks for the clarity here, but since an interaction is predicted (H2), it needs to be powered for, not just the main effect of generate vs. read. If sticking with the mixed design, many more participants than 69 (which I assume is the total) will likely be needed.</p>	<p>Sample size issues are addressed above. Rationale for detecting interaction effect is also made more clear in new draft: “The effect size included in the power analysis reflects the main effect of task type (read versus generate) which is expected to be observed. We are also testing an interaction effect between task type and sensory modality, which we expect will yield at least a medium effect size, given that both overt and covert generation tasks yield similar effect sizes (see McCurdy et al., 2020, for a review).”</p>	<p>See Method (p.12).</p>
<p>3. Inconsistencies/difficulties with the conditions explained in the Method section: I have given advice about how to make certain aspects of the Method easier to read/understand and making sure everything is clearly specified and consistent across conditions.</p>	<p>We have made substantial revisions to the Method section to remove redundant information and to make clear that all aspects of the conditions</p>	

<p>I would qualify these as relatively minor issues, and if anything a little bit fussy of me on the last point, but for sure powering for an interaction is pretty important.</p>	<p>are the same apart from the experimental manipulation. The revised Method section should be more concise and easy to follow. Thank you for pointing this out!</p>	
<p>Observed power is basically useless! I know R2 suggested it, but I thought I would point it out that this has been discussed a lot, e.g., <a href="https://daniellakens.blogspot.com/2014/12/observed-power-and-what-to-do-if-your.html">https://daniellakens.blogspot.com/2014/12/observed-power-and-what-to-do-if-your.html</a></p>	<p>We had included this in alignment with the suggestions of the last revision, although we are aware of this discussion and are happy to amend here. As of this revision, we have removed the statement of calculating observed power, as the discussion does seem to point to its lack of utility.</p>	
<p>Incorrect citations</p>	<p>These two citations have been formatted correctly and are now included in the reference page. We apologize sincerely for this oversight!</p>	
<p>Table not referenced in manuscript</p>	<p>The table is now called out in manuscript and updated to reflect new sample size and power calculation using a within-subjects design.</p>	