**Peer Review and Communication History**

**MS Title**: The Phenomenological Control Scale: Measuring the capacity for creating illusory nonvolition, hallucination and delusion

**Author Names**: Peter Lush, Ryan B. Scott, Anil K. Seth & Zoltan Dienes

**Submitted:** Apr 26, 2021

**Editor First Decision**: Revise & Resubmit

Aug 18, 2021

Dear Dr. Lush,

I have now received three reviews of your manuscript, “The Phenomenological Control Scale: Measuring the capacity for creating illusory nonvolition, hallucination and delusion.” I also independently read the manuscript before consulting these reviews.

The reviewers see value in the study, although they have many comments about the theoretical framing of the study, the reporting of the methods, the description of the results, and the interpretation of the results.

One major set of objections from the reviewers stems from the embedded theoretical assumptions that frame the research. The reviewers both disagree with some of the assumptions and also feel that these assumptions were not made explicit. As Reviewer 1 put it, “I think the authors need to be more explicit about the theoretical assumptions and commitments associated with phenomenological control including the overarching assumption that responsiveness to suggestion indexes phenomenological control, which is never actually justified.” A few of the objections may stem from misunderstanding of what you are proposing about this hypothetical construct of phenomenological control.

In some of their comments, the reviewers object to what they see as a rosy interpretation of the quantitative findings. They suggest that the viability of the phenomenological control construct is at the least highly uncertain and some of the reviewers’ view of the evidence in the broader literature is that it actually rules against the validity of phenomenological control and thus they object to the manner in which the interesting topics in here are being pursued. As Reviewer 1 put it, “I think the paper would be greatly strengthened if you provided some validation research demonstrating that responsiveness to suggestion measures what you call “phenomenological control””.

I think if a revised manuscript were to express better the degree of uncertainty around the hypothesized construct and the different views that may be warranted about them, with substantial revision of the reporting of the study, and thoughtful attention to the other comments of the reviewers, your report could be acceptable for publication in Collabra:Psychology. By softening rhetoric where it may over-claim, such as the phrase “the role of any induction or altered state of consciousness can be approached in a scientifically neutral way”, and making in some cases substantive changes in response to the reviewers’ points, I think the manuscript will be much improved.

If you do choose to undertake this revision, please respond to each comment of the reviewers, point-by-point, quoting each comment.

If you have any questions or difficulties during this process, please contact the editorial office at editorialoffice@collabra.org.

In the remainder of this letter, I provide a few comments from my own reading that I didn’t notice were mentioned by the external reviewers:

-There isn’t a capsule summary of what the study is before the Methods section starts, and a resulting problem is that it takes a lot of reading through the Methods for a reader to discern what the main methodological features of the study are

-I don’t see what the interval was between test and retest, but I might have missed it.

-the manuscript argues that “imaginative suggestibility” and “creative imagination” is no good, but “phenomenological control” probably has its own problems in leading some readers to think that you mean something that maybe you don’t, like that a lot of people can hallucinate at will (unless you do mean that?). There probably is no term that is perfect, I don’t know what you can do except work harder to clarify what you mean and add qualifier or hedge words that help convey what you don’t mean at various points in the manuscript.

-When you say “measures of changes in experience in psychological phenomena (e.g.,
the rubber hand illusion, mirror synaesthesia and visually evoked auditory illusions) are
substantially predicted by response to direct imaginative suggestion in the hypnotic context”, would it be better to say measures of changes in reported experience, not changes in experience? I think this point is related to some of the difficulties in knowing how strong your claim is about phenomenological control or what exactly is included in the construct.

-Many would object to this sentence and call it a fragment as it contains no verb: “Thus, the potential usefulness of a phenomenological control scale.”

Sincerely,

Alex Holcombe
Associate Editor

# Reviewer 1

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

Review

This is an interesting line of research and a potentially valuable contribution to the literature but I have a number of concerns with this paper that I think need to be addressed before this paper is suitable for publication. Most notably, I think the authors need to be more explicit about the theoretical assumptions and commitments associated with phenomenological control including the overarching assumption that responsiveness to suggestion indexes phenomenological control, which is never actually justified. Perhaps more importantly, I also think the authors need to be more objective about the limitations of this scale as it has a number of significant psychometric shortcomings - the authors should be more transparent about these.

Phenomenological control & suggestion
In their recent work, the authors have advanced the notion of phenomenological control. Although this is an interesting phenomenon that is worthy of study there are some problematic elements pertaining to the use of this term.

The authors assume that suggestibility (responsiveness to suggestions) is a form of phenomenological control (the capacity to generate experiences to meet one’s expectancies) and really just present this as a fact that doesn’t require justification. However, at present, this is really a hypothesis.

The central claim is that you’re study the capacity to generate experiences to meet your expectancies and thus expectancies figure prominently here. Indeed, expectancies do predict suggestibility on conventional hypnosis scales (e.g., Lynn et al., 2008, Oxford handbook of hypnosis). However, these data are correlational and sometimes the effects are actually not especially impressive (e.g., Benham et al., 2006, J Per Soc Psych). Critically, multiple studies have demonstrated that expectancy manipulations (which successfully increase expectancies) do not produce corresponding changes in suggestibility (Benham et al., 1998, J Pers Soc Psych; Lifshitz et al., 2012, Conscious Cognit). Although expectations play a role in response to suggestion, these and other data clearly indicate that expectation is unlikely to be a core mechanistic variable and this poses significant challenges for response to suggestion as a form of phenomenological control.

I think the paper would be greatly strengthened if you provided some validation research demonstrating that responsiveness to suggestion measures what you call “phenomenological control”, rather than just assuming this.

The authors maintain that phenomenological control indexes both “context-general direct and indirect suggestions” but this scale, like other hypnosis scales, is really measuring direct verbal suggestibility (Oakley et al., 2021, Conscious Cognit). For example, measures of indirect suggestibility have been shown to not reliably correlate with measures of direct suggestibility (e.g., hypnosis scales) (Polczyk, 2016, Curr Iss Pers Psych). This is relevant for other sections where you assume that direct and implicit suggestions tap the same domain.

“Social compliance (which is distinct from response to imaginative suggestion)”
This is inaccurate. Suggestibility correlates with standardised measures of compliance (Polczyk & Pasek, 2006, Int J Clin Exp Hypn) - the magnitude of these effects is similar to what the authors found with the rubber hand illusion, what they called a “substantial” relationship. In addition, many participants report responding to suggestions via basic compliant responding (e.g., Bowers, 1981, Int J Clin Exp Hypn; Bowers et al., 1988, Int J Clin Exp Hypn). Compliance is almost certainly an element of suggestibility as indexed by direct verbal suggestibility scales, albeit not the primary mechanism.

The authors should more explicitly define suggestion. Their example: “suggestion (e.g. “Did you think of trying this?”)“ references how the term is used in the general public but this is different from how it’s used in the experimental/clinical literature where it’s typically understood as a communication for an involuntary response (“your pain will go down”). Individuals who score highly on a typical suggestion scale are unlikely to be more responsive to “did you think of trying this?” so this seems irrelevant. Please more clearly define suggestion.

“which phenomenological control can surreptitiously bring about “
All of the studies you’ve done to date have been correlational - please avoid causal language.

“how people can strategically (although not knowingly) alter their conscious experience”
Provide evidence that they don’t know, especially given the high compliance rates (Bowers et al., 1988).

Context
“in a scientifically neutral way “
But the scale assumes a theoretical framework (cold control) and thus is not scientifically neutral. To have this be completely neutral, you’d have to refer to imagination-mediated suggestion.

“By contrast the phrase “the capacity for phenomenological control” does not presume anything on this matter, one way or the other.”
It does carry with it hidden assumptions regarding the underlying mechanisms (e.g., the whole model is premised on the authors’ cold control theory of suggestion/hypnosis) and thus, similar to “intentional binding” is not a theory-neutral descriptive term, especially given the way it’s defined (viz. expectation). It’s fine to have a hypothesis-driven term but I think this just needs to be more explicitly stated and not claim that the term itself is theory-neutral.

The motivation to move beyond the hypnotic context is understandable but it’s important to emphasise that you’re introducing a new context (phenomenological control) to participants and this introduces other confounds and factors into the mix. To be sure, I think it’s probably less problematic than a hypnotic context but it still may shape your results in various ways. I think this needs to be more readily acknowledged.

“…a non-hypnotic scale which better matches the contextual expectancies of other experiments and avoids the hypnotic context is potentially better suited for such investigation.”
This is a great point however this is still occurring in the context of an “imagination” task, which is very dissimilar from a lot of experimental work and carries with it various assumptions/biases/etc.

Imagination
I found references to imagination to be confusing or problematic in places.

“While the term “imaginative suggestion” accurately conveys the nature of scale items…”
Many would contest this as evidence for a role of imagination in response to suggestion is weak and not all suggestions even reference imagery.

“Previous terminology also employs the term “imagination”, which is also problematic.”
But you’re referring to your suggestions as imaginative (e.g., “the suggestions which follow as exercises in using imagination to alter subjective experience”) and using imagination as a preamble, how is this not a contradiction?

Magnitude of effects
The authors are variable in how they describe their effects and I think they need to be more consistent in their language and more objective. In the introduction, they refer to correlations between suggestibility and the rubber hand illusion as “substantial” but the correlations are actually in the weak-moderate range (~.30). In the discussion of this paper, the authors also refer to correlations of around .5-.6 as “substantial”. They also describe Cronbach’s alphas that are below commonly-used thresholds (.70) as substantial, when these are better described as “acceptable”. Later on, item-total correlations (which were around .48 for the SWASH) were described as “moderate” in the Discussion.

Current study
The motivation for this study could be made clearer. It seems that it’s about presenting norms but it’s also about comparing the PCS vs. a hypnosis scale so it really has bearing on the impact of an induction (and a different preamble) on response to suggestion. Indeed, this secondary aim about comparing the scales is the first sentence of the discussion.

Methods
Please justify the sample size for the two screenings.

Please clarify why both re-tests used the SWASH (hypnosis) scale. This means, of course, that you don’t have test-retest for the PCs. Or is this a typo?

The Materials section could be clearer, such as by explicitly describing that there were two distinct scales with a short paragraph for each. As is standard for Materials sections, more information is needed regarding the PCS and SWASH including the number of items, the nature of the audio files for delivery, the anchors for the subjective scoring, details about the preamble, instructions, and so on. There is a reference to an ‘induction’ in the procedure - is this the hypnotic induction of the SWASH or is there an induction in the PCS?

Was the subjective scale exactly the same for each item or does the wording vary? Please clarify. What were the subjective scale anchors?

Why were two subjective responses requested for two items and which ones? Why was this done? It seems peculiar to have extra scoring for only two items - this could be much clearer.

Please explicitly state in the Methods the number of days (M, SD) between the test and retest sessions - sorry if I missed this.
 Analyses
“to estimate the stability of phenomenological control capacities across the two screening procedures.”
Yes, but it should be clear that you didn’t administer the PCS both times (see above) - unless again if this was a typo.
 Results
The scale differences (Tables 1 and 2) are mostly driven by ideomotor and motor items (except taste; the mosquito hallucination often functions as a motor item because it involves a motor response; Woody et al. 2005, Psych Assess). All of the cognitive items except taste, which is the easiest one, don’t really differ between the two scales and posthypnotic suggestion is actually higher. This aligns with various spectral analyses showing that the different suggestions differentially target different underlying abilities. Of note here, compliance is more strongly related to the easy items (Polczyk & Basek, 2006, Int J Clin Exp Hypn) - increases in the PCS might be compliance effects.

It looks like you’ve done Welch t-tests (based on the df) - this should be stated.

Please comment on how the correlation between behavioural and subjective scales is greater for the SWASH than the PCS. The latter is clearly significantly lower. This suggests that the PCS is not tapping the classic suggestion effect to the same extent as the SWASH - this has important implications for the interpretation of this measure, particularly since you’re saying it indexes changes in volition.

The fact that the amnesia and posthypnotic suggestions do not correlate (behaviour-subjective association) indicates that these are poor items. This aligns with previous research, where they also have poor psychometric properties (e.g., Sadler & Woody, 2004, Int J Clin Exp Hypn; Acunzo & Terhune, 2021, Int J Clin Exp Hypn). I recommend removing these items and replacing them with better items.

“The upper limit of the 95% CI for the PC subjective scale alpha was lower than the lower limit for the SWASH, .68, 95% CI [.62, .74], but also showed substantial consistency.”
“Subjective scale alpha was substantial for both PC and SWASH.”
I’m sorry but this is inaccurate, given that the alpha was below a common threshold for what is considered “acceptable” for a psychometric instrument (alpha > .70), it’s clearly not “substantial”, especially given the other problems with the scale. At the very least, change this to “acceptable”. Also, specify that this is Cronbach’s alpha in the Discussion as some readers might find this unclear.

“Alpha did not show good reliability for the objective SWASH scale .49, 95% CI [.39, .57] …”
This was also the case with the original SWASH paper, which begs the question why this scale was not revised. The item/suggestion content of these scales should really be reconsidered.

A number of the items do not correlate well with the total scale scores on the PCS (negative visual hallucination, arm immobilisation, posthypnotic suggestion) - again, this points to the poor structure of these scales. This needs to be more readily acknowledged and is at odds with the statements in the Discussion arguing that this is a good scale.

The retest data and design are unclear given the description of the retest design in the Methods section (I assume the description in the Methods is incorrect?)

In the context of the retest again it would be nice to know the amount of time between sessions.

Why are BFs reported for some analyses but not for others? This could be clearer.

It would be valuable if the Results section more clearly distinguished a priori from exploratory analyses as it’s easy for the reader to mix up the two since the pre-registration is only mentioned in passing in the Methods section.

Please consider reporting the subjective-behavioural correlations across measures.

Discussion
It would be valuable in the Discussion to compare the test-retest reliability of the scales to other measures (e.g., the Elkins scale or the Harvard/Stanford scales).

The increase in response to suggestions in the PCS relative to the SWASH was really specific to low and medium responders - this was also found by Scacchia and De Pascalis (2020) and should be noted as it has implications for suggestion research, which is often on very high responders.

“The hypnotic context and induction may reduce correlations between hypnotisability and the use of phenomenological control in other contexts.”
Quite the opposite actually. The PCS has lower internal consistency, which places an upper boundary on the correlations you can expect to observe. All things being equal, these data suggest you’re going to get poorer correlations with the PCS than the SWASH. For example, this is probably why you observed better correlations with the subjective scale than the behavioural scale in your rubber hand illusion research (this has also been observed with other trait studies on hypnotic suggestibility).

“We suggest that the presentation of imaginative suggestions in a scientific context (the context of a scientific experiment rather than the context of hypnosis) …”
This implies that hypnosis can’t be used in scientific experiments - please rephrase.

“While we do not recommend the use of scores generated by the objective scales …”
Why have you used them in your previous studies then?

“The scale thus has a mix of the suggestions requiring possibly different subskills …”
Please reference seminal studies by Woody and McConkey (2003, Int J Clin Exp Hypn; Woody et al. 2005, Psych Assess), who were the main proponents of this idea.

“The terms “induction” and “hypnosis” presume there is a special procedure for  inducing an altered state (Coe, 1992).”
Only to an extent. Many researchers have conceptualised the induction as a preamble/introduction similar to your imagination preamble or as a simple first suggestion (e.g., Nash).
Minor
“However, hypnosis is not required for successful responding.”
To what?

“Thus, the potential usefulness of a phenomenological control scale.”
Missing word?

“to avoid potentially unpleasant effects of two WSGC suggestions.”
Unclear - more detail is needed.

Define PC at first use.

Some references are no in the reference list (e.g., Parra…)

There is some repetition (e.g., mentioning groups of 50 multiple times seems unnecessary).

The reference to taste in the analysis section is unclear as there is no mention of taste or what this means in the Materials (again, it would be nice to know which suggestions are on this scale). Same goes for urge and amnesia responses - this is unclear because it was never introduced in the materials.

There is a reference to a posthypnotic suggestion but there is no such item on the PCS - this could be clearer.

What was the difference score reported in Lush et al. (2018)? Between which two conditions? This could be clearer.

“The ratio of these ORs was 4.9 …”, sorry but I found this unclear - please provide more information.

There a handful of random typos in Table 2, Table 5, the reporting of BFs (e.g., p. 17), correlation on p. 18, etc. so please correct these.

There are no a) and b) panels in Figure 1.

Table 6: does the term hypnotisable have any meaning here if no hypnosis was used with the PCS?

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  |  | ✔ |  |  |

# Reviewer 2

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The manuscript addresses a very interesting topic, with substantive theoretical and empirical implications for the advancement of the knowledge of consciousness and its alterations. The main problem is that authors do not mention previousscientific works that already carried out almost identical investigations, but with other scales. For example, T.X. Baber created two scales (Baber Suggestibility Scale, and Creative Imagination Scale, both without any prior hypnosis inductions. The Babers’s ain was to demonstrate that all hypnotic “phenomena”, included the experience of non volition when succeeding suggestions , could be elicited without hypnotic inductions… debunking, in this way, the concept of hypnotic trance. Even more, he created Task Motivational Instructions as an alternative to hypnotic suggestions. TMI elicited objective reactions and subjective experience similar to obtained using hypnosis induction and suggestions. Therefore , again, Barber and his colleagues’ series of studies showed evidence about the weakness of theoretical perspectives about de “essence” of hypnosis based on the intrinsic properties of hypnotic induction and hypnotic suggestions. I miss references to this line of work already old (from 1960 on…), which matches with the aims and goals of the manuscript.
On the other hand, The Phenomenological Control Scale statistical analysis, at least, lacks from an exploratory factorial analysis. Much better would be to use and confirmatory factorial analysis, so that internal consistency of the scale could be addressed from another perspective, and for scales’s low level on internal consistency could be better understood.
Therefore, the manuscript aims and goals are strongly interesting, but it have to mention previous research, justify better why to use The Phenomenological Control Scale (only a change of name of the Sussex Waterloo Scale of Hypnotisability), when there are other scales (as already I mentioned) created for the same purposes… and reanalyze data with more updated statistical methodology

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  |  | ✔ |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) | ✔ |  |  |  |  |

# Reviewer 3

##### Open response questions

### Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

The current study describes a new “phenomenological control scale” that measures trait response to imaginative suggestions. The scale is a modification of a previous hypnotic suggestibility scale. The results are expected, and the statistical analysis seems to be correct. However, the interpretation of the results, the literature review, the conclusions, and the concept of “phenomenological control” are all deeply problematic. Many of the conclusions are not supported by the results, and there are many overstatements and incorrect statements in the text. The study’s limitations are not critically considered. Relevant literature is not cited, or the results of cited studies are misrepresented to fit the authors’ narrative. My overall assessment is that the current manuscript is unacceptable. Below, I list my main specific concerns.

1. I find the concept of “phenomenological control” very problematic. Heathy psychology undergraduates, like the current sample, do not hallucinate, and they are not delusional. Normal humans cannot “control their conscious perception” and see things in front of their eyes that are not there. The whole idea of phenomenological control seems rather absurd to the reader, at least from how it is presented in the current manuscript. Phenomenological control is confounded by mental imagery. The participants probably follow the instructions and suggestions and engage in various forms of mental imagery. This alternative view is not considered, and the relationship between mental imagery and "“phenomenological control” is not discussed. This is a major weakness of the study.
2. True automatic hallucinations are possible to induce with a hypnotic suggestion. However, this is a rare ability, and even very highly hypnotizable individuals are an extremely heterogeneous group. There exists strong converging support for the idea that hypnosis may involve a special state of consciousness in some rare individuals. Hypnotic hallucinations are thus a rare phenomenon (like synaesthesia), and the current way of defining it is heavily confounded by mental imagery. These critical views must be discussed.
3. The study presents no evidence that the “generated experiences” were non-voluntary. Most of the participants were likely aware that they were generating mental imagery voluntarily, at least to some degree, or their experience was a mixture of volitional and automatic. Therefore, the statements on lines 367-368 are speculative and most likely incorrect. This is a critical point because “hallucinations” should be non-voluntary and experienced as real. The latter crucial metacognitive judgment was also not investigated in the current study. To what extent did the participants experience the “hypnotic experiences” are real instead of simply imagining things or acting?
4. No evidence for genuine alternations in immediate perceptual awareness is reported in the article. All results could stem from visual, auditory, tactile, and taste imagery (so lines 369-371 are overstatements). I also do not think that any of the current students were genuinely delusional as claimed on lines 370-373; at least, no evidence for this is presented. The participants probably just responded to social compliance and engaged in mental imagery and role-playing. I do not think you are justified in referring to the current experiences as “hallucinations” or “delusions”.
5. The authors should discuss the distinction between “suggestions” and “instructions”. To what extent were the subjects simply instructed to perform mental imagery. Instructions almost always lead to a response since only ordinal social co-operation is required (and in the current study, the capacity to perform mental imagery). Automatic responses and purposeful imagining are two fundamentally different processes in highly hypnotizable individuals. I am concerned that the authors mainly gave instructions to imaging things and not “suggestions”.
6. The issue of response bias is critical in the current study. The results might not reflect “phenomenological control”, or even the capacity to engage in vivid mental imagery, but just various forms of response bias. Biases that are particularly relevant to consider are demand characteristics, social desirability, and acquiescence bias. This issue should be explicitly discussed.
7. In the abstract and introduction, the rubber hand illusion is mentioned. However, the authors should remove this text because Lush et al Nat Communications 2020 present clear evidence against a relationship between hypnotic suggestibility and the rubber hand illusion. When a control condition is used in the analysis, which is required to obtain meaningful results, there is no relationship between the rubber hand illusion and the hypnotizability score. Thus, this finding actually seems to falsify the phenomenological control theory.
8. On page 5, the authors misrepresent the “constructive nature of consciousness”. Perceptual conscious experiences are not just constructed in the brain out of nothing but results from an interplay between top-down predictions and bottom-up sensory signals. The brain tries to predict the sensory inputs in order to minimize prediction errors (e.g., predictive coding); not construct “fake” sensory data. Hallucinations generated according to the phenomenological control theory would presumably maximize prediction errors. So, I do not follow the authors’ argument here. Although psychosis and some other clinical conditions are interesting to consider in this context, nothing in the cited theories supports that healthy individuals can construct genuine perceptual-sensory experiences in the way that the authors claim in the manuscript.
9. The authors claim that “phenomenological control” and hypnotisability are single stable traits without thoroughly discussing the likely possibility that these scales capture different kinds of subskills. There might not exist a single stable trait for “phenomenological control” or hypnotic suggestibility.
10. The scale was developed outside the context of hypnotic suggestions. One drawback could be that the scale no longer captures what is really interesting: those rare, highly suggestible individuals who experience genuine hallucinations. This limitation should be discussed.
11. The last sentence (line 394) should probably be deleted.

##### Rating scale questions

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) | ✔ |  |  |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) | ✔ |  |  |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) | ✔ |  |  |  |  |

**Author Response**
Sep 22, 2021

See a separate file titles “response-letter”pdf

**Editor Final Decision: Accept**

Oct 18

Congratulations on addressing much of the confusion sparked by your initial submission. I think you have also addressed well most of the objections. There are some important issues remaining according to our reviewers, but I think your manuscript is in a good state for engaging with the community.
Thank you for submitting your work to Collabra: Psychology.

**Open response questions**

Please write your review here. The author(s) will see this review. Your identity will not be revealed to the authors unless you also include your name (i.e., sign your review) in this box. It is up to you whether to reveal your identity or not, either is fine.

I think they reviewers have done a decent job of addressing my comments and I’m not going to suggest any further revisions. I still think there are some important conceptual issues that are glossed over. The term ‘phenomenological control’ is defined with reference to expectations but there is no measurement of expectation on the scale and the authors claim that expectation is not the primary causal mechanism at play. Elsewhere phenomenological control is defined as responsiveness to suggestion. Finally, the authors’ objections to the term ‘suggestibility’ are really just based on the colloquial use of this term and misguided, not the way it was originally defined in early psychological research.

**Rating scale questions**

|  | Strongly Disagree | Disagree | Neutral | Agree | Strongly Agree |
| --- | --- | --- | --- | --- | --- |
| The study/studies in this manuscript have strong construct validity (good measures and/or manipulations of the constructs the authors wish to study). (Choose “Neutral” if this is not an empirical manuscript) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong statistical validity (appropriate statistical tests, assumptions are clear and reasonable, no statistical errors, appropriate statistical inferences, etc.). (Choose “Neutral” if this is not an empirical manuscript) |  |  |  | ✔ |  |
| The study/studies in this manuscript have strong internal validity (any causal claims or implications are well-justified, alternative explanations are thoroughly considered, etc.). (Choose “Neutral” if this is not an empirical manuscript, or no causal claims are made or even vaguely implied.) |  | ✔ |  |  |  |
| The study/studies in this manuscript have strong external validity (authors appropriately constrain their conclusions based on the limits of the generalizability of their findings to other contexts (including from lab to real world), other populations, other stimuli or measures, etc.) |  | ✔ |  |  |  |