Subject: Comment on bg-2021-34

Authors' responses are in italics

Reviewer #3

Paper packaging: The first is related to the packaging and broader context for the study. The second issue pertains to the analysis of the data.
Regarding the packaging of the study, the authors can make the paper clearer by lessening the effort to cover all bases as they describe the motivation and summary of their work. They begin this journey in the Introduction by providing context and underscoring the lack of studies on the effects of CO_2-induced acidification in freshwater taxa, especially calcifiers, relative to their marine counterparts. Fine, but one paper will not significantly change that balance. In fact, shortly after making this case, the authors add a 'but wait' because the Chinese mitten crab might not be sensitive to elevated CO_2 due to its occurrence in freshwater habitats that range widely in CO_2 concentrations. So it indeed might not be a representative freshwater taxon to study CO_2 effects after all. The reselling of mitten crab (actually, selling the paper about mitten crab) picks up again in the Discussion but from a different, all things to all people, vibe "we aimed to demonstrate for the first time the physiological and behavioural consequences of a possible future CO2 mediated freshwater acidification scenario on a juvenile calcifying invertebrate, the Chinese mitten crab." I would much prefer that the authors provide a more modest and clear appraisal of the import of their work and their working hypotheses, and let the readers assign value to their efforts.

We will modify the motivation and summary of our work to appear less like we are trying to sell this paper as a finite answer that freshwater calcifiers in general are at risk of freshwater acidification and focus more on the mitten crab. It was not our intent to state that all freshwater calcifiers are susceptible. We were more trying to get across the notion that this species is generally quite tolerant to environmental change yet is quite susceptible to future freshwater acidification and then use this to try and get the message across that this area requires greater focus so that the scientific community as a whole can make a more clear conclusion on the susceptibility of freshwater calcifiers to future freshwater acidification. We do acknowledge though that one study will not get the answer to this question but it definitely provides some important groundwork for the scientific community to build off.

The second issue is more problematic as it pertains to the experimental design (including response variables) and data analyses. The following points are interrelated and would benefit from a synthetic solution. One is suggested below.

1. It is unclear to me why the time-course data were run as a one-factor design (one-way ANOVA) rather than a two-factor design as was done for the other analyses. It would seem that a two- way analysis is appropriate and would provide the same hypothesis test

as the one-way plus more (factor 1 vs factor 2 and test for the interaction between factors 1 and 2). Regarding the presentation of results in Table 1 from the series of one-way ANOVA's, the 'Treatment' appears to be incorrect as the control FW or the acidified FW treatment was only provided as one 'type' (i.e., the crabs were exposed to control FW in the 'Control FW' treatment OR were exposed to acidified FW in the 'Acidified FW' treatment, and not different levels of each factor) so what was being tested? Was it the change in response overtime? If so, in this case 'Time' would be the 'Treatment' (akafactor) and Control FW or Acidified FW was the study condition. If I am interpreting the table correctly, the table needs to be restructured as does the language describing the tests.

We apologise for the confusion. The interpretation you have made was correct and it would have been better structured by placing the treatment as time and the control vs acidified as the study condition. We will make this correction.

2. Building on #1, it seems that many of the data types are based on repeated measures over time. Hence, a repeated measures or profile analysis would seem appropriate.

We will clarify this in the methods. This was not a repeated measures as we had 4 tanks with multiple crabs per tank and at each sampling time crabs were haphazardly selected from these 4 tanks. So, the population of crabs was repeatedly measured but as we could not guarantee that the same individuals were selected then we went with a non-repeated measures analysis.

3. Building on #1 and #2, the response data collected in the experiments are highly interrelated. Within an experiment (e.g.,locomotor behavior) the responses are clearly interdependent as well as repeated. Further, the authors make this very point in the Discussion, i.e., the responses are interrelated. For example, locomotion and energy expenditure in general are more labile in order possibly to conserve other responses more aligned with survival (Lines403-04). There is likely a degree of nesting or hierarchy of responses (e.g.,the locomotor behavior metrics are more tightly interrelated responses within the entire set of responses used) but the authors DO argue in the Discussion that there is likely a network of inter-relatedness or covariance among the responses. If true (it likely is), then employing a set of univariate tests (i.e.,the one-way and two-way univariate ANOVA's) is the wrong approach.

I would recommend revamping the statistical analyses. Run the data as two-way multivariate analyses (either MANOVA's and/or repeated measure designs). Discount the critical p-values to accommodate multiple tests on the same set of data (i.e., data on the same individuals in some cases or crabs drawn from the same tanks).

We would like to thank the reviewer for these in-depth comments about our statistical analysis. Upon reading the reviewer comments we agree that a revamp of the analysis by running a MANOVA would be more appropriate. As we did not use the same animals we would not use a repeated measure design as there was really no repeated measures.

Technical corrections suggested by reviewer 3 will be addressed