We greatly thank the reviewer for their time reading our manuscript and providing such constructive feedback. Below, we answer their questions and detail our plan for addressing their comments and concerns in the revised version. Author responses begin with R:

Reviewer #1

This study examines the biological effect of NaOH on two invertebrate species (isopod and sea hare) in a 4-day long experiment testing three different pH levels. The study presents the results that are indicative of high sensitivity of both species under elevated pH (aligned with stronger OAE treatment). There are serious data missing on the carbonate chemistry and the overlap in conditions among the experimental treatments shows that the analyses have to be conducted in a different way (see the text below).

The **introduction** addressed OA and OAE and provides a rational for the OAE testing. However, it does provide an insufficient background on the OAE effects across different species. It also fails to provide OA effects (low pH) on the two examined species, which would give the background on their sensitivity. This is pertinent to more detailed explanation in lines 92-95.

R: Additional information pertaining to OA effects on our study species and related taxa has been added to the revised manuscript at lines 115-126, stating: "Research on OA effects for the two species examined is limited, however one study (Hughes et al. 2017) investigated ocean acidification effects on *P. taylori* and *I. resecata* and found a negative quadratic relationship between change in biomass and pH. This study also observed low sea hare mortality and high isopod mortality in response to ocean acidification (Hughes et al 2017). Studies investigating ocean acidification effects on other species of sea hares and isopods can be used to provide context. When exposed to ocean acidification conditions, the California sea hare, *Aplysia californica,* experienced altered behavior and acid-base regulation (Zlatkin and Heuer 2019, Zlatkin et al. 2020). When exposed to ocean acidification conditions, behavior of sea hare *Stylocheilus striatus* was altered, showing reduced speed and foraging success, and exhibiting increased metabolic demand (Horwitz et al 2020). Another species of isopod, *Idotea balthica*, when exposed to ocean acidification conditions, exhibited 100% mortality in high pCO_2 conditions (Wood et al. 2014). Effects of OAE on other macroinvertebrates is limited and effects on the two species examined is unknown." Unfortunately, there is very little scientific literature available on the effects of OAE on macroinvertebrates (we already leverage all those we could find in the third paragraph of the introduction), and none on the two species examined.

Importantly, these are the sub-tidal species inhabiting the eel grass habitat but no data on the pH variability in the Puget sound and specifically within the eelgrass ecosystem is provided. Such information is essential for us to understand the diel variability these species encounter, as well as the pH max- namely, if subjected to high pH conditions in situ in this temporal period (Aug-Sept), then we might be expecting some natural acclimation/adaptation capacity in these species, that might be impacting the experimental results. Also, are there any changes in the month-long Aug-Sept period?

R: The revised version will include greater information about pH variability in the Puget Sound/Sequim Bay and in eelgrass habitats: "Based on Cotter et al. (2022), pH within an eelgrass meadow at the entrance of Sequim Bay varied between 8.02 and 8.22 between low and high tide and was consistently higher than outside of the meadow. Unpublished data show tidal variation of the pH in the tidal channel of Sequim Bay ranging from 7.6 to 8.2." This information was added to the manuscript at lines 130-132 and will be discussed accordingly based on the study results.

Methodology is the section where most of the clarifications are needed to evaluate the correctness of the exp design.

It needs to be noted that while the invertebrates living in the eelgrass are exposed to pH variability, none of the experiments incorporated this pattern in the exposed.

R: The experiments were designed to determine the pH level, in the range of possible OAE releases, that would be lethal to eelgrass invertebrates if exposed to these conditions continuously. For example, a sessile organism living nearest the outfall of an OAE facility that continuously releases seawater with elevated alkalinity to a defined mixing zone could be subjected to these extreme perturbations in pH and alkalinity. Some OAE deployments are currently being planned under wastewater discharge regulations (eg, NPDES under the US Clean Water Act in the US Pacific Northwest) that allow for specific water quality exceedances within a defined mixing zone. While pH must return to ambient conditions by the edge of the mixing zone, a planned deployment could potentially override natural variability (e.g., aim for an elevated pH >=9 in the mixing zone) to optimize alkalinity delivery. For the purposes of understanding extreme exposure cases, varying the pH throughout the day in each treatment group would have confounded the results. We agree that varying pH throughout the day would be similar to what a biological community currently experiences or would experience at some distance from the outfall, -- and we note that for a realistic OAE pilot deployment, continuous 24/7 release of alkalinity to maintain a constant elevated pH signal is unlikely. However, the goal of this study was to understand what constant elevated pH did to organisms as a first step to designing experiments that more closely mimic conditions related to an OAE field deployment, which we now expand in the introduction. The objectives, at the end of the introduction at lines 139-142, was rewritten as "The objective of this study was to assess the biological responses of eelgrass epifauna (Taylor's sea hares and eelgrass isopods) to increased pH and alkalinity levels and determine the pH level, in the range of possible OAE releases, that would be lethal to eelgrass invertebrates if exposed to these conditions continuously, in order to inform future mCDR field trials and identify knowledge gaps pertaining to laboratory and field trials in the context of OAE."

In addition, it is not provided how much of the eelgrass is sufficient for their nutrition and if the added eelgrasses sufficed or not- could there be some malnutrition occurring in the experiments? How was the diatom concentration determined, what equal concentration distributed within all the experimental jars? Are these two specific invertebrates feeding on diatoms and what is the heir C intake from eelgrass vs diatom?

R: As specified L125-126, the invertebrates were not fed during the 4-day experiments, per the ASTM 2000 standards. They were provided eelgrass blades with diatom clumps only during the acclimation period, in large quantities and refreshed daily. Malnourishment is one likely cause of mortality in the non-pH adjusted controls. We note that ASTM 2000 standards do not accurately represent OAE conditions that will be experienced by organisms within the mixing zone. We have added the following text to clarify this point within the methods section at lines 171-172 "This means that malnourishment may lead to some mortality, in both treatment groups and non-pH adjusted controls". This has been acknowledged in the limitations section of the manuscript at lines 538-540 stating "According to ASTM 2000 standards for acute toxicity tests (ASTM, 2000), the experiments were conducted under starvation conditions. Due to this, the observed responses in the survivability of the sea hares might have been intensified by the starvation, leading to an overstretched result from an OAE perspective."

Provide pH variability data as this represents a baseline to what these species are exposed to- lines 112-113.

R: A table will be added to the supplementary material with the data measured at each water refresh: pH, DO, salinity, temperature, pCO_2, and NaOH added. In addition, a summary table (see below) will be added to the results section about changes in pH and pCO_2, with averaged measured pH, pCO_2, and calculated alkalinity, DIC, and aragonite saturation. This is not a final table, but a quick example summary table. Full tables of each water refresh will be made available in the revision in the supplementary materials.

Organism	Time Point	organisms present	Mean Salinit	Mean Temp (°C	Target pH_{NB}	Measured pH Mean sd	pCO2 (ppm) mean sd	Alkalinity (µmol kg^{-1}) mean sd	DIC (µmol kg^{-1}) mean sd	omega aragonite mean sd
Sea Hares	initial	No	30.16	13.74	7.8	7.81 ± 0.13	646.55 ± 158.63	1955.7	1870.5	1.12
Sea Hares	initial	No	30.16	13.74	8.3	8.26 ± 0.04	317.56 ± 59.15	2146.4	1870.5	2.98
Sea Hares	initial	No	30.16	13.74	8.8	8.73 ± 0.02	205.26 ± 78.83	2530.6	1870.5	7.34
Sea Hares	initial	No	30.16	13.74	9.3	9.22 ± 0.04	197.03 ± 94.69	3135	1870.5	14.86
Sea Hares	final	No	30.28	13.45	7.8	7.75 ± 0.13	685.74 ± 120.11	1791.4	1727.1	0.9
Sea Hares	final	No	30.28	13.45	8.3	8.21 ± 0.09	308.4 ± 67.07	1961	1727.1	2.45
Sea Hares	final	No	30.28	13.45	8.8	8.54 ± 0.66	181.27 ± 69.89	2176.8	1727.1	4.77
Sea Hares	final	No	30.28	13.45	9.3	9.16 ± 0.07	203.94 ± 267.32	2840.5	1727.1	12.72
Sea Hares	final	Yes	30.9	12.56	7.8	7.5 ± 0.18	1243.86 ± 413.63	1775.9	1774.4	0.51
Sea Hares	final	Yes	30.9	12.56	8.3	8.07 ± 0.13	442.96 ± 101.43	1944.8	1774.4	1.83
Sea Hares	final	Yes	30.9	12.56	8.8	8.6 ± 0.1	244.87 ± 95.24	2279.4	1774.4	5.37
Sea Hares	final	Yes	30.9	12.56	9.3	9.13 ± 0.09	186.73 ± 166.36	2828.6	1774.4	11.95
Isopods	initial	No	32.24	13.04	7.8	7.74 ± 0.11	752.7 ± 133.55	1943.9	1875	0.97
Isopods	initial	No	32.24	13.04	8.3	8.25 ± 0.04	361.47 ± 75.9	2156.4	1875	2.96
Isopods	initial	No	32.24	13.04	8.8	8.73 ± 0.02	230.59 ± 95.36	2558	1875	7.43
Isopods	initial	No	32.24	13.04	9.3	9.23 ± 0.03	245.65 ± 130.9	3185.3	1875	15.11
Isopods	final	No	32.29	12.57	7.8	7.69 ± 0.15	778.66 ± 182.55	1782.8	1731.4	0.79
Isopods	final	No	32.29	12.57	8.3	8.2 ± 0.07	356.56 ± 70.84	1968.2	1731.4	2.42
Isopods	final	No	32.29	12.57	8.8	8.68 ± 0.06	232.63 ± 84.49	2229	1731.4	5.17
Isopods	final	No	32.29	12.57	9.3	9.17 ± 0.05	176.91 ± 85.86	2882.6	1731.4	12.9
Isopods	final	Yes	32.78	11.38	7.8	7.65 ± 0.16	982.49 ± 274.71	2050.5	2010.6	0.8
Isopods	final	Yes	32.78	11.38	8.3	8.18 ± 0.07	356.69 ± 91.8	2253.3	2010.6	2.6
Isopods	final	Yes	32.78	11.38	8.8	8.67 ± 0.05	234.32 ± 77.45	2540.4	2010.6	5.71
Isopods	final	Yes	32.78	11.38	9.3	9.16 ± 0.04	210.6 ± 142.64	3264.7	2010.6	14.47

Why was unfiltered water used in experiments? Line 126

R: Unfiltered water was used because this is what we had abundant access to at our location. Our wet lab facility has raw seawater, which comes directly from its source (the bay) on demand. Our facility also has very coarsely filtered seawater (essentially sieved). Algae and bacteria are not removed from the filtered seawater, and it sits in a holding tank until needed, so can have lower oxygen than our raw seawater. The amount of water needed each day prohibited us from manually filtering water prior to use. Text has been added to the revised manuscript at lines 152-157 and now reads: “The PNNL wet laboratory facility has abundant access to raw seawater, which comes on demand directly from the bay via the facility’s seawater intake, located near the seafloor at about 10 m depth. The facility also has very coarsely filtered seawater, which sits in a holding tank until needed. This filtered seawater

however, can have lower oxygen levels than the raw seawater, and is more prone to increased temperatures inside of the tank. The amount of water needed each day prohibited us from manually filtering water prior to use. For these reasons, unfiltered water was used in the experiments."

What is the uncertainty of the YSI probe? Line 127

R: The accuracy of the probe is ±0.2 pH_{NBS} units. We calibrated the probe daily before use and checked it with pH 7 buffer ensuring it was within ± 0.05 pH before use. Text has been added to the revised manuscript at lines 175-176 stating "The accuracy of the probe is ±0.2 pH_{NBS} units. We calibrated the probe daily before use and checked it with pH 7 buffer ensuring it was within ± 0.05 pH before use."

How does the variability of pH (7.8 +/- 0.3) impact the amount of NaOH that needs to be added? Is it possible that water with higher initial pH, where less NaOH was added, could induce less negative effects?

R: The quantity of NaOH added is a function of the initial pH, temperature, salinity, and volume, as calculated by the CO2SYS Excel macro (Pierrot et al. 2021). So indeed, less NaOH would need to be added when starting pH was higher. These parameters were measured at each water treatment refresh in order to calculate the amount of NaOH needed. Text has been modified and lines 178-181 now states: "Temperature, and salinity, and initial pH readings measured in the bucket, along with water volume, were used to calculate the amount of NaOH needed to reach the desired pH for each treatment group, using a CO2SYS Excel macro (Pierrot et al., 2021). If initial pH was higher (closer to target pH), then less NaOH was needed to reach the target pH." In terms of negative effects, our hypothesis is that pH rather than alkalinity drives negative effects, and we achieved the same target pH regardless of starting pH. Within specific treatment groups, there was no correlation between the amount of NaOH added to reach the desired pH and mortality.

Was NaoH actually produced in the electrodialysis or was it commercially available? line 133-134.

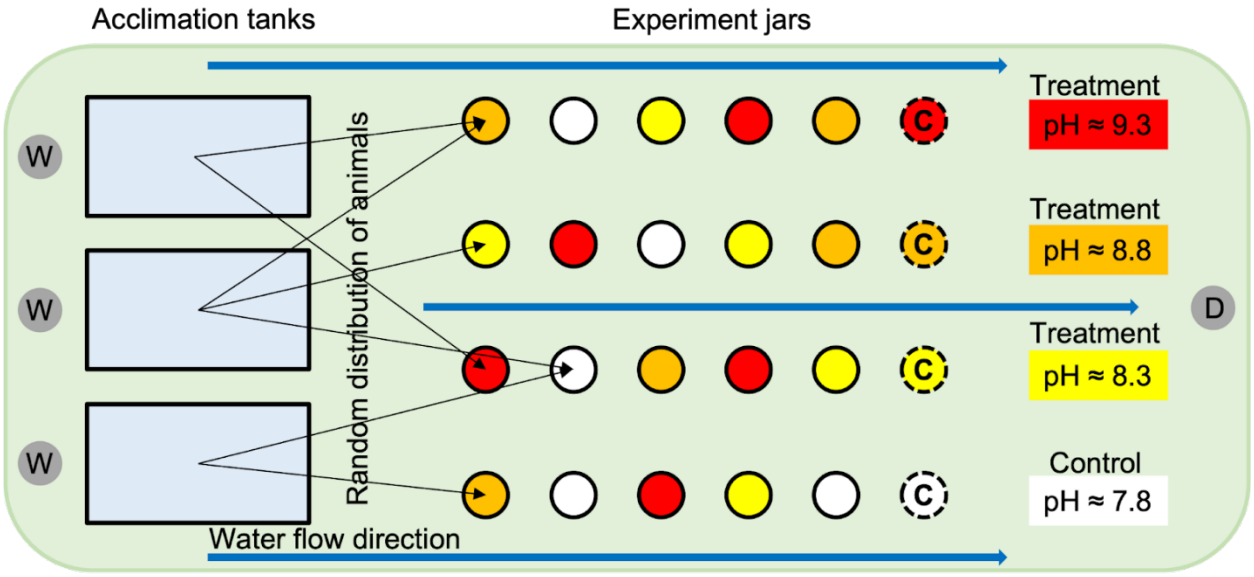
R: As stated in the manuscript L143, commercial NaOH (0.5 M, Honeywell Chemicals 352576X1L) was used in the experiment to supplement NaOH that would have been produced via electrodialysis. Text was added to the revised manuscript at lines 184-185 stating "Our facility is now capable of generating NaOH via electrodialysis, but this was not available at the beginning of the project/experiment" to provide clarification.

All OA/OAE biological studies need to report on the whole carbonate chemistry parameters and none of this is provided, not for the baseline chemistry and not for the amended (treated with NaOH) water. This absolutely needs to be added. In the same way, all the changes are only reported in pH, while we need to understand also the change in TA to understand how much NaOH was added to different treatments. Provide a clear and methodical way of representing missing data. Lines 177-180 are not solid justification and this needs to be amended.

R: Carbonate chemistry calculations will be added in revisions and will be included as a new table in the result section about changes in pH and pCO_2 (see above). We agree that it is important context for the experiments, which are primarily focused on biological responses. The table will include the amount of NaOH added and the calculated alkalinity.

Figure 1 is not clear and could be improved: 3 acclimation tanks with four treatments? How was this really conducted? Also missing is the chemical control with no NaOH added to demonstrate how the diel variability impacts diel changes in seawater before added NaOH.

R: Figure 1 has been improved for the revised version (see below). The number of acclimation tanks does not relate to the number of treatments. As explained L141-142, 5 individuals were randomly picked from each acclimation tank and moved into a jar at the start of each round of experiment. This will be clarified in the revised version and represented in the revised figure. In addition, the differences between the types of controls (non-treated water with animals, non-treated water without animals, and treated water without animals) will be clarified in the revised text at lines 197-201 by adding "Each three rounds of the experiments for both sea hares and isopods comprised a total of 24 jars: five at control pH (no NaOH added) with five animals each for a total of 25 animals, one chemical control (without animals) at control pH, five at each of the three pH treatments (NaOH added) with five animals each for a total of 25 animals per treatment, and one chemical control (without animals) for each of the three pH treatments (NaOH added).", and better represented in the revised figure. The figure caption now reads "Figure 1. Experimental setup of laboratory water table, including the 1-liter glass jars (colored circles) used in the experiment and the acclimation tanks (blue rectangles). The white circles indicate control jars with pH ≈ 7.8 (no NaOH added), yellow circles indicate low treatment with pH ≈ 8.3, orange circles indicate medium treatment with pH ≈ 8.8, and red circles indicate high treatment with pH ≈ 9.3. The dashed circles with "C" indicate the chemical control jars in which only treated, or control, seawater was added without the presence of sea hares or isopods. Animals were distributed randomly from the acclimation tanks to the experiment jars, as indicated by the black arrows. The circles "W" indicates the water input, the circle with "D" indicates the drain on the water table, and the blue arrows indicate the water flow along the table."



Why was pCO2 then measured if full carb chem was not calculated and provided (the variability in such big pH changes should still be theoretically lower than the uncertainty of the measurements).

R: As stated previously, we plan to include a summary of carbonate speciation calculations derived from pCO_2 and pH (see above). We intend to describe this data to provide context for experimental conditions that the organisms experienced.

A major drawback is reporting pooled data on water chemistry as well as on the experimental results. You should provide unpooled data and conduct analyses on this? How else was the variability determined that the within the treatment levels?

R: Unpooled data will now be provided in a table as supplementary material. However, due to the low numbers, data were pooled for the purpose of statistical analyses. Within treatment level variability was determined one-way ANOVAs.

Please, explain table 1 in more detail, I do not understand how min, max and average pH/T/DO levels determined the treatments? Why are there three values- is it form the diel variability? If so, why was the water not always taken at the same time to avoid this variability? How does this impact NaOH additions. The variability on the initial pH levels is almost as high as between treatment 1 and 2.

R: Because of the time it took to refresh the water for all 24 jars, the water drawn over 3h was subject to diel and tidal variability. Seawater was pumped from the bay directly by our facility, and the water we receive is subject to environmental variability. Data from each jar was taken into account to calculate how much NaOH was necessary to add to reach the desired treatment pH. The table caption has been clarified in the revised version to read as "Table 1. Because of the diel and tidal variability in water quality over the time the water was drawn to refresh the jars each day, this table summarizes the overall range in pH, salinity, dissolved oxygen, and temperature of ambient seawater experienced within each round of experiment before NaOH treatment was added to each treatment group."

Results section:

Why are the changes in pCO2 and pH are not a major part of the experiments in linking chemistry to the biological data? Strongly consider removing after all the parameters have been provided.

R: As mentioned above, we will add carbonate speciation calculations to address the reviewer's previous comments. Based on this comment, we will consider noting carbonate speciation in the methods section as experimental context rather than presenting chemical data as results. We thank the reviewer for this useful suggestion.

Why was there such an increase in pCO2? Line 201-204 inaccurate statements

R: The mason jars were airtight and contained active microbial communities. Although there is normal indoor lighting in our wet labs, we did not subject the jars to grow lights, nighttime lighting, or natural lights (it is a windowless lab). Therefore, algal production likely slowed/stopped while microbial respiration continued, resulting in CO_2 production and O_2 consumption. Likewise, the animals were actively respiring until dead, producing elevated pCO_2 and lowering pH. This is why we refreshed the water twice per day to maintain pH near our target treatments. This description will be added to the text in the results section.

From the Figure 2 is appears that in most rounds the treatments 8.3, 8.8 and 9.3 were the same (accounting for the variability) and there were no statistical differences between them, only 7.8 was different (not in Round 1 for no hares present, and in Round 1 and 2 for hares present). The same issue of the treatment overlap was present in the isopod treatments. This likely is the consequence of a not properly tight system. Can you provide pH data for all the rounds?

R: As explained above, this is probably a result of uncontrolled microbial activity in the jars. pH data for all water refreshes will be provided in the revised version as supplementary material.

Given the overlap in experimental treatments, it is impossible to separate the effects on the treatment levels. First, the levels need to be examined to really understand which of them are sufficiently different and only then examine biological differences. Right now section 3.2 is not valid. As it seems, the combined 8.3, 8.8 and 9.3 could be one treatment, which could be compared to 7.8, which means a 2-treatment design (at the best).

R: The treatments we were interested in testing were the levels of pH, which did not overlap even with the YSI error of +/- 0.2 in which treatments could fall into the ranges of 8.1-8.5, 8.6-9, and 9.1-9.5. In the revised manuscript we will provide a table in the supplementary material of all measured pH values to show that these were separate treatments.

What is mortality (LC50) or LOEC? This is an ecotox study so present all data, i.e. NOEC; LOEC, EC50 on the combined graphs (not only LC50).

R: Text has been added to the revised manuscript at lines 456-458 stating “The lowest observed effect concentration (LOEC) for sea hares was pH of 8.8, and the no observed effect concentration (NOEC) for sea hares was pH of 8.3. For isopods, LOEC and NOEC could not be determined because no mortality in any treatment group was statistically higher than mortality in the control group.”

I am not sure why all the regression models (Table 3) were built with multiple parameters 8salinity, DO, temp) if only the pH (CO2) was intentionally (artificially) changed?

R: Our intention was to examine if drivers other than our intended manipulation (pH) caused mortality. For example, did a lack of oxygen, difference in temperature, or change in salinity confound our results? The lack of correlation with parameters other than pH confirm that pH indeed was the cause of mortality differences across treatments assuming malnourishment equally affected each pH treatment. This will be clarified in our revised manuscript: “Multiple linear regression models were used in conjunction with the boxplots to determine whether pH was the driving factor behind mortality, or whether changes in salinity, DO levels, or temperature may have confounded the results.”

Until these changes are introduced, the text from line 380 is not applicable (I did not continue from here onwards).

R: Instead of LC50, the figures and captions will be revised to present and explain the exposure time to each treatment after which 50% mortality occurs. We will also present NOEC (no observed effect concentration) and LOEC (lowest observed effect concentration), using pH as a concentration and clarifying this in the revised text. Text has been added to lines 456-458 that states: “The lowest observed effect concentration (LOEC) for sea hares was pH of 8.8, and the no observed effect concentration

(NOEC) for sea hares was pH of 8.3. For isopods, LOEC and NOEC could not be determined because no mortality in any treatment group was statistically higher than mortality in the control group."

Why is data discussed in the Discussion (growth and reproductive behaviors) not presented in the Results?

R: These observations will be added to the results in the revised version.

Discussion should be expanded!

R: Once introduction, methods, and results are revised (see various proposed revisions above), the discussion will be expanded to reflect the revised information: e.g., LOEC and NOEC in addition to LC50, applicability of ecotox experiments to OAE effects, direct application to OAE field trials (continuous exposure in mixing zone).