Details of reply

Dear reviewers,

We would like to thank the reviewers for his/her interest in our work for their effort, constructive criticism and suggestion. We appreciate the insightful comments, as these would contribute to improving the manuscript's robustness and quality. We provide a point-by-point reply to the general and specific comments raised as follows:

REVIEWER 1:

In the manuscript, Li and Wang have developed a survey to measure flood risk perceptions and flood preparedness. They have used statistical tests and regression analysis on the survey responses to determine the effect of different variables, e.g., gender and education level, on these dependent variables. These kind of surveys and studies are highly relevant and useful for informing disaster risk education, and risk mitigation policies.

Reply: Thanks greatly, we appreciate your positive assessment of our study.

I have some major concerns regarding the comprehensiveness of authors' methodologies that make it difficult to validate some of their conclusions. I have listed both my major and other concerns below.

1.The language in the paper could be improved for readability. Especially in the abstract and introduction, the use of past tense makes it harder to follow the definitions and literature review. To be clear, the improvement in language is not a reflection on the quality of the scientific content of this paper, which was reviewed independently.

Reply: Thank you for your suggestions, and we would enhance the overall readability of this paper. We would consider the use of present tense in the section of abstract and introduction, rephrasing them for better readability.

2.The methods used in the paper are only mentioned but not described. While it is okay to use standard methods like linear regression without explaining them, it will be helpful to the audience of this journal to explain when special methods are used such as stepwise regression. Additionally, references to these methods are consistently missing. Additional comments below mention some of the methods for which it would be helpful to add descriptions and references.

Reply: Thank you for these helpful suggestions. Stepwise regression is a type of multiple linear regression that can select the best-fitted combination of independent variables for dependent variable prediction with forward-adding and backward-deleting variables. The stepping procedure begins as an initial model definition, with a stepped forward addition of a variable to the previous model. The critical F value is then used to check the eligibility of the added variable. With a new variable added, the previous variables in the model may lose their predictive ability. Thus, stepping criteria are used to check the significance of all the included variables. If the variable is insignificant, then the backward method is used to delete it. Forward selection and backward elimination are repeated until no variable is added or removed. Stepping procedure would stop until the optimized model is established. We mainly considered the way of back elimination in stepwise regression and would add more details about stepwise regression analysis in the paper. And we would add more detailed description and relevant references in this paper.

3.Since the primary contribution of this manuscript is reviewing survey responses, it will be helpful

to understand how the 844 respondents were selected across the geographic region, and the criteria for selection. Please provide additional information, such as, why 844 respondents were selected instead of more or fewer.

Reply: Thanks a lot, we used the method of random sampling to perform the survey. Random sampling is a type of probability sampling method where everyone is selected entirely on a chance, with each one having an equal probability of being chosen. The reason for adopting random sampling was to reduce selection biases in the survey and to be able to calculate sampling error. There are various methods mentioned in the literature to determine the sample size. The sample size of respondents was calculated using Yamane's formula. Using a confidence level at 95%, the sampling method proposed sample size of 844. Random sampling was chosen to carry out the survey. A total of 737 samples were finalized after discarding the incomplete questionnaires.

$$n = \frac{N}{1 + N(e)^2}$$

where n is the sample size, N denotes the size of the resident population, and e is the precision level.

4.In addition to the above, please provide additional statistics to ascertain whether the survey respondents were reflective of the socio-economic distribution of the targeted geographic region.

Reply: Thanks a lot for your useful suggestions. In our survey, women accounted for 56.2% and men accounted for 43.8%. The number of people with a bachelor's degree or above accounted for 58.1% of the total number of the survey respondents, and the rest were high school education or below, accounting for 41.9%. The age groups were: 26 to 40 years old (33.3%), 41 to 60 years old (25.9%), and 9 people over 60 years old (11.9%). The flood control knowledge survey[1] issued by Nanjing Municipal People's Government included the social and economic distribution of citizens in recent years, which was consistent with these socio-economic features of survey respondents in our study. And we would add the related illustration in this paper.

5.The authors have mentioned that a detailed description of the survey questionnaire is available in supplementary material. Given the high relevance of the survey development to the authors' conclusions, it will be helpful if some more detailed information regarding survey development, improvement, and questions was included in the main text.

Reply: Thank you for your suggestions. To identify the potential problems (e.g. unclear and ambiguous questions), preliminary tests of online questionnaire were conducted before officially issued for the survey. We collected and sorted out the survey results of interviewees and modified the questionnaire reasonably according to the feedback. Then we deleted and reduced the questionnaire options that may lead to bias and misunderstanding. These details would be added to this paper.

6.For the purpose of drawing conclusions from the surveys, all 737 valid responses are considered ground truth and representative of the entire population of the targeted geographic region. In order to substantiate the conclusions, it will be beneficial if sensitivity analysis was performed on results. One possible approach could be providing confidence intervals on the results and regression coefficients. Another approach could be bootstrapping where a subsample of responses are selected

[1] More information was shown in this website.
https://www.nanjing.gov.cn/hdjl/zjdc/wsdc/dcbg/202310/t20231012_4030100.html

multiple times, and the coefficients are generated. This would help quantify the variability of the results in order to determine whether the differences across various factors are within the error margins or not, in order to draw conclusions.

Reply: Thanks greatly for your useful suggestions. And we would add the confidence interval on the results and regression coefficients in this paper.

7.Most of the text in Section 3 lists the numbers already present in the respective tables. The section could be made more succinct by only including the key observations from the tables.

Reply: Thanks greatly for your valuable suggestions. And we would modify this section carefully and make it more concise and accessible.

8.Line 163 - Were responses also collected from online distribution of the survey and did those responses match the ones collected directly by interviewers?

Reply: Thanks a lot, we tested the online survey to identify the potential problems, such as the unclear and ambiguous questions in the questionnaire. Based on the feedback from respondents, we could improve the questionnaire reasonably and reduce the questionnaire options that may lead to bias. Because there was the different structure and design between the original online questionnaire and face-to-face questionnaire surveys, it is difficult to determine whether these responses are well matched. And all the analysis in this study was based on the results of face-to-face questionnaire surveys.

9.Line 168 - It will be helpful to include whether any analysis was done to review interviewer bias in the responses.

Reply: Thanks for your useful suggestions. When designing and improving the questionnaire, we considered the order and logical relationship of the questionnaire questions to avoid subjective bias and misleading questions. And interviewer received excellent survey skills training, such as anonymous survey, before formal interview to avoid the misunderstanding and confusion during the interview. Therefore, we didn't carry out the analysis about interviewer bias.

10.Line 179 - Please include information about how the valid and invalid responses were determined.

Reply: Thanks a lot, and we mainly excluded the invalid responses following the criteria. (1) Incomplete questionnaire, that is, a considerable part of the questionnaire was not filled in. (2) Respondents did not understand the contents of the questionnaire and answered the questionnaire incorrectly, or did not answer the questionnaire according to the requirements of the guidance. (3) Interviewees answered the questionnaire with no change, such as in the five-point Likert Scale, no matter what the question, respondents will choose the same answer all through. (4) Defective questionnaires, that is, several pages are missing or cannot be identified. (5) Inconsistent or obviously wrong questionnaires.

11.Lines 184-187 - It will be helpful to provide more information about the implementation of Mann-Whitney U and Kruskal-Wallis statistical tests, along with relevant references. The current brief description is not sufficient to understand why these tests were chosen, how they were implemented, and their objectives.

Reply: Thank you for these helpful suggestions. The Mann-Whitney U statistical test is a nonparametric statistical test used to compare the values of a variable between two independent groups. The Kruskal-Wallis statistical test is also a nonparametric statistical test used to compare the values of a variable between several independent groups. The Mann-Whitney U test was used for 'yes or no' questions, and the Kruskal-Wallis test was for

questions with three or more answer choices. These tests were used to compare the differences of flood risk perception and flood preparedness between two and several independent groups in this study. If the Mann-Whitney U statistical test returns a P-value less than 0.05, the compared categories are significantly different. As a result, they can be ranked using the Mann-Whitney U statistical test's mean rank. If the P-value of the Kruskal-Wallis statistical test is less than0.05, the difference between the compared categories is considered significant. These statistical tests are frequently used and performed via statistical software. Accordingly, these statistical tests were conducted in this study using SPSS. And we would add more detailed description and relevant references in this paper.

12.Line 191 - Briefly describe stepwise regression and provide relevant references.

Reply: Thank you for your suggestions. Stepwise regression is a type of multiple linear regression that can select the best-fitted combination of independent variables for dependent variable prediction with forward-adding and backward-deleting variables. And we would add more detailed description and relevant references in this paper.

13.Line 192 - What is Model 5?

Reply: Thanks a lot for your advice. Model 5 is a moderated mediation model. In this model, risk perception, flood preparedness, response intention and social-economic factors acted as independent, dependent, mediating and moderating variables respectively. And we would modify this illustration in the paper.

14.Line 198 - What are Cronbach's α and KMO values? Please provide brief descriptions along with relevant references.

Reply: Thank you for your suggestions. Cronbach's Alpha coefficient is the most widely used measurement method in the test of questionnaire reliability. This coefficient is a number distributed between 0 and 1. If it is greater than 0.7, the data is acceptable. If it is greater than 0.8, the data is valuable. Cronbach's Alpha coefficient greater than 0.9 indicates high internal consistency. The calculation of KMO value is a common method in the test of questionnaire validity. The KMO value is used to evaluate the correlation between the variables in the data sample. KMO value is very important for the test of questionnaire validity. If the KMO value is greater than 0.7, it that the questionnaire has good validity and can be used for factor analysis and statistical analysis. If the KMO value is less than 0.5, it may mean that there is a problem in the design of questions in the questionnaire, and the questionnaire needs to be corrected or reconstructed. And we would add more detailed description and relevant references in this paper.

15.Line 232 - What is mean rank?

Reply: Thanks a lot for your suggestions. If the Mann-Whitney U statistical test returns a P-value less than 0.05, the compared categories are significantly different. As a result, they can be ranked using the Mann-Whitney U statistical test's mean rank. If the P-value of the Kruskal-Wallis statistical test is less than0.05, the difference between the compared categories is considered significant.

16.Section 3.2 - Why was Mann-Whitney U test used for binary variables and the Kruskal-Wallis test used for multi-category variables? Can the rank values between the two tests be compared with each other?

Reply: Thanks a lot for your suggestions. The Mann-Whitney U statistical test is a nonparametric statistical test used to compare the values of a variable between two

independent groups. The Kruskal-Wallis statistical test is also a nonparametric statistical test used to compare the values of a variable between several independent groups. The Mann-Whitney U test was used for 'yes or no' questions, and the Kruskal-Wallis test was for questions with three or more answer choices. These tests were used to compare the differences of flood risk perception and flood preparedness between two and several independent groups in this study. Because the test categories of each variable were different in these two nonparametric statistical tests, the rank values between the two tests could not be compared with each other.

17.Fig 3 - The colors in the correlation plot do not appear to match the color bar. For example, the diagonal should be solid red since correlation=1.0, but it's white indicating correlation=0. Similarly, value of 0.05 is shaded light red while a value of 0.76 is shaded white. As a result, Section 3.3 could not be reviewed for accuracy, and its conclusions could not be substantiated.

Reply: Thanks greatly for your valuable suggestions. We would check and redraw this figure and modify the overall organization of Section 3.3 carefully.

18.Section 3.4 - Briefly describe models 1, 2, and 3. From Table 12, it appears that each model used a different set of features.

Reply: Thanks for your suggestions. We used stepwise regression to reveal the influencing factors of flood risk perception. First, we selected all variables for regression analysis in model 1, and found that flood risk knowledge showed the significant and positive effect while other variables exhibited relatively lower effects. And then after removing socio-economic variables, we built the model 2 with a high goodness of fit (adjusted R^2=0.788). Flood risk knowledge also maintained a higher influence (0.827) on flood risk perception. Furthermore, we excluded the variable flood risk knowledge in model 3, with a low goodness of fit (adjusted R^2=0.246). But government trust, flood experience, flood disaster education and flood risk worry significantly and positively influenced risk perception, indicated by increased regression coefficients. The effect of flood experience on flood risk perception shifted from insignificant to significant. We found that although flood risk knowledge could significantly promote risk perception, it also inhibited and decreased the positive effects of other factors. And we would add this brief and concise description.

19.Sections 3.4, 3.5 - The difference in regression coefficients between various groups (e.g., males and females) appear quite small. It will be helpful to see whether these differences are in fact indicative of reality or within the error bars (such as, confidence intervals) based on the number of survey responses.

Reply: Thanks a lot for your suggestions. Although some regression coefficients are small in different groups, this study aimed to explore the differences in the influencing factors of flood risk perception and flood preparedness between different groups. It is more important that whether these different influencing factors show the significant effect or not. These small difference in regression coefficients between various groups could be accepted in this study. And we would also add the confidence intervals of different variables in this paper.

20.Fig 4 - Why are certain coefficients missing from the figure, e.g., females and flood disaster education? Same for Fig 5.

Reply: Thanks a lot for your suggestions. We only listed the significant results of regression analysis in Fig. 4 and Fig. 5. And therefore, some insignificant coefficients were not presented in these figures, which we mentioned in this original manuscript (Line 330). We would

highlight this description about Fig. 4 and Fig. 5 in the paper again.

21.Section 3.6 - Please describe how the influence path analysis was implemented, along with relevant references.

Reply: Thanks a lot for your suggestions. We performed a moderated mediation model in PROCESS macro program of SPSS to capture the influence path between flood risk perception and flood preparedness. The PROCESS program can effectively test the moderated mediation model and help to clarify the mediating and moderating roles of different variables. All statistical analyses were conducted at a significance level of 0.05. In this model, risk perception, flood preparedness, response intention and social-economic factors acted as independent, dependent, mediating and moderating variables respectively. And we would add more detailed description and relevant references in this paper.

22.Section 3.6 - New taxonomy is presented, e.g, M-1SD, without any explanation for its meaning. As a result, it was not clear how to interpret the figures and results.

Reply: Thanks a lot for your suggestions. M-1SD means that the value of a variable is one standard deviation below the mean value. We aimed to explore the moderating effect among independent, dependent, moderating variables by increasing and decreasing the level of moderating variable. In this way, we could reveal that whether the independent variable has a significant positive predictive effect on the dependent variable or not, with moderating variable being one standard deviation below (M-1SD) or above (M+1SD) its mean value. And we would modify this section carefully and make it more concise and accessible.