Dear Cherry Lu,

Thank you for sending the reviews concerning our submission. We appreciate the comments and suggestions, which helped us to improve the manuscript. We revised the manuscript accordingly, including minor language editing, and provided point-by-point answers to the comments.

**Reviewer 1:**

This paper aims to review the most commonly used effect sizes for analyses of categorical variables that use the χ2 (chi-square) test statistic, and introduce a new effect size, Fei, which closes the gap of a missing effect size measure in a correlation like metric that is appropriate for categorical data. In addition, the authors demonstrate the implementation of these measures and their confidence intervals via the effectsize package in the R programming language.

The subject of the paper is interesting and well presented. Authors have demonstrated in a clear and easy to understand way their contribution. Based on the above the reviewer believes that the paper can be accepted for publication.

**Reviewer 2:**

1. Title of manuscript is a little weird. Given that authors proposed the Fei measure, it could be something like "A new effect size for Categorical Data that Use the Chi-Squared Statistic".

Response 1: Thanks for this suggestion. The title was meant as an allusion to the rhyme spoken by the giant in the fairy tale “Jack and the Beanstalk”, which contained two of the effect sizes under investigation in our paper. We’ve added a footnote making this explicit.

2. Line 50: Authors could add some extra lines: "For temporal data, the (cross-)autocorrelation function is considered, and significative autocorrelation is also tested with a chi-square-based statistic (Contreras-Reyes & Idrovo-Aguirre, 2020)". Reference: Contreras-Reyes, J. E., & Idrovo-Aguirre, B. J. (2020). Backcasting and forecasting time series using detrended cross-correlation analysis. Physica A 560, 125109.

Response 2: We appreciate the reviewer’s suggestion to add the above citation regarding the cross-autocorrelation function. However, we feel like it is perhaps too tangential to the goal of the current paper as to include it.

3. Line 66 & 235-236: This letter is not used in mathematical formal works. I recommend to use a Greek letter.

Response 3: We agree that this letter is not yet used in formal mathematical works. However, Hebrew letters are not uncommon in maths, like the letter aleph, or in the field of set theory. We think that all unambiguous Greek letters are already taken in statistics, some even twice over. Furthermore, the Hebrew alphabet has the advantage that its letter forms are very distinct, and not easily confused with Greek or Latin characters. Therefore, we would prefer to stick to our choice and kindly ask the reviewer to reconsider this comment.

4. Line 89: "where O\_i".

Response 4: We replaced “Here” with “where”.

5. Line 90: "k are respectively the number...".

Response 5: We added “respectively”.

6. Line 97: "where p\_{O\_i}".

Response 6: We revised that sentence accordingly.

7. I recommend to authors to move all R codes of the manuscript to an Appendix section. Source codes must not been included in the manuscript's body.

Response 7: We moved the code into an appendix file. Furthermore, the full source code to replicate the results is also available at <https://osf.io/cg64s/> (doi: 10.17605/osf.io/cg64s), as stated in the “Data availability statement”.

8. Add "." or commas where correspond at the end of each formula.

Response 8: Thanks. We added “.” and “,” to the equations where appropriate.

9. Line 147: "where V".

Response 9: We added “where”.

10. Line 163: "For example, in Table 4, each row".

Response 10: Thanks, we revised the sentence accordingly.

11. Line 183: "3. Effect size for the goodness of fit test".

Response 11: Thanks for pointing this out. We revised the headings for section 2. and 3., to make clear that those sections are about effect sizes for those tests.

12. Line 233: Proposed measure is written in terms of chi-square measure (line 87). However, what is the distribution of Fei? It could be compared with chi-square distribution with some degree of freedom? In this form, it could be tested if the association is significative or not.

Response 12: Thank you for this comment. We agree that some “benchmarks” for different effect sizes are helpful to prove our claims that Fei is an applicable effect size measure for the chi-squared statistic. We therefore conducted a simulation study and included the results in the paper (see the new section 4). We simulated data for multinomial distributions for different known true effect sizes of different sample sizes and for different expected proportions. We then compared the distribution of Fei values against the true values and compared those results to an underlying non-central chi-squared distribution. Our results suggest that indeed Fei follows a non-central chi-squared distribution, gives valid results for different multinomial distributions, and allows for power calculation by converting Fei to Cohen’s *w*. All results, figures and additional equations were added to the results section. The code used to generate the simulation data and resulting figures is included in the appendix and available at the OSF.io repository that is mentioned in the data availability statement.

13. Table 6: It is not clear what if the confidence interval. Please specify in formal form the CI.

Response 13: Thank you for this comment. We have added an explanation below Table 6 concerning how confidence intervals for Fei are estimated using the noncentrality parameter method.

14. Finally (and in relation to previous comment), authors could prove the power of the test as a way to measure the test's efficiency. Table 6 only provides 3 examples.

Response 14: See Response 12.

**Reviewer 3:**

15. Line 90. It would be better to have "rows and columns, respectively, of the contingency table"

Response 15: Thank you for this suggestion. We added “respectively” to the sentence, which was indeed remarked by another reviewer, too.

16. Line 103. Although not actually stated, the chi-square value of 454.5 given in this footnote to Table 1 is the "continuity adjusted chi-square" rather than the Pearson chi-square, which is 456.9

Response 16: Thanks, we corrected the chi-square value to 456.9 and revised the R code to produce the correct Pearson chi-square value. We decided to report the unadjusted chi-square to be in line with the formula we used in the manuscript.

17. Line 150. Heading to Table 3. The title to the table is wrong, as the table is about Class/Position, not Sex.

Response 17: Thanks, we corrected the table header.

18. Lines 176−177. More precisely, the generalizations are the square root of the proportions defined, rather than the proportions themselves.

Response 18: Thanks, we have added this in the text.

19. Line 182. Perhaps insert "but" or "and" before "is nothing more" to improve the English.

Response 19: We decided to add a “but” to the sentence.

20. Lines 217 & 221. It might be better if Otherwise started with a lower-case "o".

Response 20: Thanks! We agree with your comment and revised the equations accordingly.

**Reviewer 4:**

The authors proposed an effect size (they called this ‘Fei’) for Chi-square statistic. They normalized the goodness-of-fit test by max chi-square statistic to develop new effect size. The proposed effect size contributes to non-uniform and multinomial variables. The proposed and performance results are shown in section 3.2. For the Journal publication of this paper, the authors must solve the following problems.

21. The authors presented the keywords ‘Fo’ and ‘Fum’ in the paper title. But the authors did not explain them in main text. So, they have to explain them in this paper.

Response 21: See Response 1 above.

22. The description of the proposed effect size presented in Section 3.2 is insufficient. The authors have to illustrate more explanation on how they suggest as well as reference lists. The authors need more diverse experiments and analysis of the experimental results to show the validity of the performance evaluation for the new proposal in this paper.

Response 22: Thank you for this comment. We agree that some “benchmarks” for different effect sizes are helpful to prove our claims that Fei is an applicable effect size measure for the chi-squared statistic. We therefore conducted a simulation study and included the results in the paper (see the new section 4). We simulated data for multinomial distributions for different known true effect sizes of different sample sizes and for different expected proportions. We then compared the distribution of Fei values against the true values and compared those results to an underlying non-central chi-squared distribution. Our results suggest that indeed Fei follows a non-central chi-squared distribution, gives valid results for different multinomial distributions and allows for power calculation by converting Fei to Cohen’s *w*. All results, figures and additional equations were added to the results section. The code used to generate the simulation data and resulting figures is included in the appendix and available at the OSF.io repository that is mentioned in the data availability statement.