

Reviewer Comments-1:-

While the manuscript is framed around a very important topic, cultural / contextual consideration for “normative” practices in XYZ domain (i.e., parenting in this case; spanking, specifically), the study design and data employed rigorously explore such a nuanced, complex topic and research question.

The title is sufficient; however, greater specificity would be beneficial (see highlighted) – “Moderators of maternal spanking and externalizing problems in children”. The article has good English quality and is well-written.

Greater detail in the Methods/Results sections is needed. Additionally, the Discussion could benefit from further discussion of confounds related to the study /date itself.

For example, in the Methods, including a description of exactly how “caregivers reported on the race and gender of the child” is necessary (e.g., provide the exact item wording), given these variables are a main study focus.

Furthermore, how was parental race assessed? This information is needed.

For “child gender,” were parents asked about the child’s biological sex or “gender,” specifically? Or, as is often the case in research, was it presumed based on the child’s outward physical appearance? This information is important because “gender” is not the same as “biological sex” (e.g., chromosomal make-up). Similarly, it would be beneficial for the reader if the “spanking” item (i.e., exact wording of question) was provided, as this would allow for greater accuracy of the current study’s conceptualization (e.g., for comparison with other “spanking” conceptualizations). (Related, it would be beneficial to explore the validity of this item for spanking, or describe how it is indeed valid, given it is a sole focus of the study. Currently, on p. 5, the authors describe the validity of the CTS-OC scales but that is not the same psychometric information as the single item itself.) For the Results section, while the Tables present this information, including the statistical findings (e.g., r statistics and accompanying p values) in-text along with the narrative description (p.6 / 7) for the significant findings is needed (per APA requirements / research reporting and publishing standards).

For example, provide the significant statistics for the spanking – externalizing correlations. Further, the reader is not provided descriptive statistics for Externalizing Behavior at each age (e.g., mean, standard deviations) – this would be very helpful, given it would provide insight into the data distributions themselves.

This comment is also important in regard to the analyses involving the “clinical range on the CBCL” only (p. 7). The authors simply say, “the same pattern is also found,” but provide no data. How many children met “clinical range” T-scores for the CBCL? Was it many? Only a few? Might the relationship (in both samples, full and “clinical-only”) be driven by a handful of children with outlier externalizing scores? Or outlier households with high levels of reported spanking? More information about the data distributions are needed for the reader to better assess the quality of the data. Providing more narrative detail (in-text) about the results of the moderation analyses (e.g., what values are being provided in “Spanking x Race” column in Table A5) is needed. The moderation analyses are not well-described.

Furthermore, in the Results, added discussion about the “practical” significance (e.g., effect sizes) of these findings would be helpful – e.g., while significant, correlations are “weak” (Table A4) and differences across Black and White households (Table A2) are “small” (eta squared); therefore,

perhaps, although statistically significant, the “effect” is not all that important (i.e., “small” effect “size”) or clinically meaningful.

Finally, given the multiple analyses run to examine multiple relationships at varying ages, did the authors perform any statistical correction(s) to avoid Type I or Type II error? Perhaps the findings are not truly significant (Type I) or not truly not significant (Type II error)? In Table A4, with adjusted p-values, would those relationships still be significant at the... .01 level, for example? Last, in the Discussion section, there seems to be no discussion of the (potential in)accuracy of reporting by White households – that is, the possibility that “White households” are dishonest in their reporting of spanking (due to stigma) and spank more than reported. Could this be a confound at play? (Generally speaking, it seemed assumed that spanking was accurately being reported.)

Two minor mistakes that the Reviewer wishes to point out for the authors – For example, “Achenbach (1991)” from in-text was mis-cited in the References section as “Achenbach (1999).” This needs corrected – Achenbach 1999 is not the correct CBCL reference. Additionally, in-text, Mackenzie et al 2011 is cited but no reference provided in the References. Additionally, while all references seemed proper recent and sufficient, as an additional comment...

With any controversial, nuanced and multifaceted topic, there is always some concern about biases, including whether “all sides” and “all voices” on the topic are being represented (e.g., confirmation bias, cherry picking), especially given that the literature itself is inherently biased and rooted in historical, systemic oppression.

While no ethical issues are explicitly noticed, it is true that the authors explore a highly controversial topic and provide findings that (could potentially) be groundbreaking and/or add further noise to this controversy – in this regard, it feels especially important to ensure that the study conducted is of sound, rigorous high-quality scholarship.