29-Aug-2016

Dear Mr. Roberson:

Manuscript ID JPA-16-0121, entitled "Structural Validity of the HBSC Bullying Measure: A Self-Report Rating Scale of Youth Victimization and Perpetration Behavior", has been reviewed.  Comments provided by reviewers are presented below.

Each reviewer noted positive aspects of your research and the associated manuscript. However, numerous concerns were also noted. I invite you to respond to these concerns and revise your manuscript. While I will not restate all of the concerns raised by reviewers I would like to highlight Reviewer 1's concern regarding the apparent disconnect between the EFA results presented and the use of CFA. The correlation of .87 between perpetration and cyber-perpetration, and to a lesser extent the correlation of .7 between victimization and perpetration, indicate common variance and seem to support the two-factor model. A stronger argument is needed to support the four factor model, especially given that the fit criteria used in for CFA are very liberal. Please provide additional rationales and citations to support your use of fit criteria. For example, there are published simulation studies that can be used to support the use of fit criteria. Additionally, please address Reviewer 1's concern regarding the use of ML rather than an estimator such as WLSMV that seems to be more appropriate given items with non-normal distributions. In relation to this concern, consider presenting means, standard deviations, and estimates of skew and kurtosis for the observed variables in this study.

As noted by Reviewer 3, please add discussion of the relationship between bullying behavior and having been victimized by others. Correlations among your latent victim and latent perpetrator variables range from .46 to .59, suggesting that many people are both victims and perpetrators. Finally, the labels for your latent variables are confusing as victimization has a similar meaning as perpetration.

Reviewers' Comments to Author:

Reviewer 1:

This study identified a four-factor CFA model for the HBSC bullying measure, and tested measurement invariance across ethnicity, sex and grade level. The topic sounds to have practical importance. However, I feel that the article lacks details in describing the methods and results, and I have several concerns primarily regarding the research methodologies.

1.      In the methods section, it would be helpful to give an overview of the data analysis procedures and briefly describe each of the statistical techniques such as EFA, CFA, and measurement invariance analysis.

2.      The authors noted that items have extremely nonnormal distributions and used a maximum likelihood (ML) estimation. However, ML requires data to be multivariate normally distributed. With nonnormally distributed data in this study, ML is likely to provide an inflated chi-square statistic and biased parameter estimates. Moreover, it had not been specified whether responses on the five-point Likert scale were analyzed as continuous or categorical data. Given the categorical nature of data, a more appropriate estimation method is robust weighted least square estimation (e.g., WLSMV).

3.      Following the comment above, when ML was used for categorical data with extreme nonnormality, it was not surprised that the ML chi-square as well as fit indices calculated based on the chi-square statistic indicated poor model fit. However, it does not necessarily indicate that the two-factor model did not fit the data. If a proper estimation method was used, the model fit may improve.

4.      When reporting parameter estimates, whether point estimates were standardized or not would need to be clarified.

5.      The EFA analyses indicated two- and three-factor solutions, but a theory-driven four-factor model came up to be the best-fitting model. Statistically, it is expected that with more parameters freely estimated, the model would fit better. This makes me wonder whether the EFA analyses were necessary and worth to be reported, since the final model did not take the EFA suggestions.

6.      The authors used CFI > .85 to indicate adequate mode fit, which for me is a little questionable based on the conversion of CFI > .90 or .95. When the configural models had relatively low CFIs at .86-.87, it may not be needed to conduct subsequent measurement invariance analyses since the configural model did not fit very well. In addition, it would be helpful to conduct the chi-square difference tests for model comparisons.

7.      In Page 11, an internal reliability was mentioned but the measure for this reliability was not known. It would be helpful to clearly define the reliability and provide guidelines for using it.

8.      When conducting measurement invariance analyses, the model fit for each individual group should be tested and reported before making any parameter constraints. Also, the parameters on which constraints were made for each model should be clarified. In addition, the estimated standardized mean differences used for evaluating the factor means were not clearly defined or referred. It is thus difficult to evaluate the appropriateness of the procedure. Moreover, the significance of latent means would need further interpretations.

Reviewer 2:

My review focuses on methodology part as I am not an expert in the content area. Overall, choices of analysis techniques were adequate, analyses were well conducted, and interpretation of the results were appropriate. I have two minor comments, however.

1.      Since the four factor model specified in CFA analysis was a theoretically plausible model, I would recommend the authors should do EFA with the four factor solution to see if the resulting factor structure is consistent with the theory.

2.      In page 13, discussion of IRT (Item Response Theory) and DIF (differential item functioning) was misleading, since DIF analysis would be used for the same purpose as factorial invariance analysis done in the present study. In theory, the two techniques should have the same result with regard to invariance of items and scale across groups.

Reviewer 3:

Introduction

This manuscript is a relatively brief research report of “. . . the latent factor structure of the 22 victimization and perpetration items within the 2009–2010 Health Behavior in School-aged Children (HBSC) self-report survey.” The author(s) are commended for their efforts to add to the literature regarding bullying and the measurement of bullying. The study’s findings are significant and appear meaningful. In addition, the statistical analyses appear solid.  However, there are several concerns that need to be addressed to improve this manuscript.

Areas of Concern

1.      The manuscript is generally well-written (minor typos are noted; e.g. “bullying involvement currently exit”), but the literature review is rather brief and superficial. Although pejorative variables associated bullying have been well-established in previous literature, the authors should very briefly mention some of these deleterious effects.

2.      The authors cite a 2010 reference to suggest “bullying involvement among students frequently goes under-addressed, or completely unaddressed, in schools.” This appears to be dated information given the current public health crisis that is bullying in schools. For example, there are now more than 41 state laws regarding school bullying.

3.      The authors do not mention why they chose to examine the HBSC among the “more than 40 measures are available.” It is alluded to somewhat in the discussion, but should be addressed earlier.

4.      The HBSC assesses for bullies and victims separately (the current manuscript highlights that the HBSC is an “under-researched measure that targets specific bullying victimization and perpetration behaviors”). However, the HBSC does not include specific bully-victim items to measure whether children are simultaneously bully-victims. The authors appear to suggest that the HBSC does measure bully-victim behaviors [e.g., “Assessing perpetration in conjunction with victimization also allows for the identification of students who experience both types of bullying behaviors (i.e., perpetrator–victims). . . .”]. The authors do not address this significant limitation of the HBSC.

5.      The authors examine the HBSC bullying and victimization survey items as a “coherent measure.” Yet, they do not indicate that the HBSC survey developers designed the bullying and victimization survey items to be used as an omnibus measure. Why is it important to examine these items as an overall single scale?

6.      The authors do not indicate in sufficient detail how the sample was identified and selected.  How many schools were approached to participate in the study and how many chose not to participate? How many students were approached to participate in the study and how many chose not to participate? Are there specific exclusionary or inclusionary variables other than grade level? What is the response rate? This information is readily available. Although they may not need to be repeated at length in this manuscript, the authors should at least indicate where the comprehensive information can be obtained.

7.      A serious concern with regard to using the HBSC sample, is that a disproportionate number of participants appear to live in families with above average SES. Yet, the HBSC family affluence subscale was not referenced in this current study.

8.      Research suggests bullying may be under-reported. This was not addressed in the context of the HBSC (cf. Huang & Cornell, 2015; “The impact of definition and question order on the prevalence of bullying victimization using student self-reports.”).

9.      The authors do not acknowledge the multiple limitations associated with self-report measures.