Reviewer's report

**Title:** Psychometric behaviour of the Strengths and Difficulties Questionnaire (SDQ) in the Spanish National Health Survey 2006

**Version:** 2 **Date:** 27 September 2012

**Reviewer:** Carlos Garcia Forero

**Reviewer's report:**

General remarks

The manuscript provides evidence on the psychometric properties of the SDQ, a questionnaire which is becoming widespread in the context of child mental health assessment. Given the uncertain results concerning the SDQ factor structure, a study using a large representative sample would be very valuable and so the manuscript has the potential to be a highly cited paper. The use of a Spanish sample implies an added value to the former publications in Spanish, as the original sample is quite old and is of added value in providing a new, updated sample. However, the statistical methodology is compromised and should be subject to major revisions.

**Major Compulsory Revisions**

My main concern regards the implementation and reporting of factor analyses. The core point is that there is no information about the scale of the variables in the analysis. Were the items considered as continuous or categorical? It appears that the analyses have been considered as continuous, as PASW does not allow ordered-categorical factor analysis, and AMOS 6.0 cannot handle categorical items. So I assume that the EFA and CFA analyses are inadequate and should be repeated with considering the items as ordered categorical. It is well established in the literature (see Muthén, 1984; McDonald, 1999) that the linear common factor model is an incorrect model when dealing with categorical data, which yields biased results in terms of parameter and standard error estimation as well as fit indexes. Ignoring the metric status of the variables seriously threatens the results of the analyses.

Regarding the Exploratory Factor Analysis:

- It is essential to point out that the procedure in the paper is NOT an exploratory factor analysis, but a principal component analysis. The differences between both approaches are substantial, and might be compromising the results. A Principal Component analysis is not the adequate technique here, and the authors must be sure that they conduct an actual Exploratory Factor Analysis model on the variance-covariance matrix.

- The authors have used the Kaiser-Guttman criterion in order to decide the
number of components to retain. The Kaiser rule is among the least accurate methods for deciding on the number of factors to retain (see Humphreys and Montanelli, 1975; Velicer & Jackson, 1990). I’d advice to use of parallel analysis, although other suitable methodologies exist.

- Thirdly, given the results on the correlations among factors from the CFA, the Varimax rotation is not the most adequate rotation to decide the number of factors. An oblique rotation might be a better choice.

Regarding the Confirmatory Factor Analysis:

- A correct factor analytic approach should treat the items as categorical. Otherwise, the use of continuous items seriously threatens the quality of parameter estimates, standard errors and fit results. Factor reliability is also compromised as the latent trait variances are unlikely correct, as it can be seen from reliability estimates below the theoretical lower bound to reliability imposed by Cronbach’s Alpha (the Alpha reliability of the conduct problem scale is above the its factor reliability: this is theoretically impossible when using a correct model).

- The paper does not inform if any method of cross-validation has been used when fitting the confirmatory model. To avoid overfitting, the application of factor analysis require conducting exploratory and confirmatory analyses in at least two random independent subsamples. If analyses were not conducted in this fashion, they should be redone. Otherwise, the splitting procedure or cross-validation technique must be detailed in the methods section.

- I suspect that there might be mistake in the definition of factor reliability provided in the data analysis section, where it is said to be computed as:

$$\text{SUM}(\text{Lambda}^2)/\text{SUM}(\text{ErrorVar}) \ [1],$$

This is an incorrect formula, as it should be:

$$\text{SUM}(\text{Lambda}^2)/( \text{SUM}(\text{Lambda}^2)+\text{SUM}(\text{ErrorVar})) \ [2].$$

Where lambda refers to factor loadings. On the other side, the terminology “factor reliability” is not precise enough for this index, as there are a number of methods for computing factor reliability. If I am correct and the authors used formula (2), the reliability index they used is MacDonald's Omega (McDonald, 1978).

Regarding reliability

- Concerning the use of Cronbach’s Alpha, unidimensionality of a scale is a prerequisite for an appropriate interpretation of Cronbach’s alphas. If the unidimensionality of each scale is not entirely clear, the value of the Cronbach's alpha is not a good indicator of the internal consistency (see for example the article Cortina, 1993). If the scale is not unidimensional, then the value of the Cronbach's alpha can actually not be interpreted properly. My suggestion is to provide other reliability indexes (such as Guttman’s Lambda 2 or model-based reliability) after the factor analyses.
The reliability values are actually low for measurement standards. Specifically, the subscales “conduct problems”, and “emotional symptoms” do not meet the minimum quality standards for reliability, which are accepted to be about 0.70 (Nunnally and Bernstein, 1994). The authors should state this in the discussion, as well as provide the basis (cut-off values and references) for interpreting Alpha values. The interpretation of standard error vs. standard deviation supported by reference [16] is way too optimistic, as a ratio of 1 would imply a classification error about the average of nearly 68% of the population about the scale average. The authors must include a more conservative and sound interpretation for Alpha and standard error of measurement.

The authors computed overall-scale reliability, which is not adequate when several dimensions are involved. If the internal consistency of the scale is of interest, the procedure to compute it must be approached using other methods. I suggest the internal consistency method in Nunnally and Bernstein (1994) applied in the reliability estimations of SF-36. Another possibility is the computation of a model-based internal consistency index proposed by Bentler (2008).

I’m aware that software limitations may be impeding these analyses. However, categorical treatment is a must for obtaining correct model results. I’d suggest using FACTOR (Lorenzo-Seva & Ferrando) for EFA analyses. It would also allow applying the same estimation method as in the subsequent EFA. A nice secondary benefit is that factor reliabilities are automatically computed by the program. I’d like to say that I have no interest on this software whatsoever. However, it is a free, easy-to-use for EFA that allows many options for EFA and has native capabilities for parallel analysis. It also yields a number of very useful fit indexes for model selection. CFA analyses should be addressed using specialized Structural Equation Modelling Software with non-linear factor analysis capabilities, such as Mplus, EQS or GLLAMM. The latest, and very recent, version of R package Lavaan is also capable of handling ordered categorical data.

Minor Essential revisions

1) A limitation section clearly stating the nature of the caveats of this study must be included.

2) The Sample section should provide some detail of the computation of the Lambda weights, however briefly.

3) If the complex sample design is not taken into account in the analyses, the authors must include the understatement of the standard errors in the limitation section.

4) The authors refer to factor loadings as “factorial saturation” throughout the text. This is an incorrect direct translation from Spanish nomenclature “saturación”, where “saturation” and “loading” are different terms for “component” and “factor” loading (which, incidentally, was a differentiation introduced to avoid the confusion between PCA and FA). If the major revisions are undertaken,
“factor loading” should be used instead of “factorial saturation” in all instances.

5) The RMSEA is a standard, useful and widespread fit index, and it should be reported in addition to other fit indexes.

6) Results of the confirmatory factor analysis are not properly reported. Table 4 indicates “Standardized regression coefficient”, but the table is actually reporting standardized factor loadings. Parameter standard errors should also be included. I’d suggest replacing table 4 with a figure depicting the path model for the factor analysis. Such figure would greatly clarify the model structure and it would also inform about factor loadings, error terms, factor variances and factor intercorrelations, thus avoiding table 3.

7) Report all factor loadings in table 2, but mark the highest factor loading in boldface.

8) In page 10, second 4th paragraph, the reporting of Alphas, standard deviation and SEs is confusing. Include a new table informing of scale descriptives and Alpha values.

**Level of interest:** An article of outstanding merit and interest in its field

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**
I declare that I have no competing interests in relation to this paper