Author's response to reviews

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

Authors:

Manuel Gómez-Beneito (Manuel.Gomez-Beneyto@uv.es)
Andreu Nolasco (nolasco@ua.es)
Joaquín Moncho (joaquin.moncho@ua.es)
Pamela Pereyra-Zamora (pamela.pereyra@ua.es)
Nayara Tamayo-Fonseca (nayara.tamayo@ua.es)
Manuel Girón (giron@icali.es)
Mikel Munárriz (mikelmf@comcas.es)
José Salazar (jose.salazar@eresmas.net)
Rafael Tabarés-Seisdedos (rafael.tabares@uv.es)

Version: 3 Date: 13 February 2013

Author's response to reviews: see over
Alicante, February 12th, 2013

MS: 8339710427254223
Psychometric behaviour of the Strengths and Difficulties Questionnaire (SDQ) in the Spanish National Health Survey 2006

Dear editorial board of BMC Psychiatry:
We would like to notify that we have deeply revised the manuscript, according to referees' comments, to whom we would like to thank for the excellent revision.
As a summary of the modifications that have been made:
- The introduction has been importantly extended by adding more evidence.
- Methods section has been modified regarding factor analyses. The indexes of diagnostic efficiency of the questionnaire have also been calculated again.
- The results have been written again, according to the modification of methods.
- The discussion has been adapted to the extended bibliography and to the new results.
Strengths and limitations have been included.
- New conclusions have been added.

Further details about the changes that have been made can be found in the response addressed to the referees.

We hope the format and the content fits in well.
Thank you very much,

Andreu Nolasco

Corresponding author:
Andreu Nolasco Bonmatí at Research Unit for the Analysis of Mortality and Health Statistics, Department of Community Nursing, Preventive Medicine, Public Health and History of Science. University of Alicante. Campus de San Vicente del Raspeig s/n, Apartado 99, 03080-Alicante, Spain, e-mail: nolasco@ua.es, phone number: +34965903926, +34965903919, fax number: +34965903964.
Referee 1
Title: Psychometric behaviour of the Strengths and Difficulties Questionnaire (SDQ) in the Spanish National Health Survey 2006
Version: 2 Date: 31 July 2012
Reviewer: Mari Hysing
Reviewer's report:
The study aims to assess the psychometric properties of the SDQ in a representative sample of Spanish children. The strength of the study is the large sample and well done statistical analyses. The weakness of the study is the introduction leading up to the hypothesis including limited references. I found the article very interesting when I had read it thoroughly, but the authors need to convince the reader already from the introduction.

Introduction
The introduction is short, and while the first paragraph leads well up to the focus of the study, the introduction lacks a short summary of other studies using factor analysis. This is however well covered in the discussion, but would improve the manuscript if including more research in the introduction. Write out more of the Stones et al review results and conclusions would be of interest, and/or some results from some of the studies on factor analysis. One of the main points leading up to the aim of the study is the discrepancies in the factor structure found in a previous Spanish study. However, as my Spanish is not good enough to understand the original article studies referred to, a somewhat more specific description of this discrepancy would be preferred. If there are English versions of the main study this would also be helpful.
A more specified aim/hypothesis of the study would increase interest. The hypothesis underlying the choice of using two different statistical methods for assessing factor structure should be mentioned.

We have extended the introduction. We have analysed the data again according to the suggestions of referee 2. The analysis of the data has been restructured.

Methods
A description of the sample on characteristics such as socioeconomic status, would be informative. It is stated that the sample is representative, but on what variables this concerns would be helpful. You refer to a publication on the main study, but again this is in Spanish and with a very long webaddress, hard to reproduce. Because of this a short description of the methods, maybe including a flowchart could improve this section. If other articles in English have been written on this study you could refer to them.

The sample is representative of Spanish children aged between four and fifteen. The webaddress has been revised. Now is possible to access without problems.
Measures
The description of the SDQ is generally well written. If I understand it correctly, the impact section is not included. This could be mentioned. If the standard text is used, the symptoms referred to have occurred during the last six months. This could be stated as it is relevant in relation to the diagnostic definition you use. Concerning the mental disorder, a consequent use of terms is recommended. Disorders is generally a better term for mental problems than disease. A short description on how you dichotomized the mental disorder variable could be added after the questions.

We have revised the case definition in the data analysis section

Data analysis.
The analysis are well written. However, the description of the mental disorder variable is a little bit confusing and should be detailed. I understand it as an either of the two first questions and then you have to answer yes to the three remaining questions to be defined as having a mental disorder.

We have revised the case definition in the data analysis section

Results
A descriptive table including demographic information of the sample could be included as table 1. As the prevalence of mental disorders are very low, it would be of interest to know more about the numbers. Could gender specific and age-groups be presented on text? This would be helpful in understanding the results even if you use the whole sample in the rest of the analysis.

Our aim was to evaluate the psychometric properties of the questionnaire for the entire child population. We have included information about the distribution by sex.

Discussion
In the present study only symptom scores are included, and from one informant only and the algorithm of probability of diagnosis is not used. I would like to see a discussion of how this may have affected the results. Some limitations of the study could be included, including the definition of the mental disorders and the use of only one informant. The discussion of the factor structure is well written. How can your results affect clinical practice and how we interpret research results using this instrument?

We have extended the discussion and added the limitations

Final remarks
We want to thank sincerely the referees for all of their comments, for we think that they have contributed importantly to the improvement of the manuscript. Thank you very much.
Referee 2

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 of 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.

We are thankful for your comments. The variables must be considered in ordinal scale indeed. The analysis of the data has been performed again using the software FACTOR (suggested by the referee) for the EFA, and LISREL 9.1 for the CFA. In the section Data Analysis of Methods, we describe the new factor analysis, which takes into account variables as ordinal, basing the analysis in the polychoric correlation matrices.

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.

We totally agree with the referee. The EFA has been performed using FACTOR, and "Parallel analysis" has been the test for factor detection. PROMIN has been used as the oblique rotation between factors (methods section)

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).

As we commented before, the analysis has been performed again with an ordinal metric. The reliabilities have been recalculated (Methods section)

- 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.

The new analyses have been performed by dividing the sample in two different random subsamples of the same size, as suggested by the referee. One of them has been used for EFA and the other one for CFA (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: 
  \[ \frac{\text{SUM}(\Lambda^2)}{\text{SUM}(\text{ErrorVar})} \] 
  This is an incorrect formula, as it should be: 
  \[ \frac{\text{SUM}(\Lambda^2)}{\text{SUM}(\Lambda^2)+\text{SUM}(\text{ErrorVar})} \] 
  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
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).

Reliability of each of the factors has been calculated using McDonald’s Omega. Cronbach’s Alpha has not been calculated, because there is no unidimensionality. The calculation has been made by means of software FACTOR (EFA models) and applying McDonald’s Omega index (CFA models) (methods and results sections)

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.

We totally agree. The reference to software FACTOR has been very helpful for EFA and for reliabilities. The comments and suggestions of Dr. Ferrando-Seva (the author of FACTOR) have also been very helpful. CFA have been performed with LISREL. The use of this software has been very important to calculate the polychoric correlation matrixes, using data weightings (methods section).

Minor Essential revisions

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

Limitations are included (Discussion section)

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

There is a comment about the origin of the weights.

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.

There is reference to it in the limitations.

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.

Now we use in all instances “factor loadings”.

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

There is information about the value of RMSEA and other goodness of fit indexes (text and table 4)

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.

*Three figures have been included, as suggested by the referee (Figures 1, 2 and 3)*

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

*Tables 2 and 3 include now all factor loadings*

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.

*Alpha values have not been calculated.*

**Final remarks**

*We want to thank sincerely the referees for all of their comments, for we think that they have contributed importantly to the improvement of the manuscript. Thank you very much.*
**Referee 3**

**Title:** Psychometric behaviour of the Strengths and Difficulties Questionnaire (SDQ) in the Spanish National Health Survey 2006  
**Version:** 2  
**Date:** 28 September 2012  
**Reviewer:** George Giannakopoulos  
**Reviewer’s report:**

Minor Essential Revisions  
The present study represents one of the rather limited research efforts regarding confirmatory factor analysis techniques employed to study a commonly used screening instrument in child and adolescent populations. The methodology is sound and the paper is well-written. However, the authors should take into account more evidence concerning the issue in other European populations, since this data is of relevance, eg see Giannakopoulos et al. Ann Gen Psych 2009, 8:20.

*The introduction has been extended importantly by adding more evidence.*

**Final remarks**

*We want to thank sincerely the referees for all of their comments, for we think that they have contributed importantly to the improvement of the manuscript. Thank you very much.*