Reviewer’s report

Title: Hispanic Dental Beliefs, Dental Attendance, and Oral Status: Additional Psychometric Data for the Spanish Modified Dental Anxiety Scale, and Psychometric Data for a Spanish Version of the Revised Dental Beliefs Survey

Version: 1 Date: 29 January 2010

Reviewer: Mike John

Reviewer’s report:

The manuscript “Hispanic Dental Beliefs, Dental Attendance, and Oral Status: Additional Psychometric Data for the Spanish Modified Dental Anxiety Scale, and Psychometric Data for a Spanish Version of the Revised Dental Beliefs Survey” by Coolidge et al. assesses the reliability and validity of the Spanish version of Modified Dental Anxiety Scale (MDAS) and the Spanish version of the Revised Dental Beliefs Survey (R-DBS) in a nonpatient population.

The research topic is new, interesting, and relevant for the readership of BMC Oral Health. The major thrust of the manuscript seems to be the investigation of test-retest reliability of MDAS scores in nonpatients.

I have the following comments:

The reader could be better educated about the framework for psychometric evaluation of instrument scores. The psychometric properties, e.g., reliability and validity, of scores are linked to a particular population. There is no such thing as the test-retest reliability of a test. Only test scores' test-retest reliability in particular populations can be investigated. The manuscript gives the impression that the authors want to repeat test-retest assessment because it was poor last time. In my opinion, the authors assessed the MDAS score test-retest reliability in a different population. This is new. The authors aim for the assessment of test-retest reliability in a very important population – the nonpatient population - by collecting data from individuals attending a church, etc. Previously, scores' test-retest reliability was assessed in Spanish-speaking college students. Nothing is wrong with assessing scores' test-retest reliability in students, but it is a very particular population and very likely is not a good representation of the nonpatient population at large. To assess test-retest reliability in a more representative sample of nonpatients is a positive contribution to determining the instrument’s psychometric properties. I thought the authors did a good job in educating the reader that their nonpatient sample has a high potential to be generalized when mentioning “These similarities suggest that our study findings are likely to generalize to other Hispanics who are Mexican-Americans.”

Applying the framework that reliability belongs to scores in a particular population and not to the test in general would avoid labeling the previous results as “anomalous”. In my opinion, the previous results are what they are (in the absence of reasons why they should be wrong). The best estimate for score
test-retest reliability in students is an ICC of 0.69. The estimate for test-retest reliability in a more representative sample of nonpatients is an ICC of 0.83. Both results should be correct with an appropriate study design.

Because the major thrust of the manuscript is the assessment of test-retest reliability, additional detail needs to be provided.

First (Major Compulsory Revision), confidence intervals should accompany the point estimates. The reader needs to know to what degree the results are compatible with chance. The same recommendation applies to Cronbach’s alpha. This is also a measure of reliability and I would recommend providing 95% confidence intervals (or lower limits of the 95% CI) for this measure, too.

Second (Major Compulsory Revision), more detail should be given about the statistical model that was used to calculate the intraclass correlation coefficient. I assume subjects were treated as random factor and time was treated as fixed factor when the authors mentioned “Intraclass correlations (two-way, mixed) were used to examine the test-retest reliabilities of the MDAS and R-DBS.” For many readers, the terminology used in the submitted manuscript will not be familiar. I would recommend using the classic article by Shrout and Fleiss (Shrout, P.E., and J. L. Fleiss (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin (86): 420-428.) to inform the reader what statistical method was used.

Third (Discretionary Revision), I would recommend putting the magnitude of the ICC into perspective. I don’t know whether an ICC of 0.69 should be called “relatively low”. There are some guidelines for interpreting reliability coefficients. For example, Fleiss 2003 in Statistical Methods for Rates and Proportions, 3rd edition, p. 604 calls an ICC>0.75 “excellent” and ICCs between 0.4-0.75 as “fair to good”. Calculation of a confidence interval for the previous ICC of 0.69 would help to determine how precise the point estimate was.

The precision of results should be reduced to one decimal.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.