

### **Appendix 3. Critical issues in rating the quality of the studies**

#### *Structural validity*

The rating of structural validity was subjected to several discussions within the review team, as most of the studies did not properly address structural validity. Instead, the majority of studies interpreted the structural validation process too liberally, only testing the unidimensionality of the subscales. We considered this inappropriate when no evidence was presented that the items actually loaded on the proposed domains. While it is acceptable to generate items on specific domains based on a theoretical or conceptual model, the final structure should be tested (e.g. confirmatory factor analysis) and this was rarely undertaken. If no reference conceptual model was used (as it seems for all the PROMs given the theoretical model and the conceptualisation of the constructs/domains were not presented and specified) exploratory approaches are needed. Either way, exploratory or confirmatory factor analysis on the whole PROM and not only the single subscales would be necessary. Similarly, when a single score is obtained from a PROM that claims to measure different constructs, it is necessary that the domains belong to a higher order construct (underlying construct). This raised two concerns regarding consideration of appropriate structural validity i. First, we did not know whether the items really belonged (factor loading) in the proposed factors and if the overall structure of the PROM was confirmed. Second, we did not know if the generation of a single score was appropriate and how to interpret this total score.

#### *Internal consistency*

We rated the internal consistency less strictly as recommended in the COSMIN manual. Indeed, the internal consistency should be secondary to the structural validity since, without knowing the structure of the whole PROM, the interpretation of the consistency is difficult. The internal consistency of a PROM was rated as inappropriate only when the authors calculated the consistency for the whole instrument despite failing to confirm the unidimensionality of the PROM within the same study.

#### *Cross cultural validity*

Difficulties were found when interpreting the cross-cultural validity according to the COSMIN criteria. No studies compared different cultural groups and therefore we evaluated whether the sample was similar to the original developmental study. The data extraction Excel sheet with the specific ratings for each study and properties are presented in the Appendix 2.