

**What will questionnaire measure?**

- What construct should be measured?
- How is construct defined?

**What should questionnaire include?**

- How should content be produced (item generation)?
- How is best item set selected (item reduction)?

**Suitability for potential respondents?**

- Does it appear to measure construct appropriately (*face validity*)?
- Is content a reasonable representation of patient experience (*content validity*)?

**Does questionnaire measure desired construct?**

- Do items measure intended construct (*construct validity*)?
- Can item scores be added together validly (*scalability*)?
- Is level of measurement error acceptable (*reliability*)?
- Is scale able to measure real change in construct (*responsiveness*)?
- Are data collected free from biases unrelated to construct (*differential item functioning; DIF*)?

**Will questionnaire be required for use in new language or culture?**

- Have appropriate translation methods been employed?
- Have new language versions been tested to ensure suitability for local patients?
- Do adaptations have adequate psychometric and scaling properties?