

**Supplementary file 4: Detail of data extraction framework**

Data were extracted in Excel. These are the data extraction category headings from the Excel spreadsheet.

**1. Evidence based on test content**

- 1. Test content evaluated: yes/no/unclear
- 1. Test content: literature review
- 1. Test content: prior existing measures of the construct
- 1. Test content: expert review
- 1. Test content: participant involvement in construct / item development - structured workshops, concept mapping
- 1. Test content: participant involvement in construct / item development - interviews
- 1. Test content: participant feedback processes about items
- 1. Test content: construct description (incl. high/low descriptors)
- 1. Test content: item intent descriptions
- 1. Test content: examination of administration methods
- 1. Test content: other method (e.g., Item difficulty)

**2. Evidence based on response processes**

- 2. Response processes evaluated: yes/no/unclear
- 2. Response processes - respondents: cognitive interviews
- 2. Response processes - respondents: think aloud protocols

Authors: Hawkins M; Elsworth GR; Hoban E; Osborne RH. (2019)

- 2. Response processes - respondents: recording and timing responses to items
- 2. Response processes - users: cognitive interviews
- 2. Response processes - users: think aloud protocols
- 2. Response processes - users: recording and timing responses to items
- 2. Response processes: other method (e.g., determining construct irrelevant factors and construct underrepresentation)

### **3. Evidence based on internal structure**

- 3. Internal structure evaluated: yes/no/unclear
- 3. Internal structure: exploratory factor analysis (EFA)
- 3. Internal structure: confirmatory factor analysis (CFA)
- 3. Internal structure: multi-group factor analysis (MGFA) (SEM, measurement invariance)
- 3. Internal structure: correlation patterns and multi-trait scaling analysis (inter-item, item-total and item-remainder correlations)
- 3. Internal structure: differential item functioning (DIF)
- 3. Internal structure: other method

### **4. Evidence based on relations to other variables**

- 4. Relations to other variables evaluated: yes/no/unclear
- 4. Relations to other variables: convergent validity (between measures of the same or similar construct)
- 4. Relations to other variables: discriminant validity
- 4. Relations to other variables: test-criterion relationships (how accurately test scores predict criterion performance)

Authors: Hawkins M; Elsworth GR; Hoban E; Osborne RH. (2019)

- 4. Relations to other variables: group differences (relationships with background characteristics such as demographics information)
- 4. Relations to other variables: validity generalisation (e.g., meta-analyses / statistical summaries of past studies; cumulative databases)
- 4. Relations to other variables: nomological networks
- 4. Relations to other variables: other method

**5. Evidence based on validity and the consequences of testing**

- 5. Consequences of testing evaluated: yes/no/unclear
- 5. Consequences of testing: methods to test for consequential validity (intended consequences e.g., benefits)
- 5. Consequences of testing: methods to test for consequential validity (unintended consequences e.g., negative effects)
- 5. Consequences of testing: methods to test for construct underrepresentation
- 5. Consequences of testing: methods to test for construct-irrelevant components
- 5. Consequences of testing: methods to test for claims made beyond the intended score interpretation
- 5. Consequences of testing: methods to test for consequences for clinical implications
- 5. Consequences of testing: other methods to test consequential validity
- 5. Consequences of testing: other method (e.g., fairness - low/high-stakes consequences)

Authors: Hawkins M; Elsworth GR; Hoban E; Osborne RH. (2019)