Instrument Guide Developed by Penelope Schofield, Anna Ugalde & SCCV - 31 August 2009

Points to consider when selecting a tool / questionnaire

- Identify what you want the tool to do e.g. to flag “at risk” people, to diagnose a condition, to assist in assessment.
- Identify the areas or issues you hope / require the tool to include e.g. the 5 domains of supportive care.
- For each area / issue review the questions on the tool and decide if the questions will provide you with the level of information that you require. For example the Distress Thermometer without the Problem Checklist provides a single item scale whilst the Hospital Anxiety and Depression Scale (HADS) provides two sub-scales (for anxiety and depression separately) which would generally provide a more complete picture of the issue examined.
- Review the literature to identify what has already been developed and psychometrically validated.
- Identify the process and identify who will complete the tool i.e. the patient or the health care clinician.
- If you wish to repeat the measure at set time periods, you need to consider the time frame the questions refer to and ensure you leave enough time between repeated completions so that the time frame does not overlap.
- If the tool is to be used clinically it should be simple to understand, easy to complete, and quick to use.
- Determine how the reports from the tool can be made readily available and in an informative way for use by the clinician1.
- Identify if permission from the authors of the tool or payment (such as for the Hospital and Anxiety Depression Scale) is required prior to use.
- Determine the evidence regarding the psychometric properties (reliability and validity) of the tool (weak – strong evidence) A variety of strategies may be used to assess psychometric properties of instruments, and the strategies of selection are determined by several factors, such as measurement purposes, context (e.g. population or setting), and practical issues1.

What is validity?

Validation is the process of evaluating whether a questionnaire is measuring what it should2. There are several types of validity:

- **Construct validity** is a measure of how well the tool measures the underlying construct e.g. pain, social functioning3. There are three types of information regarding construct validity, these are convergent, divergent and discriminative validity.
  - Convergent validity is comparing scores on the new scale with scales measuring the same construct (for example, DT and HADS).
  - Divergent validity is comparing scores on similar but different construct (for example comparing the treatment related anxiety and the Anxiety subscale of the HADS).
  - Discriminative validity is comparing two distinct groups of people who are known to be different (eg comparing those who have completed curative treatment with those who are receiving palliative treatment).

- **Content validity** is appropriate particularly for newly developed measures1. Content validity can be provided by experts who identify if the questions are logical and relate to the issue being addressed.

- **Criterion validity** is another property that can be examined when a “gold standard” (such as psychiatric assessment of psychological morbidity) is available1. “Criterion validity refers to the ability of the questionnaire to measure what it should measure.”3 p.1798

What is reliability?

Reliability of a tool relates to the consistency and stability of a measure over time and covers three main areas:

- **Internal consistency** is a measure that reflects whether the questions within one domain or subset relate to each other, for example, that questions within the domain of social need all relate to social need.

- **Test-retest reliability** is measured by comparing the differences in scores when the questionnaire is administered over two time points. A high level of test-retest reliability would indicate that the questions have a level of clarity, such as when people respond in the same way to issues that remain stable over time.
Instrument construction and psychometric validation

Inter- and intrarater reliability is relevant for observational measures and is aimed to be used by different types of raters (e.g., clinicians or research team members). This type of reliability indicates that at least two people perceive the observation (such as, use of open questions) in a similar way and provide similar responses.

Points to remember if changing part of a tool previously identified as reliable and valid

Changing the time frame

If the questions refer to a person’s experience over a certain time frame (e.g., a year or a week) any change to this timeframe may adversely affect the psychometric properties of the tool.

If there is some concern about the time frame of a tool, further testing to include comparisons of results from other tools using similar time frames is required to ensure that altering the time frame for the tool does not compromise the internal consistency and validity of the tool.

Translating the tool

Translation of tools into other languages does not ensure that subtle cultural nuances are captured by the translated instrument. Instrument adaptation has to be sensitive to the various conceptual and cultural meanings. Several strategies can be used to further test the adequacy of the resulting instrument. Convergent validity demonstrated across samples of different cultural backgrounds is one strategy.

Changing the Scale(s)

If there is evidence demonstrating that a scale is reliable and valid the scale cannot be altered in length, anchor words, direction, orientation or colour without further testing of the validity and reliability of the altered scale. Further testing would require a research project to be undertaken. An example of the methodology that could be used is described in a pain project where various pain scales were compared for use in a specific target audience. The Verbal Descriptor Scale, the Faces Pain Scale (FPS), and the Numeric Rating Scale pain intensity scales were compared in a population of older adults residing in a nursing home. Results indicated accurate assessment of pain intensity in nursing home residents can be accomplished using different standardized pain-intensity tools.

Example of steps involved in constructing a reliable and valid tool / questionnaire

- Ask the target audience to identify issues to be included in the tool – usually by interview or focus group.
- Refine the number of issues generated by the interviews / focus groups into a list of items (questions) for the tool.
- Select a response scale. This involves first identifying the nature of the items as either being a categorical judgment (e.g., yes/no) or a continuous judgment (e.g., 1 to 10).
- If it is a continuous judgement then:
  - Is it unipolar (e.g., no distress to extreme distress) or bipolar (happy face - neutral face - sad face).
  - Should points on the scale should be labelled or unlabelled.
  - Establish a timeframe for the person to report on (e.g., today, in the last week, in the last month).
- If the stem (i.e., the wording of the item) response scale or instructions are altered in any way then the item(s) may be interpreted differently and may compromise the validity of the questionnaire.
- Following this, the entire instrument requires reviews to establish suitability of length, readability, acceptability, clarity and face validity.
- Formal psychometric testing is then performed to establish reliability and validity.

References