Evaluation Checklist for Clinical Measures

Prepare a handout for each member of this class (including the instructor). The handout must follow this outline including headings. If you could not find some of the requested information listed under a heading, make note of it in this handout.

I. General Information
   A. **Identifying Information** - Title, Authors, Edition, Forms (if applicable)
   B. **Instructions** - Directions for administration and scoring. This includes describing how test items are scored \(^1\) and how a total score is calculated (and, if present, calculation procedures for subscale scores should be included as well).

II. Nature and Purpose
   A. **Construct** - Does the questionnaire restrict itself to a single construct (e.g., parental self-control) or is it a multidimensional questionnaire (e.g., a parenting questionnaire with subscales measuring parental characteristics including knowledge of child development, drug and alcohol use, self-control, involvement, encouragement of maturity)?
      You should explicitly identify the construct (or constructs) the questionnaire is supposed to measure. This should include the definition of the construct as given by the designers of the questionnaire (not a dictionary definition).
      If there are subscales that measure component constructs, these constructs must be identified and defined as well.
   B. **Target Population** - Identify the group or type of individual for which this questionnaire is designed. This is usually not quite the same as the populations for which the instrument was normed.
      This refers to individuals for whom the construct being measured (e.g., parental self-control, depression, adequacy of environment) is relevant.
      However, it may include such characteristics as age range (e.g., an instrument designed for children), culture (e.g., an instrument specifically designed for members of a particular cultural group), etc.
   C. **Content** - The types of items used and the way in which the testee is supposed to respond to the items.
      Items usually are questions or declarative statements. Testees usually answer in the form of scale \(^1\) ratings.

III. Psychometric Information
   A. **Norms** - Identify the populations used in the standardization process.
      Sometimes one population is used for standardization purposes in the initial development of an instrument and it is subsequently “normed” for other populations.
      Please include all the populations evaluated.
   B. **Reliability** - Present the reliability coefficients \(^2\) reported for the questionnaire.
      For each coefficient you report, note the type of reliability coefficient (name of the procedure used to compute reliability). These should include the coefficients for the samples used to create norms for the populations you reported on in the previous section.
      Retest and Coefficient Alpha are two particularly important coefficients. Report on their relationship with each other and the implications of this for using the instrument for repeated measurement.
   C. **Validity** - Report on evaluations of validity \(^3\) for your questionnaire.
      Again, if these data are available for different “norming” samples, this information should be included.
**IV. Ease of Use** – This is influenced by the time required to fill out the questionnaire and the time needed to score and interpret it. If available, the reading level of the questionnaire should be reported here. These relate to how much effort \(^{(4)}\) would be associated with use of the instrument. If specialized training is needed to administer the instrument, the type and duration of the training should be reported here. Note whether this instrument would be suitable for use as a repeated measure. Instruments that do not require much time or effort to use are more appropriate as repeated measures. Repeated measures should also be sensitive to change \(^{(5)}\).

**V. Accessibility** – This takes in a number of issues. Tell how to order the instrument and how much it costs. Sometimes one has to order a certain number of forms so you should discuss the overall cost as well as the per form cost. If there is a manual, tell who to obtain it and its cost. Certain instruments are only available to individuals with certain credentials (e.g., clinical psychologists, education specialists) or require specific training (e.g., a course in test interpretation). Note whether or not these criteria are present and, if so, describe them.

**VI. Summary** – Summarize the major strengths and weaknesses of the instrument across all parts of the outline.

**VII. Reference** – List references consulted for this report. The reference list should follow APA style requirements.

**VIII. Appendix** – This should contain a copy of the instrument.

---

**Notes**

(1) Items in questionnaires may be scored using one of a number of scales.

- **Dichotomous scales**
  Items are scored based upon a dichotomous response (usually 0 for one answer and 1 for another, or 1 for one answer and 2 for the other).
  (e.g., 1 = true – 0 = false; 1 = applies to me – 0 = does not apply to me; 1 = agree – 0 = disagree; 1 = yes – 0 = no)

- **Likert-like scales**
  Items have several ordered response categories (usually five to seven) indicating degree of agreement with the item content.
  Items are scored according to response category where response categories indicating less agreement are given a lower score and those indicating more agreement are given a higher score.
  (e.g., 0 = strongly disagree, 1 = disagree, 2 = slightly disagree, 4 = neither agree nor disagree, 5 = slightly agree, 6 = agree, 7 = strongly agree).

- **Intensity rating scales**
  - **Categorical intensity rating scale**
    Ordered response categories are provided for each item (usually five to seven) indicating the magnitude of intensity of the thing described in the item.
    Items are scored according to the intensity category where an intensity category indicating lower intensity is given a lower score and a category indicating more intensity is given a higher score.
    (e.g., 0 = none at all, 1 = slight, 2 = moderate, 3 = considerable, 4 = extreme)
• **Numerical intensity rating scale**
The response options are a range of numerals representing minimum to maximum intensity of the thing described in the item. There are equal intervals between the numerals – from 0 to 10 represents the same change in intensity as from 10 to 20, from 20 to 30, and so on. Zero represents absence of thing described in the item. A maximum score (usually 10 or 100) represents the maximum possible intensity of the thing described in the item. (e.g., SUDs-type scale: 0 = not present, zero intensity; 100 = maximum possible intensity)

• **Psychophysical intensity scale**
A line is provided anchored at a minimum on the left and a maximum on the right. The individual filling out the questionnaire makes a mark on the line to indicate the intensity of the thing described in the item. The item score is the measured distance between the left end of the line and the mark placed by the testee (usually in decimal fractions of centimeters or inches).

<table>
<thead>
<tr>
<th>Weakest Possible Intensity</th>
<th>Strongest Possible Intensity</th>
</tr>
</thead>
</table>

• **Frequency rating scales**
  • **Categorical frequency rating scale**
  Ordered response categories are provided for each item (usually five to seven) indicating the frequency of occurrence for the thing described in the item. Items are scored according to their frequency category where an frequency category indicating less frequent occurrence is given a lower score and a category indicating more frequent occurrence is given a higher score. (e.g., 0 = none of the time, 1 = rarely, 2 = some of the time, 3 = a lot of the time, 4 = most of the time, 5 = all of the time)

• **Frequency counts** (i.e., the actual number of times the event described in the item occurred)

Of course, these do not exhaust the response possibilities, but I hope they give you an idea what to look for.

(2) The types of reliability coefficient are
  • **Retest (Test-Retest)**
  • **Alternate Forms (Parallel Forms)**
  • **Internal Consistency**
    • **Split-Half**
    • **Coefficient Alpha (Cronbach’s Alpha)**

(3) Validity refers to how well an instrument measures what it is supposed to measure.

  **Face Validity:** This is regarded as the simplest form of validity. It refers to whether the items in the questionnaire seem to be “getting at” what they should; whether the items in the questionnaire seem relevant to the construct being measured.

  **Content Validity:** This refers to evaluating the extent to which the items in the questionnaire cover the domain of content described in the definition of the construct (see the Construct and Directness section). It is a more rigorous evaluation of validity than face validity. It requires a carefully constructed definition of the construct being measured. This definition must describe the domain of content covered by the construct well enough to make clear what characteristics are included and what are not. The items in the questionnaire are evaluated to ensure that each aspect of the domain of content is represented in the questionnaire.

  **Criterion Validity:** This refers to the extent to that the results of the questionnaire’s measurement of the construct agree with the results of other measures of the same construct. Types of criterion validity include *predictive* validity and *concurrent* validity.
(4) If it takes a client more than five minutes to fill out an instrument, it would be less useful as a Rapid Assessment Instrument. Similarly, if scoring the instrument takes more than three minutes, its utility as a Rapid Assessment Instrument would be limited. (Although such instruments might be useful for other purposes.) Instruments requiring lower reading levels are usually easier to use. (Although oversimplification may create difficulties as well.)

(5) To be sensitive to change in what it is measuring, an instrument’s retest reliability coefficient must be less than its internal consistency reliability coefficient. The preferred internal consistency coefficient would be Coefficient Alpha (also called Cronbach’s Alpha). Further, the instrument ought to have high internal consistency reliability (Coefficient Alpha ≥ .90).