Answers

1. Validation is a continuous process

1. An instrument should be validated again if it is applied in another situation (e.g. for another purpose or in another target population).

2. In the case of new instruments, there may be a further development in theory, which provides opportunities to test stronger hypotheses. Moreover, there may be new empirical evidence, which enables other, perhaps more specific, hypotheses to be tested.

2. Drawing conclusions about construct validity

Option a is the only correct conclusion, because the assumption is that the hypothesis is correct. Options b and c are explanations why the hypothesis was not confirmed.

3. Formulation of conclusions about validity

The formulation of a conclusion about the validity of an instrument should at least contain information about: the type of validity studied; the population or situation, and the purpose of the instrument. Only conclusion e includes all these components. Conclusions a, b, c, and d lack one of these components.

Conclusion f is correct, because in the case of good criterion validity, it does not matter whether the instrument is used in a discriminative, evaluative or predictive application. If the gold standard can be used in a specific situation, a measurement instrument showing good correspondence with the gold standard can be used just as well in that situation.

Conclusion g may be a valuable part of the assessment of content validity. However, it does not include comprehensiveness, i.e. the assessment whether all the aspects of the construct are covered by the measurement instrument.

4. Validation of a Short Form version of the WOMAC versus the Long Form version.

a. The disadvantage of approach 1 is that there are two different study samples, so no within patient comparisons can be made. The disadvantage of approach 2 is that the SF version and the LF version are not assessed independently from each other. In fact, the SF version was totally incorporated in the LF version. So the agreement between the SF and LF versions is probably over-estimated.

b. Approach 1 involves two different sub-populations. Therefore it is not possible to construct a Bland and Altman plot which requires the subtraction of scores of the same patients for the two instruments. This is possible for approach 2, but, as stated under a, the agreement is over-estimated. For approach 1, means and SDs can be compared, assuming that, in a random split, both groups have the same level of physical functioning, and therefore the means and SDs are expected to be the same. However, the researchers did not state that they split the sample at random.

c. From earlier studies the researchers must have had ideas about the magnitude of correlations with these variables, so they should have included a statement about the magnitude of the expected correlations in the hypotheses. They could also have stated how much more correlation they expected with the SF version, and could have hypothesized that they would find a higher correlation with the assessment of physical functioning than with the pain assessment.

5. Interpretation of data on measurement invariance (DIF)

a. DIF means that people with the same ability (level or score on the construct) respond differently to an item. For example, Teresi and Fleishman (2007) found that at the same underlying level of physical functioning, African-Americans are more likely than whites to report that they have 'no trouble with a long walk', and this difference appears consistently across the continuum of physical functioning. This is called uniform DIF. It would have been non-uniform DIF if, at high levels of physical functioning, the African-Americans would have been less likely than the whites to report 'no trouble with a long walk' than whites, or, at lower levels of physical functioning, the African-Americans would have been more likely to report 'no trouble with a long walk'.

b. DIF assessment is a strong tool for cross-cultural validation, because it points to different interpretations of an item, either due to a poor translation or to cultural differences between the target population and the original population. Note that DIF may also occur when applying the instrument in another target population, because patients with different diseases may interpret items differently. For example, the term 'long walk' in the item 'Do you have trouble with a long walk?' might be interpreted differently by patients with diseases that affect their walking ability than by patients with diseases that do not affect their walking ability.

c. A '+' and a '- 'in the same cell would mean a totally different interpretation of items at the three time-points. This is difficult to explain, and would make the instrument inappropriate for longitudinal assessments, to say the least. The occurrence of '+' and 'o', or '-' and 'o' in the same cell is less serious.

Small shifts might be explained by a slight degree of DIF, around the border of significance, between the populations. Three times ‘+’ or three times ‘-‘ in the same cell indicates a consistent finding over the three time-points and strengthens the conclusion that DIF occurs in these items.