Questions and answers for Chapter 4

1. What can you do to make your instrument more valid?

   When we want to design a research instrument, such as a questionnaire, we need to make sure that we try and maximise its validity. The first thing we can do is to make sure that we are aware of the literature around the concepts we want to study, and start off with a theoretical framework that will influence the design of the survey. It is a good idea to get some experts in the area to look at and comment on the instrument. More importantly, you need to pilot the instrument with a group of respondents similar to the sample you want to use. If, for example, you were going to survey 20 year six classrooms, it would be a good idea to pilot the survey in at least three classrooms. The pilot would be used to look at any problems, such as items which don't appear to have been understood or items with no variance (everyone gives the same answer). It can be useful to also interview or conduct focus groups with your pilot pupils or a subset thereof, to get some more detailed information about possible changes to the instrument, and what they feel has or hasn't worked. If possible, it is helpful to collect data on factors that we think might be related to our instrument (for example, if we are looking at academic self-concept we could collect pupil grades), so we can test predictive validity.

2. What do you think about the effect size v. significance test debate: should we stick with significance levels, or replace them by effect size indices and confidence intervals?

There is currently quite a debate about this in the statistics community. What pretty much everyone agrees on is that it is not sufficient to rely only on significance levels, as they are determined in part by sample size and therefore are a poor guide of the strength of a relationship. Therefore we should in all cases calculate an effect size measure as well as a significance test. Some researchers say that, in part because significance tests are so often misinterpreted as effect size measures, and because of the arbitrary nature of the cut off points, it would be better to not use significance tests at all. Most researchers, however, feel that they still have a useful role to play alongside effect size measures.

3. How would you calculate whether or not your test was reliable?

There are two main ways of calculating internal consistency reliability: split half reliability and coefficient alpha. Split half reliability works as follows: say we have an attitude to teaching measure that consists of 10 items. First, we randomly split the test into two (for example, the even and uneven items). Then we calculate respondents' scores on each 'half test'. We can then see whether the two scores are related to one another. If they are both measuring the same thing, we would expect them to be strongly related, with a correlation coefficient of over .8. Coefficient alpha is another measure of internal consistency. This measure would be expected to be over .7 before we can say that our test is internally consistent.

4. Do you think a more reliable test is automatically more valid?

Reliability doesn't automatically lead to validity. One particular problem in this regard is that we can often increase reliability by ensuring that we measure narrowly and precisely. This can sometimes conflict with validity, as this narrowness may not theoretically describe the phenomenon well. However, if our measure is unreliable, it won't be valid either, as how can we measure what we want to measure if we are measuring it badly?

5. What types of error can you make when accepting the alternative hypothesis?

If we accept the alternative hypothesis, but the null hypothesis is true in the population we have type I error.

6. How can you make your instruments more reliable?

Firstly, we need to ensure that the quality of questions we ask is high and that all questions are unambiguous. Unambiguous and clear questions are likely to be more reliable, and the same goes for items on a rating scale for observers. Another way to make an instrument more reliable is by measuring it with more than one item. That way any errors that respondents make on an individual item can be cancelled out.