eleven

COLLECTING QUANTITATIVE DATA

Measurement

There are six key points I want students to get from the discussion of measurement in this chapter:

1. Students need to understand the process of measurement itself. This is important because there is so much misunderstanding of measurement, and because the issue of measurement has been at the heart of the historical disputes (see Chapter 2) between quantitative and qualitative researchers. Indeed, measurement has been called the 'bête noire' by some qualitative researchers opposed to its use. Too often (but not always), I think this sort of comment and criticism is based on ignorance and misunderstanding. Measurement is the process by which data are turned into numbers. Numbers do not occur 'naturally' – as I say to my students, the world we wish to study does not occur in the form of numbers. It is we as researchers who make the decision to turn our data into numbers, so that numbers are man-made, not 'God given'. My wish is that we make this decision consciously and deliberately, and after careful consideration.

   I stress that measurement is not some foreign or strange activity. Rather, it closely resembles – and formalises – what we do in everyday life. I encourage my students to think of it in this way: two types of activities produce numbers or turn data into numbers – one is counting and the other is scaling. Nobody finds counting problematic – we do it all the time in everyday life – and we should note that even the most dyed-in-the-wool qualitative researcher finds it useful to number (i.e. to count) the pages. It is worth pointing out, however, that we are always counting with respect to something. Scaling, on the other hand, seems to be where all the problems lie, but, here again, it resembles very closely what we do all the time in everyday life. We constantly, in our thinking and talking, use notions of scale, and the location of some object (or person, event or whatever) of interest along that scale. We don't normally, in our thinking and talking, assign numbers to these locations. Measurement does. That is, measurement formalises this normal, common, everyday process and assigns scale values – in the form of numbers – to its outcome.

   (It is good to have a 'library' of examples of the use of scaling – without the numbers – in everyday situations. Thus, for example:

   - This film is more interesting than that.
   - This dish is more tasty than that.
   - This is the most impressive paper I've read this year.

Another useful exercise is to take some writing – from a newspaper, a journal article,'a book – and locate examples, often implicit, of such 'scaling' statements.)

I want students to develop this full understanding of the process of measurement, so that they can critically assess its role in any particular piece of research, and also so that when they choose to use it, they can use it in an informed way. This is particularly important in those situations where students choose to develop their own measuring instrument (see below). I have seen too many examples of students who want to do this, without having a good understanding of measurement itself.

None of this is to say that we should or must measure in situations where we have a choice. Rather, we should rationally assess each research situation and see first whether measurement is possible and feasible, and second whether it will help us to understand the social reality we are studying.

Nor is it to suggest or imply that measurement can give us the full picture of the social reality we wish to study. On the contrary, I point out that there are limits to what we can learn about something by measuring it – or, more accurately, aspects of it. We should understand both the uses and advantages of measurement – which are very real – and its limitations. This way, we can use it more intelligently in our research. To round out the discussion on this point, it is worth spending some time looking explicitly at the advantages measurement brings. I have detailed these in the chapter itself.

2. Students also need to understand the concept of latent traits and what this concept implies for our attempts to measure in social science. It is of fundamental importance, and latent trait measurement theory lies behind the vast majority of our attempts to measure variables in social science. In passing, we should note again that it is another example of different levels of abstraction. Figure 11.1 shows that the latent trait (intelligence, in the example given) is at a higher level of abstraction than the items (or indicators) used to measure it.
3. Given the very widespread use today of Likert summated rating scaling (rather than the other forms of scaling mentioned in section 11.4, which were historically important, especially Thurstone and Guttman), it is important that students know how this method works. It is so ubiquitous that virtually all students have had some exposure to it, but it is worth going through the basic ideas behind it. Thus:

   - Multiple items are designed, or selected, as indicators of a uni-dimensional trait (this follows from latent trait theory); if the trait itself is multi-dimensional, it must be defined in terms of its uni-dimensional components, and each uni-dimensional trait is then 'operationalised' in terms of its items.
   - Each item is rated by respondents – the most common response scale is agreement–disagreement, but many other response scales are possible (see the separate point below about the number of points on the response scale).
   - The ratings are assigned numerical values.
   - These individual item vales are then summed across the uni-dimensional trait – hence the method is known as the 'method of summated ratings'.

   Describing the method in this way shows the critical importance of dimensionality and of uni-dimensionality in particular. This is another central quantitative concept it is worth spending time on, so that students understand its meaning and importance. The operation of addition, or summing, of course depends on it – we can only add 'like with like'. The technical aspects of investigating and assessing uni-dimensionality can come later. At this stage, it is important to have a conceptual understanding of it.

4. The very common situation in research planning of whether to use an already existing measuring instrument, or whether to develop one specifically for the project being planned, needs to be discussed. The points I stress with students about this issue are:

   - Developing a measuring instrument to produce good numerical data is not easy and substantially increases the amount of work involved.
   - Developing ad hoc rating scales is much easier than developing a measure of a 'major' variable and can be done effectively using the procedures described in section 11.5; again, however, substantial work is involved.
   - A very large number of measuring instruments have been developed by researchers over the years and it is always worth checking to see what exists. The problem is finding the measures which have been developed (hence the importance of *The Mental Measurements Yearbook, see p.236, 237*).
   - If an already existing measuring instrument can be found, and it fits the conceptual definition(s) of the variable(s) in this study, and it has sound psychometric properties, the advice is to use it, rather than trying to develop one's own instrument.

5. I stress the need to know the psychometric properties of a measure, as a necessary part of high quality quantitative empirical research. Coming before this, of course, is the need to know what the term 'psychometric properties' means, and its relationship to the all-important concept of the quality of data. At the most basic level, the key psychometric properties are reliability and validity, and students should know what these technical terms mean, on a logical basis. At a more advanced level, the concepts of dimensionality and sensitivity are involved. These psychometric properties are known for well-established instruments. For new instruments, they need to be assessed.
6. The quality of data issues needs discussing, especially as addressed in section 11.10. There are really two aspects to the quality of quantitative data. One concerns the psychometric properties of the instrument, especially reliability and validity, as mentioned. The other concerns the way the measuring instrument is administered and the possible (or likely) effect of this method of administration on the quality of data. This latter point applies particularly to survey questionnaire data. There is nothing esoteric about the point being made here, as it really is common sense. But it is surprising how much it 'slips under the radar'. This is probably because we do not pay enough attention to the quality of data issues in research planning with our students. I stress two points here:

   - that empirical research is only as good as the data on which it is based – a point which applies equally to qualitative and quantitative research
   - that it is better to have a small quantity of high quality data than a large quantity of poor quality data.

     I want students to think hard about the quality of data before they collect the data, while they are planning the research – to carefully consider the options as to how a survey questionnaire can be administered, to assess the strengths and weaknesses of the different possible methods, and then to decide on their method of collection, making quality of data the overriding criterion.

Sampling

There is only a short section here on sampling in quantitative research. This is not because I do not consider it to be important. It certainly is important. Rather, it is because in today's world, especially in the world of student research, we so often

have to take whatever sample(s) we can get. In view of this, I don't think it is worthwhile spending a long time on theoretically elegant 'ideal' sampling plans, when they so often have little chance of practical implementation.

In this section, I want students to:

- understand the important distinction between probability sampling and deliberate (or purposive) sampling
- know and be able to articulate their sampling strategy and be sure it fits in with the overall strategy behind the research
- deal with the three questions shown in section 11.11:
  - How big will the sample be and why?
  - How will it be chosen and why?
  - What claims will be made for its representativeness?

How many points should there be on a response scale?

There is no rule about how many points there should be. The only rule is that all points should lie on the same scale. Perhaps this is best illustrated by the following mistake which can be observed (far too often) on survey questionnaires:

> Assume that the Likert summated rating scaling is being used, with an agree–disagree response scale. The response scale is set up as follows:
>
> | | |
> |---|---|
> | Strongly agree | score 5 |
> | Agree | score 4 |
> | Don't know | score 3 |
> | Disagree | score 2 |
> | Strongly disagree | score 1 |

The problem is the middle category (don't know, score 3). Research has shown that very often this point does not lie on the same scale and therefore should not be scored 3. (Sometimes, the words used for the middle category are different – 'undecided', 'neutral' or 'neither agree nor disagree'. No matter which words are used, the problem remains.) Because this middle point does not lie on the scale, we have now introduced 'noise' into our measurement, which is not ideal. In other words, the quality of our data is not as good as it could be.

There are several ways we can avoid this problem:

- We can delete the middle category, reduce the scale to four points and make sure (through pilot testing) that we have items respondents can easily agree or disagree with.
- We can change the bipolar agree–disagree scale to a unipolar agreement scale. Instead of previous wording, we can ask: 'To what extent do you agree with the following statements?', and we can now change the response scale to:

Completely	score 5
Mostly	score 4
To some extent	score 3
Very little	score 2
Not at all	score 1

These five points now lie on the same scale.

So the issue is not about how many points lie on the response scale – it can be four, five, six or some other number; it is about making sure all points lie on the same scale.

A useful form of wording for Likert scaling, which also avoids the bipolar aspect of agree–disagree, is to phrase items and responses in terms of frequency of occurrence. Thus, if the question to respondents is 'How often would you agree with the following statements?', the response scale could be:

100% of the time	score 5
75% of the time	score 4
50% of the time	score 3
25% of the time	score 2
0% of the time	score 1

The word 'approximately' can be placed in front of each response alternative, but all points now lie on the same scale, and this scale has the additional advantage of the equal spacing of the points being quite visible.