How can we help our students be more critical? Examining the details in questionnaire studies

James Hartley

IN A RECENT PAPER a colleague and I discussed the difficulties of using questionnaires to assess the efficiency of new instructional methods (Hartley & Cabanac, submitted). In that paper we commented on nine issues. They were:

- How many researchers reported the use of questionnaires in this context.
- How many of them developed their own scales or used pre-published ones.
- What were the typical features of these questionnaires – in terms of:
  - The number of items;
  - The number of scale points;
  - The layouts of the scales;
  - The direction of the scales – from positive to negative or vice versa;
  - The use of negatively worded items and reverse scoring;
  - The proportion of badly worded items.

In this paper I continue to discuss these concerns about the use of questionnaires and suggest that psychology teachers can help improve the teaching of psychology by drawing students' attention to them *with specific examples*. In particular I illustrate my arguments by discussing three recent papers in this respect.

Paper 1: Sharp, Hemmings and Kay (2016)

Sharp, Hemmings and Kay (2016) describe the construction of a Likert-type scale to measure boredom in university students in the UK. The final version of the first part of their scale had 28 items. Respondents were asked to respond to these items using a five-point scale: *A = always, B = usually, C = occasionally, D = rarely* and *E = never*.

Unfortunately there are at least three problems here:

1. There is one item I don't understand. Item 27 reads, 'At university it seems that we do the same things all the time: it's getting old'. I thought at first that this was a misprint, but the wording is explained on p.659. It still seems an odd (and thus a poor) item to me.
2. There are at least 14 statements where there more than one issue is raised in the wording of the item – making it difficult to give a single answer. For example: 'At university, I find it easy to concentrate on my work and other activities.'
   'Having to read someone else's coursework or watch their presentation and listen to what they have to say bores me tremendously.'
   'I find it easy to entertain and motive myself at university.'
   'At university I often find myself with time on my hands and nothing to do.'
   'I often wake up with a new idea for work and other activities at university.'
3. Finally, 10 of the 28 items are negatively worded and reverse-scored by the investigators. This is a common procedure with Likert-type scales, and is recommended in current introductory textbooks (e.g. Coolican, 2014; Howitt & Cramer, 2014). However, research in other contexts has shown that the reverse-scoring of responses to negative items does not give equivalent scores to those obtained for positively worded items (Hartley, 2013; Hartley & Betts, 2013; Richardson, 2012;

Yorke, 2009) even when the negative items have been coded in reverse.

I am afraid that I think more work is needed on this scale before taking the research further, and thus that both this paper and its sequel (Sharp et al., 2016) are premature.

Paper 2. Kyriacou, Mylonakou-Keke and Stephens (2016)

Kyriacou, Mylonakou-Keke and Stephens (2016) describe the use of a 30-item Likert-type questionnaire designed to assess university students' views of bullying in schools. Here few details are provided about the scale and how it was constructed, and no indication is given of whether or not this information is provided elsewhere.

This scale again used a five-point format – *strongly agree, mildly agree, not sure, mildly disagree* and *strongly disagree*, and it was made available in three languages – English, Greek and Norwegian (with no discussion of the equivalence of these versions). The eventual ratings were collapsed into two levels of agreement – *strongly* and *mildly agree* versus *not sure, mildly* and *strongly disagree*. The discussion focuses on the proportions agreeing and disagreeing in the three countries. Over 80 per cent of the 469 respondents were female, aged 25 years or less.

As in the paper by Sharp et al. described above there are the same two problems with this questionnaire. First, some items contain more than one concern within them, making it difficult to evaluate what is and what is not being agreed or disagreed with. For example:

> 'You can only help a bully to stop bullying by dealing with their home circumstances.'

> 'Schools need to prioritise how to prevent bullying more than how to deal with cases.'

And, again two items are negatively worded (and thus negatively scored). For example:

> 'The best way to deal with bullying is to punish the victim.'

> 'Trying to help a bully stop bullying is a futile waste of time.'

But the main problem with this paper is that there is no account of how the questionnaire was developed and evaluated before its use.

Paper 3. Ebersole et al. (2016)

This paper reports the responses of approximately 3000 students to six questionnaires examining their experience of being in a subject pool. These questionnaires differed in format and in length, and were subject to the same limitations listed above. Some of the details are as in Table 1 below.

Table 1.

Questionnaire	Number of items	Format of scale	Number of points	Direction of scale	Direction of scale score
1	1	Likert	7	-ve – +ve	low – high
2	10	Likert	7	-ve – +ve	low – high
3	2	Likert	7	-ve – +ve	low – high
4	4	Likert	5	-ve – +ve	low – high
5	6	Likert	5	-ve – +ve	low – high
6	15	Likert	4	-ve – +ve	low – high

Here we can see that all of the scales were in a Likert format, with the negative poles on the left, and all were scored low-high from the left. But the number of items in each scale varied from 1–15. Again we might conclude that, despite the large sample size used in this study, the data obtained are limited by the nature of some of the six questionnaires employed.

Conclusions

It is easy to criticise questionnaires – but authors seem to ignore this when discussing their results. It would be a simple but valuable exercise in laboratory classes to examine the quality of the questionnaires used in a set of research papers. Such an enterprise would soon convince students that reported findings are not always what they seem.

What about other approaches?

Finally we might note that that the data given in questionnaires might be more convincing if they were supplemented with comments and opinions from the respondents. Basically I am arguing for a mixed-methods approach. The data on boredom, bullying, and being in a subject pool described above are undoubtedly interesting, but they would be more so if they were amplified by illustrations. For example, the individual accounts of 17 aged persons attempting to use digital technology, and what they use it for (presented by Hill et al., 2015) are mind-blowing – but they are selective and they need substantiating with some quantitative data obtained from a survey or questionnaire. Contrasting their results with those reported by 4000 survey respondents (Jelfs & Richardson, 2013) is an interesting exercise.

Students need to be made aware of the strengths and limitations of much research in psychology. What better way of doing this than letting them discuss the strengths and weaknesses of the methods they take for granted?

Acknowledgement

I am grateful to John Richardson and Nicky Hayes for helpful comments on an earlier draft of this manuscript.

James Hartley
School of Psychology
Keele University
Staffordshire
ST5 5BG
UK
Email: j.hartley@keele.ac.uk

References

Coolican, H. (2014). *Research methods and statistics in psychology* (6th edn). Hove: Psychology Press.

Ebersole, C.R. et al. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology, 67,* 68–82.

Hartley, J. (2013). Some thoughts on Likert-type scales. *International Journal of Clinical and Health Psychology, 13,* 383–386.

Hartley, J. & Betts, L. (2013). Let's be positive: The effects of the position of positive and negative values and labels or responses to Likert-type scales. *Chinese Journal of Psychology, 55,* 291–299.

Hartley, J. & Cabanac, G. (submitted). How do we know how well it works? Using surveys and questionnaires to assess the effects of new technologies.

Hill, R., Betts, L. & Gardner, S. (2015). Empowerment and enablement through digital technology in the generation of the digital age. *Computers in Human Behavior, 48,* 415–423.

Howitt, D. & Cramer, D. (2014). *Introduction to research methods in psychology* (4th edn). Harlow: Pearson Education Ltd.

Jelfs, A. & Richardson, J.T.E. (2013). The use of digital technologies across the adult lifespan in distance education. *British Journal of Educational Technology, 44*(2), 338–351.

Kyriacou, C., Mylonakou-Keke, I. & Stephens, P. (2016). Social pedagogy and bullying in schools: the views of university students in England, Greece and Norway. *British Educational Research Journal, 42*(4), 631–645.

Richardson, J. (2012). The role of response biases in the relationship between students' perceptions of their courses and their approaches to studying in higher education. *British Educational Research Journal, 38*(3), 399–418.

Sharp, J.G., Hemmings, B. & Kay, R. (2016). Towards a model for the assessment of student boredom and boredom proneness in the UK higher education context. *Journal of Further and Higher Education, 40*(5), 649–681.

Sharp, J.G., Hemmings, B., Kay, R., Murphy, B. & Elliott, S. (2016). Academic boredom among students in higher education: A mixed-methods exploration of characteristic, contributors and consequences. *Journal of Further and Higher Education*. doi: 10.1080/0309877X.2016.1159292.

Yorke, M. (2009). Student experience surveys: Some methodological considerations and an empirical investigation. *Assessment and Evaluation in Higher Education, 34,* 721–739.