LIKERT ITEMS AND SCALES

Rob Johns (University of Strathclyde)

1. The ubiquitous Likert item

The question above, taken from the 2007 British Social Attitudes survey, is an example of a **Likert item**. Almost everyone would recognise this type of survey question, even if not many people would know it by that name. This agree-disagree approach to measuring attitudes has for decades been ubiquitous in questionnaires of all kinds: market research, opinion polling, major government surveys and academic studies in fields ranging from political science to product design. Not only is it a pleasingly simple way of gauging specific opinions, but it also lends itself very easily to the construction of multiple-item measures, known as **Likert scales**, which can measure broader attitudes and values. This fact sheet opens with a brief synopsis of the landmark article in which Likert himself first set out this approach to measuring attitudes. Then we look in more detail at the construction of both individual Likert items and multiple-item Likert scales, using examples from the Survey Question Bank to illustrate the decisions facing questionnaire designers looking to use the Likert method.

2. The basis for Likert measurement

Rensis Likert was an American psychologist. (Unlike most of those who have used it since, he pronounced his name with a short 'i' sound, as in 'Lick-ert'.) What became known as the Likert method of attitude measurement was formulated in his doctoral thesis, and an abridged version appeared in a 1932 article in the Archives of Psychology. At the time, many psychologists believed that their work should be confined to the study of observable behaviour, and rejected the notion that unobservable (or 'latent') phenomena like attitudes could be measured. Like his contemporary, Louis Thurstone, Likert disagreed. They argued that attitudes vary along a dimension from negative to positive, just as heights vary along a dimension from short to tall, or wealth varies from poor to rich. For Likert, the key to successful attitude measurement was to convey this underlying dimension to survey respondents, so that they could then choose the response option that best reflects their position on that dimension. This straightforward notion is illustrated below.

Negative   *Neutral*   *Positive*

Disagree   Disagree   Undecided   Agree   Agree
strongly (1)   (2)   (3)   (4)   strongly (5)

As far as Likert was concerned, attitudes towards any object or on any issue varied along the same underlying negative-to-positive dimension. This had three significant implications. First, his method was universally applicable. In Likert's own research, he measured opinions on subjects as diverse as birth control, the Chinese, evolution, war, and the existence of God. Second, provided that the response options covered the negative-to-positive dimension, their precise wording could vary. Hence Likert's 1932 article included items worded as in the example above but also some with response scales running from 'strongly disapprove' to 'strongly approve'. Third, because responses were comparable across different questions – in each case simply reporting how positively or negatively that respondent was disposed to the attitude object in question – they could be assigned the same numerical codes, as illustrated in the diagram above. Furthermore, with multiple items on the same broad object (such as those listed just above), these codes could be summed or averaged to give an indication of each respondent's overall positive or negative orientation towards that object. This is the basis for Likert scales.

These advantages of the Likert format – above all, its simplicity and versatility – explain why this approach is ubiquitous in survey research. Yet there are a variety of

reasons why Likert measurement is not quite as simple as it looks. In the rest of this fact sheet, we examine the reasons why.

3. Designing Likert statements

Any Likert item has two parts: the 'stem' statement (e.g. "Young people today don't have enough respect for traditional British values") and the 'response scale' (that is, the answering options offered to respondents). When it comes to stem statements, most of the relevant guidelines would apply to the design of any survey question. They should be simple (and preferably quite short), clear and as unambiguous as possible. Three rules call for particular attention, however.

First, **double-barrelled** questions – that is, those that contain two attitude objects and are therefore potentially asking about two different attitudes – should be avoided. Although this is a well-known rule, it is often and easily broken, as a couple of examples from the British Social Attitudes survey illustrate:

```
Q1220  [CanCrime]
       CARD M3
       Now thinking about the drug cannabis. How much do you agree or
       disagree that... cannabis is a cause of crime and violence?
```

[ScFailCl]
   a. Schools that fail to attract enough
      pupils should be closed and the
      teachers lose their jobs.

Respondents might reasonably think that cannabis leads to crime (indeed they might think that it follows logically from cannabis use being criminalised) without believing that it leads to violence. Equally, they might believe that unpopular schools should be closed but that teachers, rather than losing their jobs, should instead be transferred to the more popular schools. Double-barrelled questions create problems for respondents, who are forced to choose which part of the statement to address, and for researchers, who have no means of knowing which part the respondents chose.

The second rule is to avoid **quantitative** statements. This is also best illustrated by some examples from the British Social Attitudes survey.

```
Q1224   [CanSelPr]
        CARD M3 AGAIN
        (How much do you agree or disagree that ...) people who sell
        cannabis should always be prosecuted?
1       Agree strongly
2       Agree
3       Neither agree nor disagree
4       Disagree
5       Disagree strongly
8       (Don't know)
9       (Not answered)
```

----

24. Please tick <u>one</u> box to show how much you agree or disagree with each of these statements.

PLEASE TICK **ONE** BOX ON EACH LINE	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	Can't choose
[RelSch3] c. Single religion schools have a better quality of education than other schools	☐	☐	☐	☐	☐	☐

It is the quantitative terms in those questions, 'always' and 'better', that cause the problems by introducing ambiguity into 'disagree' responses. Take someone who chooses 'Disagree strongly' with the first statement. Do they strongly disagree only with the policy of prosecuting *all* dealers, or should we infer that they think cannabis dealers should *never* be prosecuted? Meanwhile, someone disagreeing with the second statement may think that the quality of education is no better in faith schools, or they may think it is actually worse. The key point is that Likert items are intended to capture the extent of agreement or disagreement with an idea, and not to measure some sort of quantity or 'hidden variable'. If the latter is the purpose of an item, then it should be recast with response options designed to make that hidden variable explicit. In the second example above, the variable is the 'relative quality of education in faith schools', and the response scale should therefore run from 'much better' to 'much worse' (via 'no different').

The third rule concerns **leading** questions. Normally, questionnaire designers are urged to be even-handed in their approach, asking questions from a neutral standpoint and avoiding leading respondents towards a particular answer or opinion. An easily overlooked aspect of Likert items is that, by their very nature, they break this rule. The stem statements are clear and potentially persuasive assertions. For example, the above statement about faith schools could be argued to lead respondents towards a positive evaluation of the education that those schools provide. This matters because there is ample evidence that respondents are indeed led in this way. **Acquiescence bias** – a tendency to agree with statements, to some extent irrespective of their content – has long been known to be a serious problem with the Likert format. Its impact is vividly illustrated by a question wording experiment reported by Schuman and Presser (1981, ch. 8).

	Agree (%)	Disagree (%)
Version A: "Individuals are more to blame than social conditions for crime and lawlessness in this country"	60	40
Version B: "Social conditions are more to blame than individuals for crime and lawlessness in this country"	57	43

Survey respondents were randomly allocated to one of two versions of the stem statement. These versions were, as the table shows, direct reversals of one another. Hence, since 60% agreed with version A, we would expect only 40% to have agreed with version B. In fact, though, comfortably over half of respondents agreed on both versions. This suggests not only that Likert statements can indeed persuade respondents of the argument that they present, but also that the scale of such acquiescence bias is considerable. Schuman and Presser therefore advise questionnaire designers to avoid the Likert format where possible. In this case, the obvious alternative is a question asking respondents "Which do you think is more to blame for crime: individuals or social conditions?"

4. Designing the Likert response scale

The example set out at the beginning of this fact sheet uses what is probably the most common formulation of the Likert response scale. As noted earlier, the man himself also used an approve-disapprove format, and it has become quite common for people to use the term Likert to refer to almost any rating scale designed to measure attitudes. Here, though, we will limit our attention to agree-disagree questions. That nonetheless leaves a number of decisions facing question designers.

The first concerns the **number of scale points**. While Likert opted for five, there is no theoretical reason to rule out different lengths of response scale. (After all, as noted above, the options are supposed to reflect an underlying continuum rather than a finite number of possible attitudes.) And, in survey practice, various lengths – from two points up to eleven or even more – have been used. The reason why five has become the norm is probably because it strikes a compromise between the conflicting goals of offering enough choice (since only two or three options means measuring only direction rather than also strength of opinion) and making things manageable for respondents (since few people will have a clear idea of the difference between, say, the eighth and ninth point on an eleven-point agree-disagree scale). Research confirms that data from Likert items (and those with similar rating scales) becomes significantly less accurate when the number of scale points drops below five or above seven. However, these studies provide no grounds for preferring five rather than seven-point scales.

One simple way of illustrating the problems with long scales is that **labelling the response options** becomes extremely difficult. Typically, for seven-point scales, options labelled 'slightly agree/disagree' are introduced either side of the neutral point. Much beyond that, though, the shades of agreement become as hard for survey designers to express as they are for respondents to distinguish. One way round this, common with longer scales though not unheard of even with the five-point version, is to confine verbal labels to the extreme points on the response scale. The points between 'Strongly agree' and 'Strongly disagree' are simply given numerical labels. The problem with this simple strategy is that the evidence from studies of survey response is unequivocal: full labelling enables respondents to deliver much higher quality data. (They also prefer it.)

So far, we have referred almost exclusively to response scales with an odd number of points. That is because the standard practice, again following Likert's original example, is to include a **neutral midpoint**. While Likert labelled this point as 'Undecided', the more common version is now 'Neither agree nor disagree'. The purpose of this option is evidently to avoid forcing respondents into expressing agreement or disagreement when they may lack such a clear opinion. Not only might this annoy respondents, but it also risks data quality. It has long been recognised that people often lack clear views on even the hottest political topics; at the same time, they are usually reluctant to use a genuine **non-response option** like 'Don't know'. In that context, the midpoint is a useful means of deterring what might otherwise be a more or less random choice between agreement and disagreement. This helps to explain why the labels typically given to Likert midpoints are compatible both with ambivalence (i.e. definite but mixed feelings) and indifference (i.e. no particular feelings about the statement).

That said, there may sometimes be a case for forcing respondents to come down on one side or the other. The reason is that some people use the midpoint to avoid reporting what they see as less socially acceptable answers. Another survey experiment, taken from Johns (2005) illustrates the point.

		Agree (%)	Disagree (%)
"It is the government's responsibility to provide a job for everyone who wants one"	Version A: midpoint offered	57	43
	Version B: midpoint omitted	48	52

In this case, the two versions of the question varied according to whether a midpoint was included. When that option was offered, one quarter of respondents chose it. Of the rest, a majority (57%) reported agreeing with the statement. In contrast, among respondents were denied a midpoint, disagreement was (narrowly) the majority position (52%). The most plausible explanation for the difference is that, when a midpoint was offered, it attracted many of those who actually disagreed but were reluctant to admit as much, perhaps because they felt disagreement could appear callous, implying a lack of sympathy for the unemployed. When questions are on controversial topics, such as immigration or racial matters, the most accurate gauge of public opinion might therefore be obtained by omitting the neutral category (and perhaps offering a 'Don't know' option for those who really cannot choose between agreement or disagreement).

5. Constructing Likert scales

A Likert scale is a composite, or 'battery', of multiple Likert items. (The terminology can be confusing because the list of answer categories is, as in this fact sheet, usually referred to as the 'response scale'. But the precise term 'Likert scale' should always refer to a collection of items.) The example presented at the outset, about young people respecting traditional British values, is part of a six-item Likert scale that has long been included in British Social Attitudes surveys to measure libertarian-authoritarian values.

45. Please tick one box for each statement below to show how much you agree or disagree with it.

*PLEASE TICK **ONE** BOX ON EACH LINE*

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly
[TradVals] a. Young people today don't have enough respect for traditional British values.	☐	☐	☐	☐	☐
[StifSent] b. People who break the law should be given stiffer sentences.	☐	☐	☐	☐	☐
[DeathApp] c. For some crimes, the death penalty is the most appropriate sentence.	☐	☐	☐	☐	☐
[Obey] d. Schools should teach children to obey authority.	☐	☐	☐	☐	☐
[WrongLaw] e. The law should always be obeyed, even if a particular law is wrong.	☐	☐	☐	☐	☐
[Censor] f. Censorship of films and magazines is necessary to uphold moral standards.	☐	☐	☐	☐	☐
	(1)	(2)	(3)	(4)	(5)

Likert scales are 'summated' scales, so called because a respondent's answers on each question are summed to give their overall score on the attitude or value. In the example above, 'Agree strongly' is coded as 1 and 'Disagree strongly' is coded as 5. It could just as well be the other way round; the key point is to work out how your chosen way relates to the attitude to be measured. Here, since the items are worded so that agreement is the authoritarian position, high scale scores actually denote libertarian opinions while it is lower-scoring respondents who are the authoritarians. Once the scores are summed, they can be 'rescaled' into a more intuitive range. For example, scores on this scale would range between 6 (all 1s) and 30 (all 5s). This could easily be recoded to a 1 (maximally authoritarian) to 25 (maximally libertarian) scale. Another, probably preferable, alternative is to calculate the mean rather than the sum of a respondent's scores. These averages will fall on the same 1-5 scale as the individual items. (A further advantage of the averaging approach is that, should respondents have skipped or answered 'Don't know' on one or two of the questions, a mean can still be calculated based on those questions that were answered. Dealing with missing data is less straightforward with the summing approach.)

One objection that could be raised about the procedure above is that it involves some very dubious assumptions about the possibility of translating attitudes – that is, complex and slippery psychological phenomena – into numbers. The numerical distances between 1 and 2, and between 2 and 3, are equal. Can anyone really say the same about the distances between 'Agree strongly' and 'Agree', and between 'Agree' and 'Neither agree nor disagree'? There is also something odd about declaring, say, that someone has a 'mean attitude' of 1.94 (let alone that a different respondent with a mean score of 3.88 was therefore 'twice as libertarian'). In statistical terms, these objections amount to arguing that the **level of measurement** of the Likert response scale is 'ordinal' rather than 'interval': that is, we can make assumptions about the order but not the spacing of the response options. One response to this is to refer back to Likert's original argument, which was that survey respondents actually construe the response scale in terms of evenly-spaced points along an underlying attitude continuum. Fortunately, we can sidestep a complicated argument about quite how people interpret and envisage response scales. The advantages of Likert scales are available even if scores on the scale are seen as placing respondents in order of authoritarianism, rather than generating precise and comparable scores on that attitude. Multiple-item measurements give more accurate readings, whether rankings or ratings, than could be obtained from any individual item.

One reason for this greater accuracy is that Likert scales allow us to cover the various facets of what are often complex and multidimensional attitudes or values. In this instance, the items take in what social psychologists have identified as the three core elements of authoritarianism: punitiveness, submissiveness to authorities, and moral traditionalism. A related advantage of Likert scales is that they dilutes the impact of the 'random' error to which any individual item is inevitably subject. For example, a respondent might have broadly libertarian values but, having recently read about or even been the victim of a serious crime, be temporarily converted to the need for stiffer sentences. If that were the only question asked, this respondent would be wrongly categorised as authoritarian.

It follows from this that, when it comes to Likert scales, the longer the better. One pioneering measure of authoritarianism, the California F-scale, contained 30 items. Unfortunately, the luxury of that much questionnaire time is very seldom available to survey designers. This means that the process of **item selection** becomes crucial. If only a small number of items can be included, it is doubly important that they are carefully chosen so that they all measure the target attitude, and between them cover all its relevant dimensions. Whole books have therefore been devoted to the subject of forming and evaluating Likert scales (e.g. DeVellis, 2003), and the six-item libertarian-authoritarianism scale set out above is itself the result of a lengthy scale development procedure. Here, we focus on two key attributes of a good Likert scale: internal consistency and balance.

9

The **internal consistency** of a scale refers to the extent to which its items can be said to be measuring the same thing. This can be judged simply by looking at the items and comparing their content with previous theoretical accounts of the attitude in question, as we did above when discussing the British Social Attitudes authoritarianism scale. More commonly, however, internal consistency is measured statistically. The logic is that two items can be deemed to be measuring the same attitude if those who agree with one item also tend to agree with the other – that is, if responses to those items are closely correlated. And the standard measure of internal consistency in a Likert scale is the 'alpha coefficient', which is based on the correlations between each pair of items in the scale. The closer alpha is to its maximum value of 1, the more internally consistent the scale. Any attempt to specify a required level for alpha is inevitably rather arbitrary, but the coefficient calculated for that British Social Attitudes libertarian-authoritarian scale in 2007, 0.72, comfortably exceeds 0.6 which is probably the most commonly suggested benchmark.

One less satisfactory feature of that example scale is that it is **unbalanced**. In this context, that refers to the fact that the items are all worded in the same – in this case, the authoritarian – direction. In a balanced scale, at least some (and preferably half) of the items would be worded so that agreement was the libertarian position. For example, in previous British Social Attitudes surveys, items like "The death penalty is never an appropriate sentence" and "Censorship has no place in a free society" have replaced the items on those topics in the unbalanced version set out above. There are various reasons why balanced scales are superior. First, respondents with a particular viewpoint are liable to resent a succession of statements, all asserting the contrary position. Moreover, if Likert statements can sway respondents towards agreement, then an unbalanced scale will distort the results. That is not only suspect from an ethical point of view, but it defeats the object of the scale which is to obtain an accurate measure of authoritarianism.

This brings us back to the notion of acquiescence. All Likert items are subject to this tendency to agree with declarative statements. The problem with unbalanced scales is that they make it impossible to distinguish acquiescence from the attitude that is supposed to be measured. Suppose that a respondent agrees with all six of the British Social Attitudes authoritarianism items. We do not know whether that respondent is an authoritarian or simply prone to agree. If the scale was balanced, however, and a respondent agreed throughout, then we could much more confidently diagnose acquiescence. There are much more statistically sophisticated means of measuring acquiescence bias than that, of course. But all of them require balanced scales.

6. Conclusion

Handbooks on questionnaire design usually implore researchers to keep things as straightforward as possible. On the face of it, little is simpler than a Likert item. The format is also very versatile: it is difficult to find an attitude item of any type that could not be converted into a Likert item, and usually without gross clumsiness. Hence, the format can be used repeatedly in the same questionnaire. That is especially important given that multiple items are often required to do justice to the multi-faceted attitudes to be measured. Consistent use of the Likert response scale is likely to prove much more popular with respondents than consistent switching of more complex formats.

Yet there are strong arguments for avoiding the Likert format. Put simply, questions are better than statements. That is partly because disagreement with Likert statements is often ambiguous, especially when they contain some sort of quantitative element. It is also because such statements are inherently persuasive, which in turn means that agreement with Likert items is also often ambiguous: is the respondent agreeing with the idea, or because he tends to agree anyway? In terms of the accuracy of responses, a preferable strategy is to find the 'hidden variable' and to ask an even-handed question about that.

As quite often in survey design, then, there is a trade-off between convenience – for both researchers and respondents – and data quality. The dictates of the former mean that the Likert method will probably remain the workhorse of public opinion research. In that case, it is all the more important to follow the simple rules –both for individual items and for composite scales – set out here.

7. References and further reading

Carifio, James and Perla, Rocco J. (2007): Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes, *Journal of Social Sciences*, 3(3), 106-116 **(on some of the key controversies about Likert measurement)**.

DeVellis, Robert F. (2003): *Scale Development: Theory and Applications*, Thousand Oaks, CA: Sage **(on building Likert scales)**.

Likert, Rensis (1932): A technique for the measurement of attitudes, *Archives of Psychology*, 140(1), 44-53 **(the original article)**.

Johns, Robert (2005): One size doesn't fit all: Selecting response scales for attitude items, *Journal of Elections, Public Opinion and Parties*, 15(2), 237-64 **(on neutral options)**.

Schuman, Howard and Presser, Stanley (1981): *Questions and Answers in Attitude Surveys*, New York: Academic Press, ch. 8 **(on acquiescence)**.