Survey Responses Using Scale Categories Follow a "Double Jeopardy" Pattern

John Dawes, University of South Australia

Abstract

This paper examines the use of scale category responses (i.e. a response of "6 out of 10" on a scale of zero to ten) by survey respondents. It adopts the terminology of brand buying research and discusses *penetration* (number of respondents who use a scale category response at least once in a survey) and *frequency of usage* (how many times those respondents who do use a particular scale category response do actually use it. The study finds a strong 'double jeopardy' effect. That is, responses that are used by many people tend to be used more often by the people who use them. Responses that are used by fewer people tend to be used less often by the people who use them. This is explained by the fact that respondents tend to use a range of scale category responses, typically about half the number of available scale points, and mostly in the upper half of the scale. Occasionally they use a scale response category outside of their normal range. The scale response categories used in such instances differ from person to person. Therefore any particular respondent uses some scale response categories more than others. Scale response categories that appear in more respondent's set of responses, by default are also used more often by the respondents that use them. This finding provides some norms for what survey data obtained using Likert type or numerical scales "should look like".

Introduction and Background

The issue of how people respond to rating scale questions is of interest in a variety of disciplines including marketing, sociology, education and psychology. This issue is usually discussed in the literature under the general topic of response styles. Research issues in response styles include the incidence and effect of extreme responses - 'yea saying' or 'nay saying,' the tendency for some respondents to use the middle of the scale, and aquiescence bias.

Much effort has been expended to examine how extreme response styles can be detected and compensated for (Greenleaf, 1992b), the possible associations between background variables such as age, education, culture, and extreme response style (Clarke III, 2000), and the correlates or effects of acquiescence bias (Javeline, 1999). Other work has examined the effect of response styles on multivariate analysis techniques (Heide and Grønhaug, 1992).

This paper addresses a much more basic issue, which is simply to investigate and describe possible generalisable patterns in responses to survey questions. In particular, this paper examines responses to questions that ask a respondents to give a rating using a scale, or indicate their extent of agreement, such as "7 out of 10" or "agree" – the latter which is often then coded as an interval scale response.

I begin by making an observation and asking a seemingly innocuous question. In the following table are response category frequencies from a commercially gathered market research data set. 187 respondents were asked 15 questions pertaining to service quality. The data is therefore r (for respondents) x c columns (for questions) = rc or the total number of responses to all questions, tabulated according to the response category used.

Table 1: Responses to 15 questions

Response category	N responses	% of all responses
0 (0 percent agreement)	91	3
1	31	1
2	23	1
3	53	2
4	53	2
5	309	11
6	138	5
7	409	15
8	590	21
9	396	14
10 (100 percent agreement)	662	24
total	2775	100

The most popular response is "10" which was used for 662 of the total of 2775 responses. Why is this so? Obviously there was an overall high level of agreement to this battery of questions. But is the popularity of "10" mainly due to many or most respondents using it (for which we adopt the terminology of brand usage research (Ehrenberg, 2000) and call a *penetration* effect), or was there perhaps a smaller group of respondents who used it very, very often, (which we could call a *frequency of usage* effect). It is possible, for instance, that the 662 responses of "10" are comprised of 662 / 15 questions = 44 respondents who answered "10" for every question, or at least a significant group who used "10" for a large part of the whole survey. This would mean that the popularity of "10" was primarily a frequency of usage effect. Alternatively, it is possible that most respondents use "10" at some stage in their series of responses. The converse situation can be considered for response categories such as "1" that have a low incidence in the data set. Plainly this must be primarily a lack of penetration - few respondents could use that response category with it being used for only 1% of the total responses! However, is it the case that low incidence response categories are used by *very* few people but those who use them use them very often (i.e. 'contrarian respondents'). Alternatively, they may be simply used by few people and not used very often even by the respondents who used them.

If there is some generalisable answer to these issues in terms of a pattern in the incidence of certain response categories, this could be useful for those who regularly conduct such research. It may be possible to describe "normal conditions" for such survey data. Therefore, further analysis was conducted on a series of data sets.

The aim was to firstly determine *how many* scale categories respondents typically use (relative to the total available). Second, to examine the amount of variation in the range of scale *values* around their mean response over a series of questions. We know from published evidence that some respondents tend to have extreme response styles (high, low, or middle). However, there is less descriptive knowledge in terms of the general verbal behaviour of respondents to market research surveys using numerical or Likert type scales. Following on from that, the third objective was to examine any relationships between the incidence of a certain response category in a given data set, the number of respondents who used it (penetration) and the number of times those users did use it (frequency of usage).

Data

Thirteen data sets of varying topics, questionnaire length, and number of scale categories were obtained. All were from commercial market research studies conducted from 1996 to 2001. Details of the data sets are shown below, along with the average number of scale responses used by respondents in the right column.

Data set	Sample size	Respondents	Method	No. of questions	# scale categories	Avg. # scale points used
1	331	consumers	Face-to-face	39	11	5.5
2	225	business	Telephone	14	11	4.0
3	1802	cons & bus	Telephone	25	11	5.1
4	496	business	Telephone	20	11	5.0
5	1966	business	Telephone	17	11	5.6
6	580	business	Telephone	20	11	6.4
7	880	consumers	Telephone	3	5	1.8
8	1024	consumers	Telephone	6	5	3.0
9	1000	consumers	Telephone	6	5	3.0
10	77	business	mail	3	5	1.7
11	152	consumers	Telephone	7	5	2.5
12	119	business	Telephone	47	11	5.5
13	697	consumers	Telephone	50	11	7.8

On *average*, respondents use half the available scale range for 5 point or 11 point scales. This is weakly (r=0.4) correlated to the number of questions asked. The distribution of scale point usage is also approximately normally distributed. For example, in data set 1, 70% of respondents used 6 different scale response categories, while 7% used only 2 and only 5% used 9 different scale response categories.

These findings suggest that the difference in response categories in terms of the proportion of total responses they account for in a survey, is more likely due to either a penetration effect, or a combined penetration and frequency of usage effect, rather than primarily being a usage effect. Since we find that the distribution of scale point usage is approximately normally distributed, then not many respondents restrict their responses to a very small proportion of the total scale range. In other words, if a zero to ten scale is used, with a questionnaire of reasonable length, not many people only use one or two scale response categories throughout the survey. Therefore the fact that certain scale response categories account for a much larger proportion of total responses is unlikely to be due to a minority of respondents using those categories for almost all of their responses. The question can now be narrowed down to whether the difference in usage between scale response categories is a penetration effect or a combined penetration and frequency of usage effect. To answer this question we plotted the proportion of people who used a certain response category against the number of times they used it. If there is a positive association between the two, this suggests a combined penetration and frequency of usage effect. If there is no association, this suggests a purely penetration effect. For details of the plots, see the last section of the paper.

Findings

There is a consistent double jeopardy-like (McPhee, 1963 , Ehrenberg et al., 1990) relationship between penetration and usage of scale response categories in each data set. In other words, responses that are used by many respondents also tend to be used more often by the people that use them. Responses that are used by only few respondents tend to be used less often by the people that use them.

Explanation

Why is this so? It appears that in any battery of questions, of the type typically used in service quality or customer satisfaction questions, that (a) most respondents use a subset of the scale, usually toward the upper end of the scale than the lower end (Peterson and Wilson, 1992), and (b) respondents will occasionally give a response outside their normal range of responses, but this response is likely to be different for different people. In other words, for many questions, respondents will typically give an answer of say, 7, 8 or 9 (out of 10). For some questions, *some* respondent might give a different response – suppose they disagree with a statement so they say "3". But another respondent might respond with a "4". So there will be many respondents using response categories 7,8,9 (hypothetically) and using them for most questions, while a few respondents give a different level of response (some might use 3, some might use 4, again hypothetically) occasionally through the survey. Note that this finding is *not* simply a reflection of using questions that vary in their wording extremity or likely aggregate levels of response. This pattern may occur with questions that have the *same* mean average level (i.e., several questions or statements that result in average scores of "6.5 out of 10".

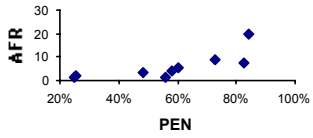
Usefulness

The usefulness of this finding is that it tells research practitioners what to *expect* when constructing surveys using agree .. disagree or numerical scales. It also provides an unusual insight into aggregate level (verbal) consumer behaviour. In aggregate, respondent's use of response categories resembles the patterns shown in repeat purchase of brands (the double jeopardy effect in terms of brand penetration and repeat buying). This is an amazing coincidence, given that double jeopardy is explained as being due to 'asymmetry in people's familiarity with or exposure to items that are similar but differ in popularity' (Ehrenberg et al., 1990 p.85). In comparison, respondents should not differ in familiarity across the range of numbers from zero to ten, etc. - yet the pattern is the same. This finding may help to create norms in terms of what data comprising *x* questions of *y* scale type *should look like*. We see that the patterns in the scatterplots on the series of graphs over the page differ (the relationship is flatter / steeper in some cases). There are also some 'outliers' in the pattern, which is that some response categories are used 'too many times' given their penetration. In several cases, the response category that is used 'too many times' is the response 5 (out of 10). It may be that this is used by respondents as a surrogate for *don't know*. The finding is useful in terms of diagnosing how well the questionnaire battery or scale type is working, and departures from 'normality' can be investigated. This research complements prior work on response styles (e.g. Greenleaf, 1992a). Understanding individual level response *styles* can help in the improvement of scale measures. Understanding *aggregate* level response patterns, as in this research, can help scale development and analysis by identifying which scale response categories appear 'normal' in terms of how often they are used relative to how many respondents use them. Further research is planned to identify how this general pattern might

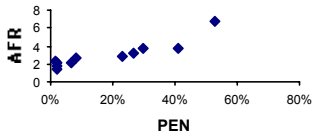
differ across surveys. For example we may find that when scale anchors are used, this distorts the general 'double jeopardy' pattern. Also, the overall level of agreement across the battery of questions may influence the arithmetic relationship between the penetration of a scale response category and the number of times it is used.

Graphs

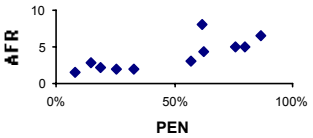
This section show the graphs and correlations of penetration (no. of respondents using that response category) and average frequency of response for those who did use that response category. The X axis is penetration (number of respondents using that response category) and the Y axis is the average number of times that response category is used. Note that Data set 13 is not shown due to space limitations.



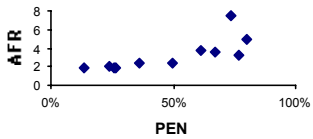
Data set 1 r = .74



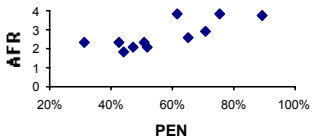
Data set 2 r = .92



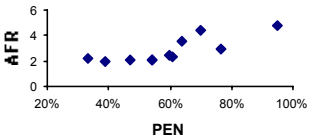
Data set 3 r = .79



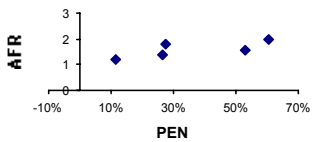
Data set 4 r = .77



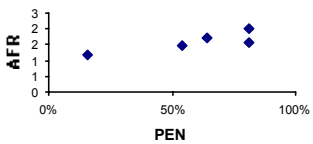
Data set 5 r = .79



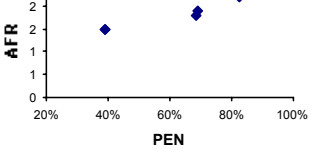
Data set 6 r = .81



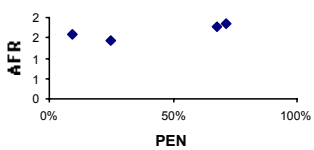
Data set 7 r = 73



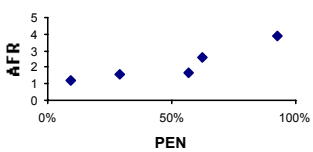
Data set 8 r = .85



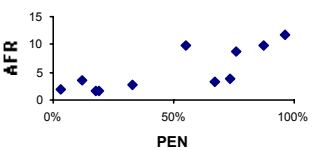
Data set 9 r = .97



Data set 10 r = .86



Data set 11 r = .92



Data set 12 r = .79

References

Clarke III, I., 2000. Extreme Response Style in Cross-Cultural Research: An Empirical Investigation. Journal of Social Behaviour & Personality 15 (1), 137-152.

Ehrenberg, A.S.C., 2000. Repeat Buying - Facts, Theory and Applications. Journal of Empirical Generalisations in Marketing Science 5, 392-770.

Ehrenberg, A.S.C., Goodhardt, G.J. and Barwise, T.P., 1990. Double Jeopardy Revisited. Journal of Marketing 54 (July), 82-91.

Greenleaf, E.A., 1992a. Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. Journal of Marketing Research 29 (May), 176-188.

Greenleaf, E.A., 1992b. Measuring Extreme Response Style. Public Opinion Quarterly 56, 328-351.

Heide, M. and Grønhaug, K., 1992. The Impact of Response Styles in Surveys: A Simulation Study. Journal of the Market Research Society 34 (number 3), 215-230.

Javeline, D., 1999. Response Effects in Polite Cultures. Public Opinion Quarterly 63, 1-28.

McPhee, W.N., 1963. Formal Theories of Mass Behaviour, New York: The Free Press of Glencoe.

Peterson, R.A. and Wilson, W.R., 1992. Measuring Customer Satisfaction: Fact and Artifact. Journal of the Academy of Marketing Science 20 (No. 1), 61-71.