Do data characteristics change according to the number of scale points used ? An experiment using 5 point, 7 point and 10 point scales.

John Dawes
Ehrenberg-Bass Institute for Marketing Science
University of South Australia

Brief Biography

John Dawes is an Associate Professor at the Ehrenberg-Bass Institute for Marketing Science, University of South Australia. His research interests centre on competitive market structure and the effects of price changes on buyer behaviour. In addition he has conducted a number of studies on scaling issues.

Keywords: scales, measurement scales, Likert scale, scale points, re-scaling, psychometrics.

Published in: International Journal of Market Research, Vol 50, No. 1, 2008.

The data used for this study is publicly available from www.johndawes.info

Do data characteristics change according to the number of scale points used ? An experiment using 5 point, 7 point and 10 point scales.

Abstract

To what extent does the number of response categories in a Likert-type scale influence the resultant data ? Surprisingly little attention has been paid to the issue of whether the response category format has any influence on data characteristics such as the mean, coefficient of variation, skewness and kurtosis. This issue is important for several reasons. The first is that decisions are made based on outcomes such as the mean score. For example, marketing organizations and research providers use Likert type scales to measure constructs such as customer satisfaction. In this situation a higher score is better. Could the score have been comparatively better if a different scale format had been used ? There is an absence of evidence on this issue. The second reason is that scale formats that are used in on-going market research projects such as tracking studies occasionally change. Can the old results be re-scaled or transformed to be comparable to data from a new scale format ? Again, little is known about this. The third reason concerns data characteristics such as variation about the mean, skewness and kurtosis. Analysis tools such as regression are often used on data of this type to explain the variation in certain variables. If there is little variance in the data, this is harder to do. How does scale format affect these characteristics ? The answers would be useful to both market researchers as well as academics. A literature review found that little work has been done on this issue. Therefore, this study set out to investigate the impact of scale format on data characteristics. It examined how using Likert-type scales with varying numbers of response categories affects the resultant data in terms of mean scores, and measures of dispersion and shape. Three groups of respondents were administered a series of eight questions (group n's = 300, 250,185). Respondents were randomly selected members of the general public. A different scale format was administered to each group – either a five-point, seven-point or ten-point scale. The surveys were conducted by a professional market research organisation via telephone interview. Data characteristics of mean score, standard deviation, skewness and kurtosis were analysed according to scale format. The five and seven-point scales were re-scaled to a comparable mean score out of ten. The study found that the five and seven-point scales produced the same mean score as each other, once they were re-scaled. However the ten-point format tended to produce slightly lower relative means than either the 5 or 7-point scales (after the latter were re-scaled). The overall mean score of the eight questions was 0.3 scale points lower for the 10-point format compared to the 5 and 7-point format. This difference was statistically significant at $p=0.04$. In terms of the other data characteristics, there was very little difference among the scale formats in terms of variation about the mean, skewness or kurtosis. Therefore each of the three formats appears comparable for the type of research project in which multiple-item scales are analyzed with multivariate statistical methods. This study is also 'good news' for research departments or agencies who ponder whether changing scale format will destroy the comparability of historical data. Five and seven-point scales can easily be re-scaled with the resultant data being quite comparable. In the case of comparing five or seven-point data to 10-point data, a straightforward re-scaling and arithmetic adjustment easily facilitates the comparison. Finally, it appears that indicators of customer sentiment – such as satisfaction surveys – may be partially dependent on the choice of scale format. A five or seven-point scale is likely to produce slightly higher mean scores relative to the highest possible attainable score, compared to that produced from a ten-point scale.

Introduction

Rating scales are one of the most widely used tools in marketing research and commercial market research. Rating scales are used to capture information on a range of phenomena. In consumer research, respondents may be asked about their attitudes, perceptions, or evaluations of products, brands, or messages - among many other possibilities. In other marketing research streams, respondents such as managers or marketing personnel may be asked to rate their company's performance, type of strategic focus, personnel, degree of marketing excellence, training regimes and so forth using such scales.

Rating scales typically require the respondent to select their answer from a range of verbal statements or numbers. Scales that use verbal statements include semantic differential scales and Likert scales. An example of the semantic differential scale is *very good … very bad,* or *pleasant … unpleasant*. An example of the Likert response scale is as follows: *strongly disagree, disagree, neither disagree or agree, agree, strongly agree*. This particular example is a five-point Likert scale utilising verbal response descriptors. Likert scales may also use numerical descriptors where the respondent selects an appropriate number to denote their level of agreement. For example, a question could be worded like this: "indicate your agreement from 1 to 5 where 1 equals strongly disagree and 5 equals strongly agree".

The range of possible responses for a scale can vary. Textbooks on the subject typically portray five or seven-point formats as the most common (e.g. Malhotra and Peterson 2006 ch. 10). Ten or eleven-point scales are also frequently used (Loken, Pirie et al. 1987). Hereafter in this study the term 'scale format' is used to refer to scales with differing numbers of response categories.

In terms of the interface between the respondent and the interviewer in a telephone survey, there are some advantages and disadvantages of each scale format. With a five-point scale, it is quite simple for the interviewer to read out the complete list of scale descriptors (1 equals strongly disagree, 2 equals disagree ….). This clarification is lengthier for the seven-point format. Such a verbal clarification becomes quite impractical for a 10-point format as the gradations of agreement become too fine to easily express in words. In this case, the interviewer normally reads out the verbal meaning of the end points. The 10-point format therefore places greater reliance on the respondent using a numerical response for which the precise meaning has not been precisely defined. However, this disadvantage is balanced against the fact that many people are familiar with the notion of rating 'out of ten'.

There have been numerous studies on the topic of how scale format affects scale reliability and validity. Far less attention has been paid to how it influences data characteristics such as mean and variance. The issues of reliability and validity are outside the scope of this study. Suffice to say, simulation studies and empirical studies have generally concurred that reliability and validity are improved by using five to seven-point scales rather than coarser ones (those with fewer scale points). But more finely graded scales do not improve reliability and validity further.

The next section presents some theoretical reasons for why the scale format might influence the mean score, variance and skewness. The small number of empirical studies that have examined this issue are then reviewed.

Why *would* scale format influence data characteristics ?

One of the most basic summary data characteristics is the mean. Scores for Likert-type questions are often 'negatively skewed' (e.g. Dawes 2002b; Peterson and Wilson 1992). This term is counterintuitive and refers to the fact that more responses are at the positive end of the scale and the 'tail' is at the negative end. If more respondents tend to give positive responses, then a finer scale, with more response options, could result in a slightly lower mean score. This can be illustrated by considering the range of positive response options for five, seven and ten-point formats. Firstly consider a five-point scale. There are only two options for a positive response: points four and five. If we average those two responses and re-scale to the equivalent score on a ten-point scale (using the method described and used later under 're-scaling') the result is 8.9 /10. If we undertake the same procedure for a seven-point scale the positive responses are 5, 6, and 7 for an average of 6, which re-scales to a score of 8.3/10. The positive responses for a ten-point scale are 6, 7, 8, 9, and 10 which averages to 8/10. Therefore, based on the arithmetic properties of the scales, the three scale formats would produce somewhat different comparative mean scores if the majority of responses were on the positive side of the mid-point. The potential of the different formats to produce comparatively different mean scores seems worthwhile to investigate.

In relation to the distribution of data about the mean, more scale points, by definition, provide more options for the respondent. Therefore, finer scales could result in a greater spread of the data. This would result in a larger standard deviation, and possibly more positive kurtosis as kurtosis is related to, although not the same thing as, variance.

More scale response options may also conceivably result in less skewed data. This is illustrated using the situation whereby a scale is used to measure a construct that most respondents give a particularly positive response for. A coarse scale will provide few options for this positive sentiment and so the responses may be 'bunched up' at the positive end of the scale. A finer scale could reduce this negative skew by allowing for more gradations of positive response. This could also reduce the overall mean score, for the reasons outlined above.

The empirical studies examining scale format and its association with data characteristics are now reviewed.

Studies examining level and shape of data

There are only a small number of studies on this issue. One is by Finn (1972) which reported means and variances for 3, 5, 7 and 9-point scales. They were 1.6, 2.2, 4.1 and 4.9 for means and .32, .60, 1.32 and 4.0 for variances respectively. Applying a re-scaling formula from Preston and Colman (to be discussed in more detail later in the analysis section), I transformed these reported means to a score out of 100. The transformed scores are 30, 30, 52 and 49 respectively. This suggests the 7 and 9-point formats produced comparatively higher scores. This is counter to the theoretical expectation outlined above. In terms of the variance, taking its square root and dividing this by the original mean score gives the coefficient of variation. This is a standardised measure of variance that controls for the differing number of scale points. The coefficient of variation for Finn's four scale formats is calculated to be .35, .35, .28 and .41 for the 3, 5, 7 and 9-point scales respectively. It appears the nine-point format produced higher comparative variance in that study compared to the coarser scales.

Two other studies are pertinent to the issue of how the number of scale points affects data characteristics such as the mean score. One of these was many years ago, in which Ghiselli (1939) conducted an experiment using undergraduate students who were asked to indicate whether they thought the advertising for 41 different brands was *sincere*. One group answered using a two-point (yes / no) scale, the other group answered using a four-point scale (very sincere … very insincere). The four-point scale resulted in higher ratings for the perceived sincerity of the advertising than the two-point scale.

Another study was by Dawes (2002a) who analysed two split-sample experiments in which groups of respondents were administered questions with either 5-point or 11-point scales. He found that once the 5-point scale was re-scaled to 11-point equivalence, that the means from the 11-point scale were slightly higher by an average of 0.25 points, although no inferential test was applied. This result could be partially attributable to the eleven-point scale having an anchor value of zero (i.e., a zero to ten scale). This characteristic may have artificially lowered the mean score for the 5-point data from the rescaling process. For example, a score of one out of five was rescaled to zero out of ten. In that study, the 11-point scale also produced slightly more dispersion in the data as measured by the coefficient of variation, but there was no difference in skewness or kurtosis between the two scales. Only one other study has examined the issue of scale format and skewness, which was by Johnson, Smith and Tucker (1982). They found that a 2-point format produced more skewness than a five-point format.

This review makes it apparent that basic issues to do with the mean and distribution of the data, and how they are affected by scale format have not been closely studied. There also seems to be some variation in the results from previous studies. While Dawes (2002a) found that re-scaled means from five to eleven-point scales were almost the same; an inspection of the data reported in Finn (1972) showed more marked differences. Likewise, one prior study found that coarse scales resulted in more skewness (Johnson, Smith and Tucker 1982), albeit between two and five-point scales, the former of which is rarely used in marketing studies. Another study found no appreciable difference between five and eleven-point scales in this regard (Dawes 2002a).

Research questions and rationale

We know that scales are ubiquitous in both market research and academic marketing research. But there is a less than comprehensive amount of documented knowledge on the topic. Therefore further investigation of the way scale format might influence the data is warranted. There are at least three reasons for this.

First, the sophistication of analytical methods is increasing. Techniques such as confirmatory factor analysis and structural equation modeling are now commonplace in marketing research. These tools are sensitive to the characteristics of the data, such as variance, kurtosis and skewness (e.g. Bentler 1995). Therefore more knowledge about how scale format affects these characteristics would be desirable.

Second, in many cases the data from a survey is not just reported, it is analysed with the objective of 'explaining' or accounting for the variance in a dependent variable. Examples of the dependent variable might be overall customer satisfaction, probability of purchase, or attitudes towards a brand or organization. The analyst wishes to find out what other variables might be strongly related to higher or lower scores on the dependent variable. In this situation, more variance in the dependent variable is desirable. This is illustrated with an

example. If, hypothetically, all respondents gave the same score for customer satisfaction there would be no variance to explain. If there was very little variance, for example all responses were either 6 or 7 on a seven-point scale then normal OLS regression is not an appropriate analysis method. More complex techniques such as logistic regression would be needed.

The third reason is that in industry, many organizations periodically track consumer sentiment, and often, scales of the type discussed here are a major part of the research. For example, many service organizations such as banks, telecommunication companies or insurance companies routinely survey customers about their perceived levels of service quality or customer satisfaction. For a variety of reasons, the choice of scale is sometimes changed, say, from a five-point scale to a seven-point scale. The reasons for this could be personnel changes, the appointment of a different research provider, department mergers and so on. Obviously the information gleaned from the data, such as mean scores, is based on the number of scale responses used. But is the data dependent on the scale to the extent that the mean score relative to the highest possible score is different for one scale compared to another ? There are some theoretical grounds for thinking scale format might affect the data, as outlined earlier. Also, there is little guidance on this apparently practical and important issue and indeed some conflict in prior results. More knowledge in this area would therefore seem desirable.

This study therefore sought to compare the aggregate-level data characteristics derived from attitudinal questions with either 5, 7 and 11-point numerical scales as the response categories. The specific research question is:

> If data on the same construct is gathered using three scale formats (5-point, 7-point and 10-point numerical scales) and the data from the 5 and 7-point formats are re-scaled to a common 10-point format, are there any differences in terms of mean, variance, kurtosis and skewness ?

This question presumes to treat the data as if they were at least interval quality. There is some evidence that the psychological 'distances' between Likert-type scale points are not equal, for example Bendixen and Sandler (1994) and Kennedy, Riquier and Sharp (1996). That said, the relation between the original scale values and the 'real' identified scale values is very close in these studies. For example in Kennedy, Riquier and Sharp (1996) the notional scale values of 1, 2, 3, 4 and 5 equated to 1, 2.2, 3.1, 4.1 and 5 respectively. The leading texts in the field support the treatment of such scales as if they are equal-interval (e.g. Aaker, Kumar and Day 2004 p. 285; Burns and Bush 2000 p. 314 ; Dillon, Madden and Firtle 1993 p. 276; Hair, Bush and Ortinau 2006 p. 365-366). Based on the empirical studies showing a reasonably close approximation to equal-interval, and the apparent precedent shown in the leading texts, I analysed the data as if it were equal-interval.

Data

In accordance with the research objective, data were gathered via a survey of consumers drawn at random from the electronic telephone directory. The survey was conducted over the 2005-2006 period by a professional market research organisation using CATI (Computer Assisted Telephone Interviewing).

The questionnaire items were derived from existing 'price consciousness' scales (Bruner and Hensel 1992). Price consciousness is an example of a subject-centered scale, and it appeared to contain content that respondents could readily understand and easily answer. The scale comprised eight items, which are shown below. Respondents were asked to answer the questions with the instruction 'please answer using the scale from 1 to *X* where 1 equals strongly disagree and *X* equals strongly agree'. *X* was either 5, 7 or 10 depending on the treatment group. The precise meaning of each scale point was not read out to respondents for any of the three scale formats, whereas normally one would do so for the 5 or 7-point formats. This potentially lowered the utility of those two scale formats, but I wished to have a consistent approach to administering all three of the scales.

TABLE 1 HERE

The number of respondents in each experimental group was: 10-point scale: n=300; 7-point scale n=185; 5-point scale n= 250. The reason for the varying sample sizes for each group is that the study had other unrelated objectives, and the questionnaire programming used to direct respondents into the three scale format groups was also used to direct sample numbers to other question sets, and those groups required different sample sizes. The other survey content did not affect the results reported here.

I considered whether the sample sizes were adequate by calculating how large the difference in mean scores across groups would need to be to achieve statistical significance. I set a difference of half a scale point as the magnitude of difference that the experiment should be able to identify as statistically significant. I had three treatment groups, with the smallest group numbering 185 respondents. To ascertain their adequacy, I conducted a conservative inferential test using simulated data of three groups of n=185. I first generated a series of 185 scores with a mean of 6.0 and a standard deviation of 2.0 using Microsoft excel. These data characteristics were taken from the results of a previous study (Dawes 2002a). I then generated two other series, such that I had three data series that differed by 0.5 scale points to each other. This process was repeated with data series that exhibited progressively smaller differences in mean scores. I found that if there was a mean difference of 0.3 scale points (or more) between each of three groups of this size, an analysis of variance would be statistically significant at the p=0.05 level. Since two of my groups had larger number of respondents than this, the sample sizes appeared to be adequate for the purpose.

The survey sample was broadly representative of the general population, excepting that younger respondents were under-represented. The age breakdown of the sample is shown below. The gender breakdown was 42% male and 58% female. Ideally the survey would have obtained an age and gender breakdown closer to the general population, but in order to do so data collection costs would have increased, which was not feasible for this study. Later I discuss whether the results are comparable for age and gender sub-groups to ensure the results are not biased by the sample.

TABLE 2 HERE

Analysis

Re-scaling

To examine the various data characteristics of interest, it is convenient to re-scale the data so that each scale format is comparable, each with the same upper limit such as out of ten or out of 100. Note that the purpose of this re-scaling is to facilitate comparison between the scale

formats, not to find a specific functional transformation that will minimize any re-scaled differences.

There are a number of straightforward methods by which this could be done. One method is based on a formula used by Preston and Colman (2000). They used the formula (rating -1) / (number of response categories -1)*100. This re-scales to a common score out of 100. For the purpose of this paper we could use the same formula but adapted to be (rating -1) / (number of response categories -1)*10 which re-scales all scale formats to a score out of ten. A feature of this method is that any score using the lowest scale point of any scale becomes zero. For example a score of 1 on a five-point scale would become (1-1) / (5-1) *10 = zero.

Another method is one employed by Dawes (2002a). This is a simple arithmetic procedure whereby the scale end points for the 5 and 7-point versions are anchored to the end points of the ten-point scale. The intervening scale values are inserted at equal numerical intervals. For example, to re-scale the 5-point scale to ten points, one remains as one, five is re-scaled to ten, the mid-point of 3 on the 5-point scale is adjusted to be as per the mid-point between 1 and 10 (namely 5.5) and so on. This is shown below in Table 3.

The second approach has the appealing feature for the present research that the ten-point scale remains unchanged, and the other scales are altered to be comparable to it. However, it results in a slight biasing effect for the lowest scale point. This is because it takes a score of 1 out of 5 or 1 out of 7 and re-scales it to be equivalent to 1 out of 10 – the latter being a lower score in proportional terms. Therefore if there are any responses using these lowest scale points for the five or seven-point formats, the re-scaled score expressed as a mean out of ten will be lower than it was originally. However, preliminary analysis showed that the method based on Preston and Colman (2000) and Dawes (2002a) produced virtually identical results. The Dawes (2002a) method was used because it was slightly simpler.

TABLE 3 HERE

Results

Mean scores
The re-scaled mean scores for each item are shown below, for each of the three scale formats. The data are ordered according to the mean score on the 10-point scale for clarity.

TABLE 4 HERE

The re-scaled 5-point and 7-point scales produced more instances of higher scores compared to the ten-point format. For seven out of the eight questions, the 5-point format (once re-scaled) produced slightly higher scores than the 10-point format. For six out of the eight questions, the 7-point format (once re-scaled) produced slightly higher scores than the 10-point format. There appeared to be little difference between the 5-point and 7-point format.

To test if the overall mean scores from the eight items were statistically significantly different according to scale format, I ran a one-way ANOVA. Since there was virtually no difference between the 5-point and 7-point formats I combined them as one factor. The average re-scaled scores of the eight scale items were the dependent variable, and the factors were the scale formats (5-point re-scaled and 7-point re-scaled combined as one factor, and 10-point

scale comprising the other factor). The result was statistically significant (F=4.1; d.f. 1,733; p=0.04).

Based on this result, it seems that a 10-point scale format will produce slightly lower scores compared to the scores generated from 5-point or 7-point formats, once the latter are re-scaled for comparability.

Variance

I next examined the standard deviation for the re-scaled 5 and 7-point data compared to the 10-point data. If the data is not dependent on the choice of scale format, then once the data is re-scaled to a score out of ten, all three scale formats should exhibit the same standard deviation.

TABLE 5 HERE

Looking across the three scale formats in Table 5, the differences in standard deviation for the individual scale items are quite small, in the order of zero to 0.2. The average difference is only -0.1 when comparing either 7 to 10-point or 5 to 7-point data (with 5 and 7-point formats rescaled). It appears that scale format does not have a marked influence on variation about the mean. To clarify this formally, I tested the overall average score for each format using the Levene test for homogeneity of variance. The test was not significant (Levene Statistic=0.21; d.f. 2,732, p=0.80). Scale format therefore did not have an association with variance in this experiment.

An examination of the standard deviation tells us about the dispersion of scores about the mean for a particular questionnaire item, or variable. It does not, however, tell us about how individual respondents have used the scale. For example, if we ask respondents to answer eight questions using a one to five scale, how many different scale points will they use ? Obviously the precise answer depends on what the questions pertain to. However, researchers would generally want respondents to use more response options over a series of questions, rather than less. The reason is that this indicates those questions are generating discrimination in responses. Therefore, as a supplementary analysis I also examined how many different scale points respondents actually used, and whether this differed according to the scale format. I found that over the eight questions, the average number of scale points used for the five-point scale was 2.9, for the seven-point scale it was 3.6 and for the ten-point scale, respondents used 4.0 different scale points on average. An analysis of variance confirmed that there was a statistically significant difference between the scale formats in terms of the number of scale points used (F=54; d.f. 2,732; $p<0.01$). Therefore, there is evidence that respondents do use more scale points when given a scale format with more response options.

Skewness

Data may be normally distributed, or may be positively skewed or negatively skewed. If the data is negatively skewed this means the data tends to cluster at the ‘high’ end of the scale with a long tail to the lower scale values. The figures for skewness are shown below in Table 6.

TABLE 6 HERE

The data from all three scale formats is negatively skewed. There are some differences among the individual scale items according to scale format, but nothing systematic. In terms

of the skewness of the overall mean score, there is less than one standard error difference between each scale format, therefore this is not statistically significant.

Kurtosis

Kurtosis refers to the shape of the data around the mean and the tails of the distribution. A normal distribution has a kurtosis value of zero. Data that exhibit positive kurtosis are more clustered about the mean ('peaked') and the tails of the distribution are longer. A negative kurtosis score occurs when the data are clustered less around the mean and have shorter tails. A distribution may have the same mean and standard deviation but exhibit different levels of kurtosis. Hypothetical examples of distributions with the same mean and standard deviation but with either zero, positive and negative kurtosis are shown below to help elaborate the term.

FIGURE 1 HERE

The analysis of kurtosis is shown in Table 7. All three scale formats tend to produce data with negative kurtosis scores. There are only minor differences between the scale formats for the individual scale items. The overall score from each scale format exhibits negative kurtosis, and the differences between them are not managerially or statistically significant.

TABLE 7 HERE

Sub-group analysis

As mentioned earlier, the sample used for this experiment was biased somewhat towards older respondents and females. To ensure the results are not influenced by this sample bias, I re-ran the analysis for two sets of subgroups: older vs. younger respondents and male vs. female respondents. There were no significant differences in mean score, variance, skewness or kurtosis within these subgroups. Therefore there is no reason to think that the slight gender bias in the composition of the sample has influenced the overall results.

Discussion and Conclusions

This study conducted a split-sample experiment to assess the impact of scale categories on responses to questions. The study compared data obtained from using 5-point, 7-point and 10-point numerical scale formats. The 5-point and 7-point data were re-scaled to scores out of ten. Once rescaled, the five and seven-point formats tended to produce more instances of higher mean (rescaled) scores compared to the ten-point format. Indeed, an analysis of the aggregated score over the eight question items found the ten-point format produced a 0.3 point lower score, which was statistically significantly different to the other two formats at under the $p=0.05$ level. In terms of the other data characteristics, the three different scale formats exhibited no appreciable differences in terms of standard variation, skewness or kurtosis. The study also found that if a scale with more response options was administered, respondents used more response options.

Based on these findings it seems reasonable to conclude that data gathered from a five-point format can be readily transferred to seven-point equivalency using a simple re-scaling method. If the analyst wishes to compare data from five or seven-point formats to data in a ten-point format, a simple arithmetic adjustment and re-scaling using the method described here produces comparable data. This outcome may be welcome news to those market research departments who ponder whether data gathered using one scale format can be

transformed to make it comparable to another. It also answers a potential question regarding whether results might have conceivably been better (e.g. a higher relative score) had a different scale format been used. The answer appears to be a scale with more response options produces slightly lower scores relative to the upper limit of the scale.

In terms of the other data characteristics, no scale format produced data with markedly lower variances about the mean. This suggests that none of the three formats is less desirable from the viewpoint of obtaining data for that will be used for regression analysis. Kurtosis and skewness were likewise all very similar for each format, therefore either 5, 7 or 10-point scales are all comparable for analytical tools such as confirmatory factor analysis or structural equation models in this respect.

Directions for Future Research

This study examined scale formats that differed in the number of response categories but were all numerical scales. They all required respondents to nominate a number within a specified range. Such numerical scales are but one type of response scale, it is also common for market researchers and academics to ask respondents to use scales that employ only verbal anchors. This paper, therefore, has tackled only one aspect of a wider issue pertaining to the use and comparability of rating scales in market research. More insight into the effect of the number of response categories on the resultant data when using scales that are only verbally anchored would also be a useful addition to current knowledge. Likewise, this study only examined the effect of scale format using telephone survey methodology. There is scope to examine whether the results found here would generalize to other data collection methods such as self-completion or face-to-face.

Acknowledgements

I thank the field team at the Ehrenberg-Bass Institute for their dedicated effort in data collection.

References

Aaker, D.,V. Kumar, and George S. Day. 2004. Marketing Research. New York: John Wiley and Sons.

Bendixen, Mike, and Merle Sandler. 1994. Converting Verbal Scales to Interval Scales Using Correspondence Analysis. Johannesburg: University of Witwatersrand.

Bruner, G. C., and P. J. Hensel. 1992. Marketing Scales Handbook. Chicago: American Marketing Association.

Burns, Alvin C., and Ronald F. Bush. 2000. Marketing Research: Prentice Hall.

Dawes, John G. 2002a. Five point vs. eleven point scales: does it make a difference to data characteristics? Australasian Journal of Market Research 10 (1): 39 - 47.

Dawes, John G. 2002b. Survey Responses Using Scale Categories Follow a "Double Jeopardy" Pattern. In *ANZMAC*. Eds. R. N. Shaw, S. Adam and H. McDonald. Melbourne.

Dillon, William R, Thomas J Madden, and Neil H Firtle. 1993. Essentials of Marketing Research. Homewood: Irwin.

Finn, R.H. 1972. Effects of Some Variations in Rating Scale Characteristics on the Means and Reliabilities of Ratings. Educational and Psychological Measurement 32: 255-265.

Ghiselli, Edwin E. 1939. All or none versus graded response questionnaires. Journal of Applied Psychology 23 (June): 405-415.

Hair, Joseph F., Jr. , Robert P. Bush, and David J. Ortinau. 2006. Marketing Research. Boston: McGraw Hill.

Johnson, Steven M. , Patricia C. Smith, and Susan M. Tucker. 1982. Response format of the job descriptive index: Assessment of reliability and validity by the multitrait-multimethod matrix. Journal of Applied Psychology 67 (4): 500-505.

Kennedy, Rachel, Christopher Riquier, and Byron Sharp. 1996. Practical Applications of Correspondence Analysis to Categorical Data in Market Research. Journal of Targeting, Measurement and Analysis for Marketing 5 (No. 1): 56-70.

Loken, Barbara, Phyllis Pirie, Karen A. Virnig, Ronald L. Hinkle, and Charles T. Salmon. 1987. The Use of 0-10 Scales in Telephone Surveys. Journal of the Market Research Society 29 (No. 3, July): 353-362.

Malhotra, Naresh, and Mark Peterson. 2006. Basic marketing research: A decision-making approach. Upper Saddle River, NJ: Prentice Hall.

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

Preston, Carolyn C., and Andrew M. Colman. 2000. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. Acta Psychologica 104: 1-15.

Figure 1. Examples of distributions with same mean but different kurtosis

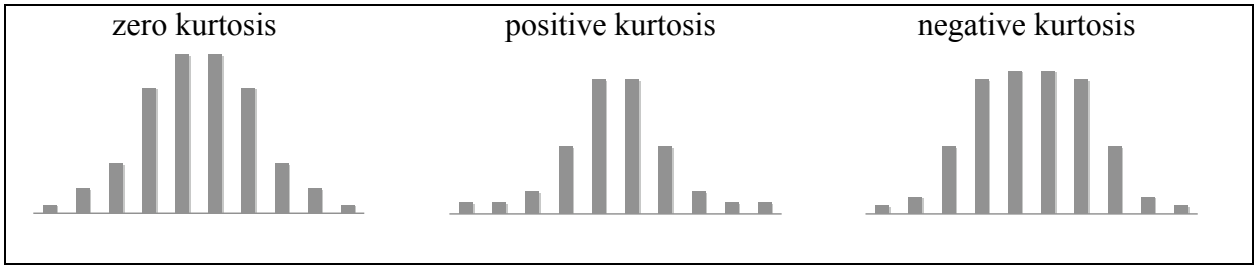

Table 1. Scale items

Item #	Statement
1	When I am in a shop I will always check prices on alternatives before I buy
2	When I buy or shop, I really look for specials
3	I usually watch ads for announcements of sales
4	I believe a person can save a lot of money by shopping around for bargains
5	In a store, I check the prices, even when I am buying inexpensive items
6	I pay attention to sales and specials
7	Clothing, furniture, or appliances ... whatever I buy, I shop around to get the best prices
8	I often wait to purchase items, so I can get them on sale

Table 2. Survey Breakdown

Age category	Sample N	% of sample
Under 21 years	33	5
21 to 30 years	85	12
31 to 40 years	136	19
41 to 50 years	184	25
51 to 60 years	148	20
Over 60 years	149	20
Total	**734**	

Table 3. Re-scaling method for this study

Five-point scale		Seven-point scale		Ten-point scale	
Original value	Re-scaled value	Original value	Re-scaled value	Original value	Scale value
1	1.0	1	1.0	1	unaltered
2	3.25	2	2.5	2	”
3	5.5	3	4.0	3	”
4	7.75	4	5.5	4	”
5	10	5	7.0	5	”
		6	8.5	6	”
		7	10	7	”
				8	”
				9	”
				10	”

Table 4. Mean Scores according to Scale Format

Scale Item	Mean score: 5-point data re-scaled to /10	Mean score: 7-point data re-scaled to /10	10-point data	Mean score 5 point rescaled minus 10 point	Mean score 7 point rescaled minus 10 point	Mean score 5 point rescaled minus 7 point
1	7.8	8.1	7.4	0.4	0.7	-0.3
2	7.4	7.3	6.9	0.5	0.4	0.1
3	5.1	4.6	4.8	0.3	-0.2	0.5
4	7.9	8.1	7.4	0.5	0.7	-0.2
5	6.8	6.9	6.6	0.2	0.3	-0.1
6	7.0	6.9	6.6	0.4	0.3	0.1
7	7.1	7.3	7.6	-0.5	-0.3	-0.2
8	5.9	6.0	5.3	0.6	0.7	-0.1
Overall score (average of all eight items)	**6.9**	**6.9**	**6.6**	**0.3**	**0.3**	**0.0**

** statistically significant difference to the other two formats at p=0.04.*

Table 5. Standard Deviation according to Scale Format

Scale Item	Standard Deviation: 5 point re-scaled /10	Standard Deviation: 7 point re-scaled /10	Standard Deviation: original 10-point data	Std. Dev. 5 point rescaled minus 10 point	Std. Dev. 7 point rescaled minus 10 point	Std. Dev. 5 point rescaled minus 7 point rescaled
1	2.7	2.4	2.5	0.2	-0.1	0.3
2	2.7	2.9	2.7	0.0	0.2	-0.2
3	3.2	3.1	3.1	0.1	0.0	0.1
4	2.4	2.3	2.6	-0.2	-0.3	0.1
5	3.0	2.9	2.8	0.2	0.1	0.1
6	2.7	2.7	2.7	0.0	0.0	0.0
7	2.7	2.6	2.4	0.3	0.2	0.1
8	2.9	2.9	2.8	0.1	0.1	0.0
Overall score (average of all eight items)	**2.0**	**1.9**	**2.0**	**0.0**	**-0.1**	**0.1**

Table 6. Skewness according to Scale Format

Scale Item	Skewness 5 point re-scaled /10 (std error = .14 all items)	Skewness 7 point re-scaled /10 (std error = .16 all items)	Skewness original 10-point data (std error = .16 all items)	Skewness 5 point rescaled minus 10 point	Skewness 7 point rescaled minus 10 point	Skewness 5 point rescaled minus 7 point rescaled
1	-1.2	-1.4	-0.8	-0.4	-0.6	0.2
2	-0.8	-0.8	-0.6	-0.2	-0.2	0.0
3	0.2	0.4	0.3	-0.1	0.1	-0.2
4	-1.1	-1.4	-0.9	-0.2	-0.5	0.3
5	-0.7	-0.7	-0.5	-0.2	-0.2	0.0
6	-0.7	-0.6	-0.5	-0.2	-0.1	-0.1
7	-0.6	-0.8	-1	0.4	0.2	0.2
8	-0.1	-0.2	0.0	-0.1	-0.2	0.1
Overall score (average of all 8 items)	**-0.5**	**-0.4**	**-0.4**	**-0.1**	**0.0**	**-0.1**

Table 7. Kurtosis according to Scale Format

Scale Item	Kurtosis 5 point re-scaled /10 (std error = 0.47 all items)	Kurtosis 7 point re-scaled /10 (std error = 0.47 all items)	Kurtosis original 10- point data (std error = 0.28 all items)	Kurtosis: 5 point rescaled minus 10 point	Kurtosis: 7 point rescaled minus 10 point	Kurtosis: 5 point rescaled minus 7 point rescaled
1	0.6	1.3	-0.1	0.7	1.4	-0.7
2	-0.4	-0.5	-0.7	0.3	0.2	0.1
3	-1.3	-1.1	-1.3	0.0	0.2	-0.2
4	0.3	1.3	-0.2	0.5	1.5	-1.0
5	-0.7	-0.6	-0.8	0.1	0.2	-0.1
6	-0.4	-0.5	-0.8	0.4	0.3	0.1
7	-0.8	-0.2	0.2	-1.0	-0.4	-0.6
8	-1.1	-1.0	-1.1	0.0	0.1	-0.1
Overall score (average of all 8 items)	**-0.4**	**-0.5**	**-0.4**	**0.0**	**-0.1**	**0.1**