

## Assessing the Applicability of Self-Anchoring Scales in Web Surveys

Frederik Van Acker, Peter Theuns

Work and Organizational Psychology, Vrije Universiteit Brussel, Brussels, 1050, Belgium

### Abstract

Self-anchoring scales were first mentioned by Kilpatrick and Cantril (1960) and Cantril (1965) as rating instruments in which the end anchors are defined by the respondent himself, basing on his own assumptions, perceptions, goals and values. The uses of these scales are legion and they have proven to be very useful in reducing measurement bias in cross-cultural research (Cantril, 1965; Bernheim, Theuns, Mazaheri, Hofmans, Fliege & Rose, 2006). In the first part of this study we investigate whether context effects can be lessened or eliminated by using self-anchored scales. For this purpose, an experiment similar to the one by Tourangeau, Couper and Conrad (2004), in which they manipulated images that figured in a web survey, was conducted. The hypothesis that self-anchoring scales can reduce contextual bias, is not supported by our data. The second part of the study investigates if and how self-anchoring scales affect drop-out during the filling out of questionnaires. It is found that (compared to a regular rating scale, a larger proportion of respondents drop-out. Moreover subjective preferences for one or the other scale do not seem to differ.

**KEY WORDS:** self anchoring scales, web surveys, context effects

### 1. Introduction

#### 1.1 Rating scales

When researchers seek answers to how people feel or think about certain aspects of life or society, the most practical method to perform a quantitative study is to ask respondents to express their opinion on a rating scale. Rating scales are widely used instruments to record subjective data. Gannon and Ostrom (1996) discern four types of rating scales, namely bipolar, unipolar, agree/disagree and probability scales. These scales differ only in the labels that are attached to the scale points. For most rating scales these labels are fixed. If, for example, we would ask people how happy they are with their life as a whole, we could use a seven point rating scale with the labels “extremely happy” and “extremely unhappy” printed at the scale ends. Using such rating scale, Nigeria was found to be the country with the highest percentage of happy people amongst 65 countries that were surveyed (Inglehart, 2004). However, this first place is rather questionable given the current rates of violence, poverty

and emigration in this region (Erubami, & Young, 2003). One problem with fixed verbal scale labels is that these labels may be interpreted differently by different respondents. Common labels in subjective wellbeing (SWB) studies such as ‘best’ and ‘worst imaginable’ can have several meanings, depending on the frame of reference of the respondents (Diener and Diener, 1995). For the one person “best” may refer to good health, whereas and for some other person “best” may mean “wealth”. This difference in frame of reference would make comparisons between SWB ratings difficult to interpret.

An alternative to scales with fixed anchors are self-anchoring scales. These rating scales were first defined and used by Kilpatrick and Cantril (1960) and Cantril (1965). When applying these scales, a person is asked to define the two extremes, basing on his own assumptions, goals and values. In order to establish how happy someone feels, a researcher could ask this person to define the one extreme as the worst period in his life and the other extreme as the best period in his life. Bernheim et al. (2006) mention that most respondents refer to one of a small collection of situations to define either their best or their worst period in life. The best period of one’s life is mostly associated with love, marriage or the birth of a child. The worst period would then be defined as the death of a close relative, memories of war or a serious illness. There is no reason to believe that these events would be different across people or cultures, although we do not know of any studies confirming this. The expectation is that by using self defined anchors, the frames of reference of most respondents will be more alike and thus comparisons between different groups would be made possible.

#### 1.2 Rating in the context of the web

Web surveys have become extremely popular because they are a cost effective tool to quickly collect data. As is the case with paper and pencil questionnaires, most responses in web studies are obtained by means of rating scales. There is however one important difference between traditional surveys and electronic surveys. In traditional mail or face to face surveys researchers have some control over the appearance of the survey on the one hand, and over the context of the survey on the other hand. For example, we know that when a mail survey is sent to people’s home address, they will probably fill in

the survey at this location. Connectivity to the internet has boomed this seriously in the last few years that with web surveys there is no way of knowing where people actually take the questionnaire. A virtual mailbox can be consulted wherever one has internet access, whether it be on a fixed computer, a laptop or even more portable devices such as mobile phones and handhelds. Each of these devices enables potential participants to take a web survey.

It seems that little research has paid attention to this possible effect of location although there are reasons enough to believe that this could impact on ratings. Tourangeau, Rips, and Rasinski (2000, p 206) note that “one of the key assumptions of the belief-sampling model for answers is that respondents only tap a small portion of the potentially relevant considerations in formulating their answers”. Respondents tend to process mainly the information that comes readily to mind. Considerations that are easily accessible will be used to formulate answers. This means that prior questions on a certain topic influence answers on later questions as this information has been made more easily accessible (Tourangeau, Rasinski, & D’Andrade, 1991). However, we believe that all relevant information that is readily available to the respondent may bias his response behavior. Participating in a survey on purchasing behavior inside a store may elicit different responses than when the same questions are considered at home. In general, all available contextual cues may impact on responses. In an experiment where questions on purchasing behavior were either accompanied by a picture of low frequency instances of this behavior (e.g. going to a shoe store) or high frequency instances (e.g. going to a grocery store) it was found that participants were influenced by these images and reported higher frequencies of shopping when pictures showed high frequency instances (Couper, Tourangeau, and Kenyon, 2004). In the same study, questions on subjective quality of life were presented to participants. These were either accompanied by a negatively contrasting image (a healthy young woman jogging) or a positively contrasting picture (a woman lying in a hospital bed). Here too respondents’ mean subjective wellbeing ratings were influenced by the available contextual information Respondents who were presented the picture of a healthy woman jogging had lower ratings than those who saw the woman lying in a hospital bed (Couper, Tourangeau, & Kenyon, 2004). Such context effects can be described as “directional” since they cause a shift of responses in a certain direction (Tourangeau, Rips, & Rasinski, 2000). It seems that providing people with relevant information can influence their frame of reference and hence causes a shift in their response.

Previous research has shown that questions and images can bias responses, but even the weather conditions can

impact on how people rate certain questions (Schwarz and Clore, 1983). In one experiment it was shown that mood, manipulated by calling respondents either on a sunny or a rainy day, influences ratings of subjective wellbeing. The authors showed that rainy weather induced a negative mood which impacted negatively on happiness ratings.

There seems to be legion evidence that ratings are subject to contextual influences, thereby endangering the validity of the rating method. The present study investigates whether self anchoring scales can reduce contextual bias and so increase the validity of ratings. In the case of subjective wellbeing the influence of, for example, bad weather is expected to be less pronounced with self anchoring scales, given that both extremely negative and extremely positive situations need to be described by each participant. It is expected that the anchoring procedure in itself creates a (personalized and therefore strong) context for the rating which may be less prone to external contextual cues. Since we expect the anchoring procedure to be responsible for lessened context effects, this should not only be found in items rated on a category scale, but also on open ended questions. A possible disadvantage of using qualitative descriptions, as could be expected according to the availability heuristic (Tversky and Kahneman, 1973) is that participants may overestimate the frequency of the described events, possibly resulting in biased ratings.

### 1.3 Nonresponse in web surveys

Since the beginning of the 21<sup>st</sup> century researchers have tried to identify strategies to augment the number participants in web studies. From social exchange theory (Homans, 1958) we could predict that the perceived cost of responding to a survey impacts negatively on response rates. This has been confirmed in a series of experiments from which was concluded that a higher perceived burden will have a negative effect on the number of completely filled-in questionnaires (Crawford, Couper, and Lamias, 2001, Heerwegh, 2006).

Filling in open ended questions typically requires more cognitive effort than marking an answer on a rating scale. Using self anchoring scales typically requires participants to give detailed qualitative descriptions of the end anchors. Although self anchoring scales may be able to reduce bias, they are more demanding for respondents. This would lead us to think that more respondents will abandon the survey if they are confronted with self anchoring scales compared to fixed anchor rating scales. The aim of most researchers is to reduce measurement error in their surveys by applying the right scale and asking the right questions. However, measurement error is not the only issue: data quality also depends on unit or item nonresponse. Therefore, this study not only focuses

on the ability of self anchoring scales to reduce context bias, but also considers whether this potential benefit outweighs the cost of possibly losing respondents.

## 2. Method

A snowball sample of 266 students and non students was invited by e-mail to participate in a web survey on three topics: quality of life, shopping experiences and dining out. Participants were aged 28.65 on average ( $sd = 12.17$ ) and 79.4% were women.

Two factors were manipulated in this experiment according to a  $2 \times 4$  (rating scale  $\times$  context) factorial design and by clicking the link in the e-mail invitation, participants were assigned to one of the eight conditions. Half of the participants were asked to rate several

pictures were added after the respondent had completed the self anchoring procedure.

Next, a series of questions on shopping and eating out, derived from Preston and Colman (2000) were presented to the participants, either without any images, or combined with a picture of high frequency behavior, low frequency behavior or both. Participants who were assigned to the “no picture”, the “negatively contrasting”, the “positively contrasting” and the “both pictures” condition for the questions on SWB, were respectively assigned to the no picture, “low frequency”, “high frequency” and “both pictures” condition for this series of questions.

The main interest was on whether using self anchoring scales in web surveys could reduce context bias. A main

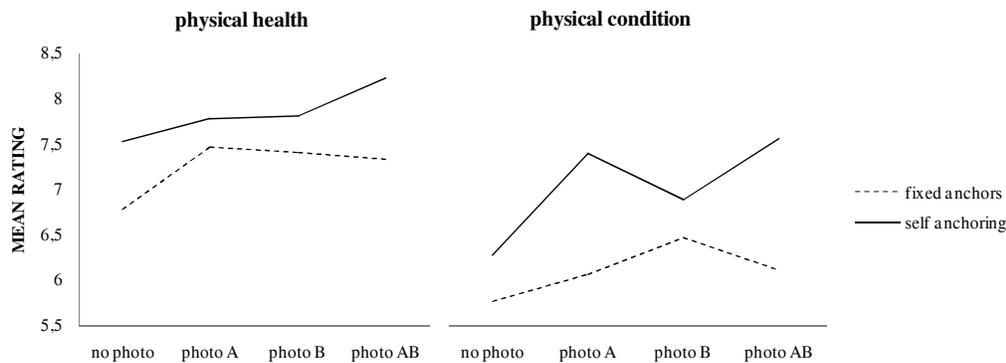


Figure 1: plots of mean ratings across conditions for Q4 (left panel) and Q5 (right panel); photo A was a picture of a hospitalized woman, photo B presented a healthy woman doing sports.

statements on a fixed anchor 10-point rating scale with end anchors defined as “very dissatisfied” and “very satisfied”. The other half were instructed to describe their own end anchors, and thus used a 10-point self anchoring scale. The latter participants were specifically asked to write down how their worst and best possible, but still realistic life standard would be like. These descriptions were then used as end anchors for the questions to be answered on subjective wellbeing. For the other questions, participants had to describe their best and worst possible shopping and eating out experience.

Subjects were first instructed to respond to several questions about subjective wellbeing. These items were adopted from Mazaheri and Theuns (2006) and had proven to discriminate well between respondents. The context factor comprised four levels and was manipulated by including pictures in the survey. In one condition there was no picture in the questionnaire. In the other conditions there was either a picture that was negatively contrasting, positively contrasting or both were put at the top of the webpage. In the self anchoring condition

effect of context was expected as was found in the experiments of Couper, Tourangeau, & Kenyon (2004). Moreover, if self anchoring scales can reduce this effect, a significant interaction between context and scale is expected. The analysis will focus on several questions as dependent variable. An overview of these questions can be found in Table 1. Two of these questions were to be rated on a self anchoring scale. For the three other questions, respondents were instructed to enter a number representing their answer.

Finally, near the end of the survey, participants were asked to judge three aspects of the questionnaire with a rating out of 100. One of these aspects was the ease with which they were able to fill in the survey. A second rating was the subjective speed with which they took the survey. A final question inquired upon the extent to which they were able to express their personal opinion in the survey.

Table 1: Items that were included in the analysis. Questions marked with a \* were rated either on a self anchoring rating scale or a fixed anchor rating scale

Q1: How many times have you been eating out in the past month?	Context × Scale	3	209	1.76	.16
Q2: How many times have you been shopping in the past month?	Q3: Could you rate your general life satisfaction with a number out of 100?				
Q3: Could you rate your general life satisfaction with a number out of 100?	Context	1	214	.06	.98
Q4: How satisfied are you with your physical health?*	Scale	3	214	1.16	.28
Q5: How satisfied are you with your physical condition?*	Context × Scale	3	214	.58	.63

### 3. Results

First, we examined whether the questions rated on a self anchoring or a fixed anchors scale showed the expected main effect of context. This effect was found for question 5 ( $F(3, 225) = 8.41, p = .05$ ) but not for question 4 ( $F(3, 225) = 4.40, p = .22$ ). No significant scale × context interaction was found for either question (for Q4  $F(3, 225) = .32, p = .81$  and for Q5  $F(3, 225) = 1.22, p = .30$ ). However, the type of rating scale used had a significant main effect for both items (for Q4  $F(1, 225) = 6.36, p = .01$  and for Q5  $F(1, 225) = 14.837, p < .001$ ). Figure 1 gives the factorial plots for these findings.

Figure 2 and Table 2 show the results of similar analyses that were performed on items Q1 through Q3 which were open ended.

Next, we were interested in whether defining the end anchors induced a larger number of respondent drop-out. In the fixed anchor condition, only 7.86% quit the survey before completing it entirely, compared to 26.98% in the self anchoring condition, which was a significant increase ( $\chi^2(1) = 17.26, p < .001; n = 266$ ). The time required to complete the survey differed significantly between the self anchoring ( $m = 15.18; sd = 9.17$ ) and the fixed anchors version ( $m = 4.81; sd = 2.88$ ) ( $t(101.84) = 10.44, p < .001$ ). Respondents rated the survey speed as significantly lower on the self anchored questionnaire than on the fixed anchor version ( $t(120.42) = -7.31, p < .001$ ). The ease with which they could respond to the survey questions was also rated significantly lower in the self anchored condition than in the fixed anchors condition ( $t(132.55) = -7.31, p < .001$ ). The extent to which respondents thought they could express their

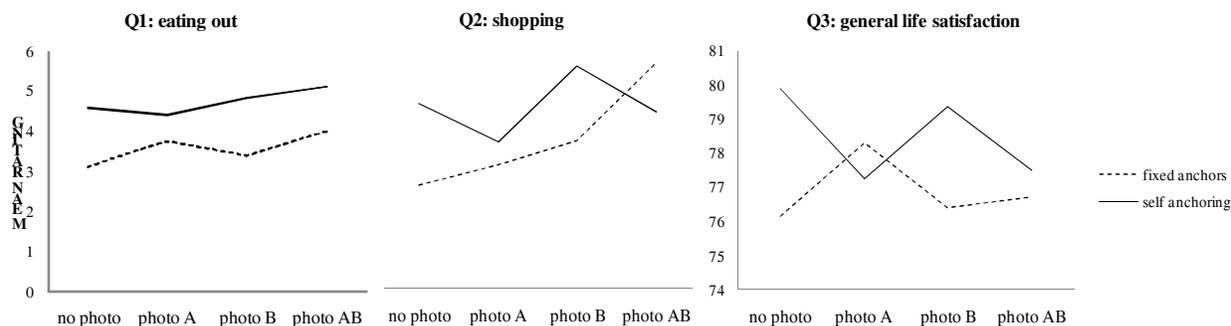


Figure 2: plots of mean ratings across conditions for questions Q1, Q2 & Q3.

Table 2: Results from the ANOVA on Q1, Q2 & Q3.

	<i>df</i> <sub>effect</sub>	<i>df</i> <sub>error</sub>	<i>F</i>	<i>sign.</i>
Q1: How many times have you been eating out in the past month?				
Context	1	233	.36	.78
Scale	3	233	6.16	.01
Context × Scale	3	233	.17	.92
Q2: How many times have you been shopping in the past month?				
Context	1	209	2.10	.10
Scale	3	209	2.23	.14

personal opinion was only marginally influenced by the scale type. Ratings on the survey with self anchoring scales were slightly higher than those obtained with fixed anchors ( $t(216) = 1.80, p = .07$ ). An overview of these findings can be found in Table 3.

Table 3: Means and standard deviations of ratings concerning the survey in terms of speed of completion, ease of completion and extent to which one could express his or her opinion.

Item	Mean (sd)	
	self anchored	fixed anchors
ease	72.32 ( $\pm 20.54$ )	87.40 ( $\pm 11.65$ )
speed	74.03 ( $\pm 17.84$ )	88.81 ( $\pm 8.68$ )
express opinion	81.24 ( $\pm 17.38$ )	76.97 ( $\pm 17.27$ )

### 3. Discussion

The main research question of this study was whether self-anchoring scales can help researchers in dealing with context effects. These context effects have been found with several cues, such as preceding questions, images and weather conditions (Tourangeau, Rasinski, & D'Andrade, 1991; Couper, Tourangeau, & Kenyon, 2004; Schwarz & Clore, 1983). We will discuss the outcome of our experiment for the consecutive topics that were covered in the survey

#### *Subjective wellbeing*

The research instrument contained three items of interest for this study, two of which were items to be rated on a self anchoring scale. The third item was an open ended (fill in a number) question. We expected that the images provided in six out of the eight conditions of the experiment would produce a context effect. Participants would contrast their situation against the information presented in the pictures (healthy person doing sports, a hospitalized woman or both) and this should produce a shift in mean responses. There was a marginal effect of context for the question on physical condition. For the other two questions we were not able to reproduce a main effect for context. To see whether ratings on self-anchoring scales produced smaller shifts, the context  $\times$  scale interactions were tested statistically. For none of the questions on well-being the interaction reached statistical significance, thereby providing no evidence that self anchoring scales would be less susceptible to contextual biases.

A rather interesting finding is that in both the question on physical condition and the one on physical health the scale produced a significant main effect with ratings on self anchoring scales being on average significantly higher. This effect was however not found for the open ended question. It seems that the reflection upon the end anchors shifts ratings of physical health and condition towards the positive end. This shift could be explained by the design of the survey. Since all participants described their worst period first, followed by the best period, this latter information was more readily available during the ratings. Respondents could then, according to the availability heuristic, falsely judge that positive events occur more frequently than is really the case (Tversky, & Kahneman, 1973). Another possible explanation is that the ratings are contrasted against the negative description, thus producing more positive responses. It would be

interesting to see in future research if counterbalancing the order in which the end anchors are described can control for this effect, ruling out the first explanation.

#### *Shopping and eating out*

We were interested to see whether using self anchoring scales could possibly reduce context effects produced by images for other topics. Therefore, participants were presented with two open ended questions on their shopping and eating-out-behavior. Contrary to our expectations, we did not find an effect of context, although Figure 2 clearly shows a trend, especially for the shopping behavior. We were not able to demonstrate the expected interaction either. However, in the open ended question on eating out there was a shift in responses caused by the rating scale, as was found for the two questions on physical health and condition. It seems that reflection upon the end anchors does not only alter responses given on a rating scale, but also reporting of subjective frequency.

#### *Nonresponse and subjective ratings*

Nonresponse greatly increased in the self anchoring group. We observed about 20% more break-off in the self anchoring condition. Basing on social exchange theory we expected that the perceived cost would be higher in the self anchoring condition. The cognitive effort required to respond to the anchoring questions is relatively high, which was also reflected in the response times and the subjective ratings given at the end of the survey. Although the survey was relatively short, respondents needed an extra 10 minutes on average to complete the self anchoring survey compared to the fixed anchor version. Evaluations of the subjective speed and ease with which the survey was taken were significantly lower. Although respondents have the chance to elaborate on the choice of the end anchors, this did not influence the way they reported to what extent they were able to express their opinion, compared to fixed anchor scales.

#### *General conclusion*

In the current study we found no support for the hypothesis that self anchoring scales can reduce contextual influences caused by adding images to a survey. We must note however, that the expected context effect was not found for each topic in the survey, although this was expected, basing on Couper, Tourangeau, & Kenyon (2004). By applying self anchoring scales to our survey, however, it seemed that another context effect was found. Some ratings performed after end anchors had been described were higher than when using fixed anchor scales. An interesting question that remains is which condition yields the most valid response. However, given the large proportion of respondents that decided to leave the survey, it seems that self anchoring scales have no real benefits in web surveys compared to fixed anchor scales.

### Acknowledgements

This research was supported by a grant of the Onderzoeksrraad from the Vrije Universiteit Brussel OZR1386BOF.

### References

- Bernheim, J.L., Theuns, P., Mazaheri, M., Hofmans, J., Fliege, H. & Rose, M. (2006). The potential of Anamnestic Comparative Self-Assessment (ACSA) to reduce bias in the measurement of Subjective Well-being. *Journal of Happiness Studies*, 7, 227-250.
- Cantril, H. (1965). *The pattern of human concerns*. New Jersey: Rutgers University Press.
- Couper, M. P., Tourangeau, R., & Kenyon, K. (2004). Picture this! Exploring visual effects in web surveys. *Public Opinion Quarterly*, 68, 2, 255-266.
- Crawford, S. P., Couper, M. P., & Lamias M. J. (2001). Web surveys, perceptions of burden. *Social Science Computer Review*, 19, 146-162.
- Diener, E., & Diener, M. (1995). Cross-cultural correlates of life satisfaction and self-esteem. *Journal of Personality and Social Psychology*, 68, 653-663.
- Erubami, M., & Young, I.R. (2003). Nigerian violence: A review of statistics and perceptions. Centre for Human Rights Research and Development, CHRRD RESEARCH REVIEW No. 4, <http://www.chrrd.kabissa.org/nigeria-violence-review.htm>.
- Gannon, K. M., & Ostrom, T. M. (1996). How meaning is given to rating scales: the effects of response language on category activation. *Journal of Experimental Social Psychology*, 32, 337-360.
- Heerwegh, D., Loosveldt, G. (2006). An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in web surveys. *Journal of Official Statistics*, 22, 191-210.
- Homans, G. C. (1958). Social Behavior as Exchange. *American Journal of Sociology*, 63, 597-606.
- Inglehart, R. (2004). World Values Survey. <http://www.worldvaluessurvey.org/news/index.html>.
- Kilpatrick, F.P., & Cantril, H. (1960). Self-anchoring scale: A measure of the individual's unique reality world. *Journal of Individual Psychology*, 16, 158-170.
- Mazaheri, M., & Theuns, P. (2006). A comparison of different formats of the anamnestic comparative self assessment (ACSA) for the assessment of subjective well-being. *PRO Newsletter*, 36, 10-13.
- Preston, C. C., & Colman, M. C. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica*, 104, 1-15.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgements of well-being: informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513-523.
- Tourangeau, R., Rasinski, K., & D'Andrade, R. (1991). Attitude structure and belief accessibility. *Journal of Experimental Social Psychology*, 27, 48-75.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press: New York.
- Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.