ARTICLES

Exploring Animated Faces Scales in Web Surveys: Drawbacks and Prospects

Matthias Emde[1], Marek Fuchs[1]

[1] Darmstadt University of Technology

Web surveys have been mimicking paper questionnaires with respect to their layout and appearance for a long time. Even though the rapid development of the internet and internet data collection methods offers various graphic and multimedia design features, very little is known about the influence of animated web survey questions on the question answering process. In a web survey among university journal readers, we conducted an experiment exploring the effects of implementing animated faces in scale questions. By varying the visual appearance of faces in a scale question, we enhance their influence on the question answering process.

Introduction

Rating scales with faces or "smileys" as symbolic labels are frequently used in questionnaires on job satisfaction (Herman, Dunham, and Hulin 1975; Jäger and Bortz 2001; Kunin 1998) and global well-being (Andrews and Withey, 1976; Wanous, Reichers, and Hudy 1997). They are also considered especially suitable for surveying children, as they are more easily understood in comparison to text based self-report measures (Chambers et al. 1999). The advantage of these scales is mostly seen in the easier formatting of affective answers. Global self-report measures ask respondents about complex constructs like general satisfaction on broad categories and over long periods of time like one's lifetime. Applied to the question answering process the retrieval of relevant information on global questions is nearly impossible (Schwarz and Strack 1991). Instead of relevant information, accessible information is used to generate an answer. In this case, answers are to a greater extent affective and fit easier into an affective answer scale like a faces scale. The translation of feelings into words is not necessary and the respondent only has to "check the face which looks like he feels" (Kunin 1998, 824).

Even though faces scales are used in web surveys, the present findings for these scales are mostly based on paper questionnaire experiments. The visual design advantages of web surveys can be seen as a valuable addition to the present purpose of these scales. Using faces scales implies using graphical elements. Pictures in surveys attract attention (Couper, Tourangeau, and Conrad 2007) and affect answers particularly when visual and verbal information does not

match the presented question (Couper 2001). Moreover, surveys consist of words, but they also imply a visual language including symbols, numbers and graphics, which influence answers to survey questions (Christian and Dillman 2004). Even though visual content might increase respondent enjoyment, Couper, Tourangeau, and Kenyon (2004) found only little support for this hypothesis.

Our study was designed to explore whether faces scales are appropriate to measure general satisfaction. We hypothesized that the easier formatting of affective answers to a faces scale would apply especially when the faces are animated and change their visual appearance. To better understand the characteristics of a faces scale we strengthen their affective aspect by animating the faces visual appearance. Apart from the easier formatting we employ faces scales to attract attention and increase respondents' enjoyment. If they do, respondents will spend more time to answer the question, which allows for deeper question processing and therefore increases data quality.

Methods

Our study was carried out in a survey among university journal readers and non-readers (N=1042) using a mixed-mode design of paper and web based questionnaires. Results reported were based on the web survey (N=611). Web survey respondents who read the journal answered a question using radio buttons concerning their global satisfaction with the journal in the middle of the questionnaire. Furthermore, respondents were randomly assigned to one out of three versions of the same satisfaction question measuring the overall satisfaction with the university journal, at the end of the questionnaire: a fixed design, an affective design, and a cognitive design of a faces scale.

Figure 1 illustrates the three faces scale designs and the radio button control question. The fixed design included no animation at all and mimics commonly used faces scales in paper and pencil questionnaires. In the affective design, the faces changed their color (blazing red to red-orange to dark orange to light orange to light green to grass-green) and increased their size with the cursor hovering over, while in the cognitive version the faces did not change their color and zoomed out and a text answer category was displayed. In the affective design we enhanced the emotional context and redeployed attention to the faces, while in the cognitive setting we accentuated cognitive processing by downsizing faces and offering an additional text label for the answer category. The radio button control design used the same answer categories as the cognitive design but included no animation at all. As the faces scales were part of a contract work survey, we were forced to use a 6 point scale even though a middle response option seemed to be more appropriate. On the other hand, we avoided the use of a neutral face of questionable adequacy with a straight mouth line (Elfering and Grebner 2010).
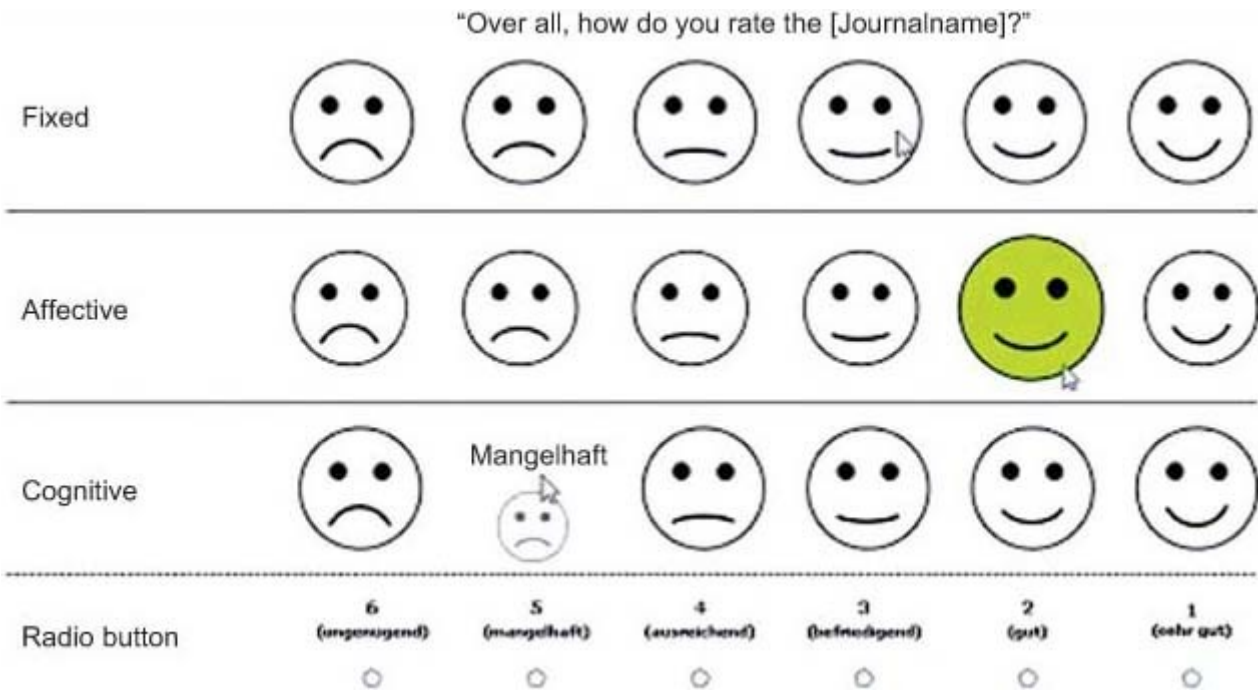
Figure 1 Faces scale design.

Results

Table 1 shows the distribution of responses for the three faces scale designs and the radio button version. We found no significant differences within the three face scale designs. Comparing the three faces to the radio button scale, we found significant differences between the fixed and the affective design, while the cognitive design was not significantly different from the radio button version. Moreover, answers to the cognitive and the radio button design were slightly more positive (mean = 2.5) than answers to the fixed and affective design (mean = 2.6).

Table 2 reveals the time span (measured in seconds) respondents needed to answer the questions. When answering one of the three faces designs (fixed/affective/cognitive) respondents needed about 14 seconds on average to select their answer and to click the submit button. Again we found no significant differences among the three faces scales. Comparing the radio button version to each of the faces scales yielded significant differences: the radio button version was on average four seconds faster.

In summary, the results of our study show similarities in the answer distributions for the fixed and the affective design on the one hand and for the cognitive and the radio button design on the other hand. Furthermore, all faces scale versions took respondents longer to answer in comparison to the radio button design.

Table 1   Faces scales vs. radio button.

	Fixed	Affective	Cognitive	Radio
1 (very good)	12.1%	7.6%	4.5%	4.6%
	(12)	(7)	(4)	(13)
2 (good)	36.4%	41.3%	52.8%	52.8%
	(36)	(38)	(47)	(150)
3 (satisfactory)	36.4%	40.2%	32.6%	31.3%
	(36)	(37)	(29)	(89)
4 (adequate)	10.1%	6.5%	5.6%	7.0%
	(10)	(6)	(5)	(20)
5 (inadequate)	4.0%	2.2%	4.5%	3.2%
	(4)	(2)	(4)	(9)
6 (unsatisfactory)	1.0%	2.2%	0.0%	1.1%
	(1)	(2)	(0)	(3)
Total	100.0%	100.0%	100.0%	100.0%
	(99)	(92)	(89)	(284)
Mean	2.6	2.6	2.5	2.5

Note. Chi-squared test comparing fixed, affective and cognitive faces scale: n.s.

Chi-squared test comparing radio button and faces scales:

$p < 0.01$ radio button compared to fixed design;

$p < 0.05$ radio button compared to affective design;

n.s. radio button compared to cognitive design.

Table 2   Time needed to answer (outliers excluded).

	Fixed	Affective	Cognitive	Radio
Min. time	5 sec.	3 sec.	3 sec.	2 sec.
Max. time	52 sec.	46 sec.	42 sec.	48 sec.
Mean	14.7 sec.	13.9 sec.	13.6 sec.	9.4 sec.
Median	12 sec.	13 sec.	12 sec.	8 sec.
Total	78	92	88	289

Note. T-test comparing fixed, affective and cognitive faces scale: n.s.

T-test comparing radio button and faces scales:

$p < 0.01$ radio button compared to fixed design;

$p < 0.01$ radio button compared to affective design;

$p < 0.01$ radio button compared to cognitive design.

Discussion

Even though we found differences between the faces scale designs, there was surprisingly no significant influence of face color and size on the answers provided in comparison to the fixed faces scale design. We hypothesize that a change in color and size might be not enough to make a faces scale a more affective measure. The differences in the distributions reveal slightly lower

satisfaction for the fixed and affective faces scale designs. However, the key finding of this study is that the cognitive faces scale design provides corresponding answers to the radio button question. We therefore suggest that the cognitive design (using faces and text) can be used instead of a radio button scale for questions on global satisfaction.

Due to the fact that respondents needed more time to answer each of our faces scale designs, we assume that faces scales trigger more attention. If the focus of the respondents is needed especially on the answer categories, we assume faces scales could provide that. On the other hand, faces scales might draw attention away from the question itself, which may cause problems, especially with complex worded questions. Based on the survey design (we had to put the faces scales at the very end of the questionnaire), we are not able to assess break-offs and item nonresponse reliably. As including images to a survey slightly increases respondents' enjoyment (Couper, Tourangeau, and Kenyon 2004; Toepoel and Couper 2011), we furthermore assume that the sparing use of faces scales might increase enjoyment and reduce non response.

Nevertheless, our study has some limitations. The use of a 6 point scale is not ideal; and a middle response option seems to be more appropriate. In addition, the question wording itself was quite specific for a faces scale. Moreover, there is slight uncertainty about the ideal design of the facial shape and the utilization of the mouth line as an indicator of well-being or satisfaction, which appears to be less adequate in eastern cultures where emotional expression is primarily coded in the eye section of the face (Yuki, Maddux, and Masuda 2007).

Overall, results suggest that faces scales using the fixed and the affective design yield response distributions that differ from the responses obtained by a cognitive faces scale or a radio button question. Based on our findings, if a faces scale has to be used, the cognitive design provides the best trade-off between entertainment, attention and adequate measurement.

Note

An earlier version of this paper was presented at the AAPOR conference, Phoenix, AZ, May 2011.

REFERENCES

Chambers, C., K. Giesbrecht, K.D. Craig, S. Bennett, and E.A. Huntsman. 1999. "Comparison of Faces Scales for the Measurement of Pediatric Pain: Children's and Parents' Ratings." *Pain* 83: 25–35.

Christian, L.M., and D.A. Dillman. 2004. "The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions." *Public Opinion Quarterly* 68 (1): 57–80.

Couper, M.P. 2001. "Web Surveys: The Questionnaire Design Challenge." In *Paper Presented at the International Statistical Institute*. Seoul, Korea.

Couper, M.P., R. Tourangeau, and F.G. Conrad. 2007. "Visual Context Effects in Web Surveys." *Public Opinion Quarterly* 71 (4): 623–34.

Couper, M.P., R. Tourangeau, and K. Kenyon. 2004. "Picture This! Exploring Visual Design Effects in Web Surveys." *Public Opinion Quarterly* 68 (2): 255–66.

Elfering, A., and S. Grebner. 2010. "A Smile Is Just a Smile: But Only for Men. Sex Differences in Meaning of Faces Scales." *Journal of Happiness Studies* 11: 179–91.

Herman, J.B., R.B. Dunham, and C.L. Hulin. 1975. "Organizational Structure, Demographic Characteristics, and Employee Responses." *Organizational Behavior and Human Performance* 13: 206–32.

Jäger, R., and J. Bortz. 2001. "Rating Scales with Smilies as Symbolic Labels - Determined and Checked by Methods of Psychophysics." In *Paper Presented at the Annual Meeting of the International Society for Psychophysics*.

Kunin, T. 1998. "The Construction of a New Type of Attitude Measure." *Personnel Psychology* 51: 823–24.

Schwarz, N., and F. Strack. 1991. "Evaluating One's Life: A Judgment Mode of Subjective Well-Being." In *Subjective Well-Being*, edited by F. Strack, M. Argyle , and N. Schwarz, 1st ed., 27–48. Kronberg: Pergamon Press GmbH.

Toepoel, V., and M.P. Couper. 2011. "Can Verbal Instructions Counteract Visual Context Effects in Web Surveys?" *Public Opinion Quarterly* 75 (1): 1–18.

Wanous, J.P., A.E. Reichers, and M.J. Hudy. 1997. "Overall Job Satisfaction: How Good Are Single-Item Measures?" *Journal of Applied Psychology* 82 (2): 247–52.

Yuki, M., W.W. Maddux, and T. Masuda. 2007. "Are the Windows to the Soul the Same in the East and West? Cultural Differences in Using the Eyes and Mouth as Cues to Recognize Emotions in Japan and the United States." *Journal of Experimental Social Psychology* 43: 303–11.