

# Smileys, Stars, Hearts, Buttons, Tiles or Grids: Influence of Response Format on Substantive Response, Questionnaire Experience and Response Time

Bulletin de Méthodologie Sociologique

2019, Vol. 142 57–74

© The Author(s) 2019



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/0759106319834665

[journals.sagepub.com/home/bms](https://journals.sagepub.com/home/bms)**Vera Toepoel***Utrecht University, The Netherlands***Brenda Vermeeren***Erasmus University Rotterdam and ICTU-InternetSpiegel, The Netherlands***Baran Metin***Utrecht University, The Netherlands***Résumé****Smileys, étoiles, cœurs, boutons, grilles ou tableaux : Influence du format de la question sur la réponse, expérience pour le répondant et temps de passation.**

Les études portant sur les manières de poser les questions suggèrent que le choix (de présentation) des catégories de réponses a des effets significatifs sur les réponses des interviewés. Depuis quelques années, les images telles que les émôticones ou les étoiles sont souvent utilisées dans les études marketing en ligne. Remplacer les instruments de mesure traditionnels des enquêtes (*items* de réponse verbaux) par des catégories de réponses imagées ne produit pas des effets clairs. Dans cet article, nous expérimentons plusieurs *designs* dans le cas d'échelles de Likert, afin d'observer dans quelle mesure les résultats convergent et si l'expérience des répondants varie. Les données proviennent du *Flitspanel* (échantillon non probabiliste néerlandais). L'utilisation de cœurs ou d'étoiles aboutit à des scores de réponses comparables à ceux obtenus à partir de boutons à cliquer. Les répondants accueillent plus favorablement

---

**Corresponding Author:**

Vera Toepoel, Utrecht University, Padualaan 14, 3584 CH Utrecht, the Netherlands

Email: [v.toepoel@uu.nl](mailto:v.toepoel@uu.nl)

les designs *designs* recourant aux *smileys* tandis que les présentations sous forme de tableaux (grilles d'*items*) sont plus mal reçus. Le chercheur désirant comparer les résultats d'enquêtes devront être attentifs à ces effets et ne comparer que les réponses issues de *designs* de questions identiques.

### **Abstract**

Studies of the processes underlying question answering in surveys suggest that the choice of (layout for) response categories can have a significant effect on respondent answers. In recent years, the use of pictures, such as emojis or stars, is often used in online communication. It is unclear if pictorial answer categories can replace traditional verbal formats as measurement instruments in surveys. In this article we investigate different versions of a Likert-scale to see if they generate similar results and user experiences. Data comes from the non-probability based Flitspanel in the Netherlands. The hearts and stars designs received lower average scores compared to the other formats. Smileys produced average answer scores in line with traditional radio buttons. Respondents evaluated the smiley design most positively. Grid designs were evaluated more negatively. People wanting to compare survey outcomes should be aware of these effects and only compare results when similar response formats are used.

### **Mots clés**

Échelle de Lickert, Émoticônes, Formats des *items* de réponse, *Items* imagés, *Survey research*

### **Keywords**

Emojis, Likert-scale, Pictorial Answer Formats, Response Formats, Survey Research

### **Introduction**

Response formats in surveys are often chosen on the basis of the knowledge or intuition of the researcher. Studies about the cognitive and communicative processes underlying question answering in surveys suggest that the choice of (layout for) response categories can have a significant effect on respondent answers (see for example Couper, 2008; Toepoel and Dillman, 2011a). It is well known that differences in response options can lead to substantial differences in responses (Christian and Dillman, 2004; Tourangeau et al., 2004, 2007; Christian et al., 2007; Toepoel et al., 2009). Dillman (2007) distinguishes between verbal and nonverbal language in surveys. Verbal and nonverbal cues can independently and jointly influence the survey answers. For example, Redline et al. (2003) provide evidence that the visual and verbal complexity of information in a questionnaire affects what respondents read, the order in which they read it and, ultimately, their comprehension of the information. Dillman (2007) suggest that writing effective questions for Web surveys may depend at least as much on the presentation of the answer categories ("visual language") as on the question wording itself.

Some researchers use pictures instead of text in situations where reading ability might create barriers (Reynolds-Keefe and Johnson, 2011); for example, in cross-cultural

studies where (part of) the population can be low-literate, or in studies with respondents with intellectual disability. In recent years, the use of pictures – such as emojis or stars – is often used in online communication. It is unclear if pictorial answer categories can replace traditional verbal formats as measurement instruments in surveys. In this paper, we will investigate different versions of a Likert-scale (smileys, stars, hearts, radio buttons with and without color, grids with and without color tiles, and grid with traditional radio buttons) to see if they generate similar results and user experiences.

## **Background**

How surveys are displayed and completed in online research has changed in recent years. Nowadays, people do not only complete surveys on desktop PCs or laptops, but also on subnotebooks, tablets or smartphones (Lugtig and Toepoel, 2016). Toepoel and Lugtig (2015) argue that Web surveys should now be thought of as mixed-device surveys. This implies that survey researchers have to design Web surveys to be user-friendly (see for example Revilla et al., 2016). A user-friendly design typically uses large buttons or tiles (Arn et al., 2015), no scrolling or only down scroll (Johnson, 2015), graphics or pictures (Johnson, 2015), no grids (de Bruijne and Wijnant, 2014), and a design for varying screen sizes (see also Couper et al., 2017).

### *Formats for Ordinal Scales*

Ordinal scale questions are probably the most widely-used measurement instrument in Web surveys. These questions can be presented in various ways: answer categories fully labeled or for the endpoint categories only, with radio buttons as standalones or in a grid/matrix, with slider bars or visual analogue scales etc. Pictorial icons such as smiley faces, or emojis are user-friendly in the sense that they are commonly used in computer-mediated communication and instant messaging. In addition, mobile devices typically have small screen sizes, and pictorial icons may save space on a screen compared to text labels and radio buttons. Therefore, they may serve as a user-friendly measurement instrument for ordinal scale questions.

### *Radio Buttons and Matrix Questions*

Radio buttons are circles in which a respondent clicks to provide an answer. Radio buttons use standard HTML and work with all browsers. They are a low-tech response format. A problem with radio buttons is that they are not very efficient in use of space on a screen because they are not scalable. There are many ways to present radio buttons on a screen.

Sometimes, shades of green (for positive) and red (for negative) are added to radio buttons. Toepoel and Dillman (2011b) demonstrated that respondents are more reluctant to select negative answer options when color is added to the radio button format. This effect was only apparent in a polar-point labeled scale, not in fully-labeled scales, however. The authors argue that respondents follow simple heuristics in interpreting the visual features of a question. Options that are of similar appearance are considered

conceptually closer than when they are dissimilar in appearance. However, verbal labels seem to overrule visual cues.

Another heuristic that respondents use in answering questions is the “near means related heuristic” (see Tourangeau et al., 2004, 2007). This heuristic implies that questions that are grouped together, as in grid or matrix questions, will be seen as conceptually related. Presenting questions in a grid or matrix is a way to save space on a screen, preserve context, and reduce the number of clicks/taps. Research shows that items are more likely to be seen as related if grouped on one screen, reflecting a natural assumption that blocks of questions bear on related issues, much as they would during ordinary conversations (Schwarz and Sudman, 1996; Sudman et al., 1996). Couper et al. (2001) concluded that correlations are consistently higher among items appearing together on a screen than for items presented across several screens. Grouping questions in a matrix can hence affect responses. Revilla et al. (2017), in their comparison on PC and smartphone surveys, suggest using an item-by-item format to improve comparability between devices.

### *Pictorial Scales*

Pictorial scales are commonly used in surveying children (de Leeuw, 2001; Hall et al., 2016), to assess levels of pain (Toepoel and Funke, 2018), experiences (Yang, 2002), job satisfaction (Kunin, 1955), or to replace text response options in surveys (Elfering and Grebner, 2010). In smiley face scales, respondents match their emotions or attitudes on a scale showing faces with only the curvature of the mouth line varying systematically from a large smile to a grimace. The smiley face has proven to be related to the recognition of the happy versus sad emotion scale (Ekman, 1999). Stange et al. (2016) demonstrate that smiley faces can be used in Web surveys in addition to text labels. The faces speed up processing of questions, especially for low-literate respondents. From their study however, it remains unclear whether faster processing means the question was cognitively easier to process or respondents took shortcuts. Smiley face scales help low-literate people in answering survey questions without having to read and understand verbal text. In addition, respondents can experience the question-answering process as more enjoyable (Emde and Fuchs, 2012).

Thomas and Barlas (2017) compared smileys and thumbs up to text and found no differences in task duration and mean scores. They found more categories meant longer completion times for text, but not for smiley or thumbs up. They do, however, warn that emojis are not suitable for all type of questions. For example, using a smiley for an item such as “A person who plans a murder and carries it out should be put to death” does not seem to be appropriate for these types of pictorial answer formats. They also suggest not using more than five categories since it can be difficult to portray meaningful gradations in emojis with more categories. Stange et al. (2016) report results of two eye-tracking experiments in which satisfaction questions were asked with and without smiley faces. Respondents to the questions with smileys spent less time reading the question stem and response option text than respondents to the questions without smileys. The response distributions did not vary per version. Stange et al. find

support that lower literacy respondents rely more on the smiley faces than their counterparts. In addition, Reynolds-Keefer and Johnson (2011) noted that students refrained from using the pictures representing more moderate responses when the questionnaire options were less realistic and more exaggerated pictures. Exaggeration of emotions seems to polarize responses.

Cultural differences in the perception of facial emotion can affect the cross-cultural applicability of a scale. Masuda et al. (2008) investigated cartoons depicting a happy, sad, angry, or neutral person surrounded by other people expressing the same emotion as the central person or a different one. The surrounding people's emotions influenced Japanese but not Westerners' perceptions of the central person, indicating that Japanese respondents pay more attention to social context. Kilbride and Yarczower (1980) compared US and Zambian students in the imitation of facial expressions and found that Zambian students were less accurate in detecting the facial expressions. This could be due to cultural differences in recognition of facial expressions.

In an American study, Reynolds-Keefer et al. (2009) did not find variability in responses when varying a picture (sad/happy face, clouds/sun). SurveyMonkey (2017), one of the larger survey software providers, asked 12 questions in their panel using a satisfaction scale with five response options on radio buttons. In addition to these, they asked the same questions using stars, smiley faces, hearts and thumbs. They published results on their Web site showing that all formats produced similar responses.

In this paper, we analyze the variability of answers over different types of verbal (radio buttons) and pictorial Likert scales in terms of substantive response, "don't know" options, questionnaire experience and response time.

## **Methodology**

### ***Sample and Response***

For the so-called Flitspanel survey, conducted in June 2017, a Web panel was used. The Flitspanel was established by the Dutch Ministry of Interior and Kingdom Relations with the aim of enabling quick and effective information collection. This panel consists of more than 20,000 Dutch public sector employees. In the past, these employees have signed on for the panel themselves. Approximately six times a year, they receive a short questionnaire. For our survey, all 21,059 public sector employees that participated in one or more studies in the past year (the so-called active panel) were invited by email to take part. Subsequently, 20 sub-samples were randomly drawn from this total group, all of which were presented with a different version of the questionnaire, varying in design, length and direction of the answer options. The sub-samples were drawn in such a way that they are comparable in terms of age, gender and educational level. After three weeks, 7,096 employees replied (a response rate of 34%). Halfway through the fieldwork period, a reminder email was sent. For this paper, we use eight subsamples that only varied in design. The response rates of these eight designs ranged between 31.7% (neutral tiles grid) and 35.1% (coloured radio buttons).

## Measures

In this article, eight different designs were studied: smileys, stars, hearts, coloured radio buttons, neutral radio buttons (no color), coloured grid with tiles, neutral grid with tiles, and a radio button grid. See Appendix 1 for the screenshots (in Dutch). In tiled designs, the whole rectangular tile of the answer format was clickable. For non-grid designs, we used a scrollable design with an auto-forward function. For the grid designs, the “don’t know” option was placed at the right hand-side of the substantive answer options as is commonly the case with grid questions. For the non-grid designs, the “don’t know” option was placed below the substantive answer options.

The designs were compared on four different aspects namely 1) substantive response, 2) “don’t know” options, 3) questionnaire experience, and 4) response time.

**Substantive Response.** Respondents were asked to answer several work-related questions regarding their satisfaction, engagement, commitment, role clarity, autonomy, alignment and employability. There were different answer categories, namely 1) totally disagree to totally agree, and 2) very dissatisfied to very satisfied. For reasons of comparability, we selected the 14 questions regarding employee satisfaction (See Appendix 2 for the selected questions), and we used the average score across these 14 questions in this study.

**“Don’t know” Option.** For the 14 questions discussed above, we calculated a binary variable indicating whether or not the response option “don’t know” was selected in any of the 14 questions.

**Questionnaire Experience.** Respondents were asked to answer five questions regarding their questionnaire experience, namely 1) “the questionnaire was nice to fill in”, 2) “the questions were clear”, 3) “I like the completion time of the questionnaire”, 4) “I find it easy to fill in the questionnaire”, and 5) “I like the layout and appearance of the questionnaire”. The respondents could answer on a five-point Likert scale ranging from 1) totally disagree to 5) totally agree.

**Response Time.** The response time is a continuous variable ranging from 0.13 to 1,435.75 seconds. The time was calculated from opening to finishing the questionnaire. We truncated everyone with a completion time above two standard deviations to the value of 2 standard deviations. This applied to 194 respondents.

**Control Variables.** We controlled for three personal characteristics. We coded gender as a dummy variable (1 = female). Age was a continuous variable subdivided into three classes. Young is 15-34, Middle is 35-54 and Old is 55 years and older. Educational level was subdivided into three classes, namely low (primary education and low vocational education), middle (higher general secondary education, preparatory academic education, vocational education), and high (higher vocational education, candidate exam; scientific education, PhD).

**Table 1.** Mean scores on 14 substantive questions for the eight different designs

	Satisfaction	N
Smileys	3.53	317
Stars	3.44	316
Hearts	3.33	331
Coloured radio buttons	3.51	334
Neutral radio buttons	3.44	322
Coloured tiles grid	3.64	316
Neutral tiles grid	3.60	308
Radio button grid	3.55	313
F value	8.30	
P	.00	

**Table 2.** Tukey HSD test for 14 substantive questions regarding employee satisfaction

Designs	N	Subset for alpha = 0.05		
		1	2	3
Tukey HSD				
Hearts	331	3.33		
Stars	316	3.44	3.44	
Coloured radio buttons	322	3.44	3.44	
Neutral radio buttons	334		3.51	3.51
Smileys	317		3.53	3.53
Radio button grid	313		3.55	3.55
Neutral tiles grid	308			3.60
Coloured tiles grid	316			3.64

## Results

### Descriptives

In our study sample, 60% of the respondents were male. The average age was 53.8 years old and the predominant educational level was higher education. Age, gender and education did not significantly differ per design condition.

### Substantive Response

The results in Table 1 show that there are statistically significant differences in the answers of the respondents according to the designs. The average satisfaction with the job and the organization was highest for the (coloured and neutral) rectangles grids.

Post-hoc comparisons using the Tukey HSD test in Table 2 indicated three homogeneous subsets, where hearts have the lowest outcome scores and were substantially different from radio buttons, grids and smileys.

**Table 3.** Use of “don’t know” option

Design	N	Don’t know
Smileys	359	42 (11,7%)
Stars	356	40 (11,2%)
Hearts	360	29 (8,1%)
Coloured radio buttons	370	36 (9,7%)
Neutral radio buttons	353	31 (8,8%)
Coloured tiles grid	341	8 (2,3%)
Neutral tiles grid	334	10 (3,0%)
Radio button grid	353	14 (4,0%)
<i>Pearson chi square</i>	49.52	p=.00

**Table 4.** Mean scores on questionnaire experience for the eight different designs

	“The questionnaire was nice to fill in”	“The questions were clear”	“I like the completion time of the questionnaire”	“I find it easy to fill in the questionnaire”	“I like the layout and appearance of the questionnaire”	“Average survey experience”
Smileys	4.14	4.37	4.46	4.53	4.55	4.42
Stars	4.05	4.28	4.37	4.42	4.39	4.31
Hearts	4.04	4.21	4.32	4.40	4.30	4.26
Coloured radio buttons	4.08	4.29	4.35	4.41	4.41	4.31
Neutral radio buttons	4.02	4.32	4.38	4.48	4.38	4.32
Coloured tilesgrid	3.92	4.23	4.30	4.35	4.14	4.19
Neutral tiles grid	3.94	4.23	4.32	4.37	4.12	4.21
Radio button grid	3.79	4.17	4.24	4.27	3.75	4.04
F value	7.14	3.82	4.22	5.98	40.29	16.26
P	.00	.00	.00	.00	.00	.00

### “Don’t know” Option

Table 3 shows the number of people that selected at least one “don’t know” option in the 14 questions. In the grid designs, fewer people selected the “don’t know” option compared to the other designs. For the grid questions, the “don’t know” option was presented at the end of the list of substantive answer options, and not visually separate (see Appendix 1). In the other designs, the “don’t know” option was a separate button placed below the substantive answer options. The results in Table 3 suggest that higher visibility results in more respondents selecting the “don’t know” option.

**Table 5.** Tukey HSD test for “I like the layout and appearance of the questionnaire”

Designs	N	Subset for alpha = 0.05				
		1	2	3	4	5
Radio button grid	342	3.75				
Neutral tiles grid	316		4.12			
Coloured tiles grid	321		4.14	4.14		
Hearts	356			4.3	4.3	
Neutral radio buttons	347				4.38	
Stars	350				4.39	4.39
Coloured radio buttons	366				4.41	4.41
Smileys	356					4.55

**Table 6.** Regression analysis questionnaire experience (standardized coefficients)

	Questionnaire Experience Beta
Gender (1=female)	-.03
Age_young	.01
Age_old	.03
Educational level_low	-.02
Educational level_high	-.03
Smileys	.24*
Stars	.17*
Hearts	.14*
Colored radio buttons	.17*
Neutral radio buttons	.18*
Colored tiles grid	.09*
Neutral tiles grid	.10*
Radio button grid (ref)	
R square	
F value	10.21
P	.00

\* =  $p < .001$ 

### Questionnaire Experience

The results in Table 4 show that there are statistically significant differences in the questionnaire experience of the respondents across the designs. On average, the respondents have the best experience with the smiley design and less positive experiences with the radio button grid.

Post-hoc comparisons using the Tukey HSD test in Table 5 indicate five subsets showing the distinction between radio, grid, pictorial/colour and smiley in the evaluation question about the layout of the survey. Appendix 3 shows post-hoc comparisons for the other evaluation questions.

**Table 7.** Mean scores on response time for the eight different designs

	Response time in minutes	N
Smileys	22.67	359
Stars	30.28	356
Hearts	23.97	360
Coloured radio buttons	26.08	370
Neutral radio buttons	22.10	353
Coloured tiles grid	16.99	341
Neutral tiles grid	16.99	334
Radio button grid	20.10	353
F value	1.57	
Significance	.14	

Note: variances are high between the designs. When we truncate maximum duration more strict than mentioned in Section 3.2, results remain insignificant.

Table 6 shows the results of a linear regression analyses on the average score of the five evaluation questions with demographical and design variables. The regression analysis demonstrates that demographics do not have an effect on the average evaluation score. However, the design variables all have an effect, with respondents being less positive when they answered in a grid design. Respondents were most satisfied with the smiley layout.

### *Response Time*

The results in Table 7 show that there are no statistically significant differences in response time between the different designs.

## **Conclusion and Discussion**

In this paper, we have evaluated eight different designs in terms of average answer scores, selection of “don’t know” options, respondents’ evaluation and duration of the survey. We used grid questions (with and without tiles), radio buttons, pictorial answer formats, and shades of color. Data come from the non-probability based Flitspanel in the Netherlands. Panel members were used to all kinds of answer formats.

We found that grids have higher average scores than designs without grids. This could be explained by the fact that in the grid questions, the “don’t know” option was placed at the right hand-side of the substantive answer options, adding a sixth response option to the scale. Respondents may misinterpret the conceptual midpoint of the scale with the visual midpoint of the scale, resulting in higher answer scores. This is in line with visual heuristics, in particular the “Middle Means Typical” heuristic (see Tourangeau et al., 2004; Toepoel and Dillman, 2011b). Fewer people selected the “don’t know” option in the grid designs compared to the non-grid designs where the “don’t know” option was placed below the substantive answer options and hence was more visually pronounced.

In addition, the hearts and stars designs received lower average scores compared to the other formats. The star and heart designs both seem to map better to unipolar

response scales. Whereas the sad face shows negative ratings and the smiling face shows positive ratings, there is no way to show negative stars or negative hearts. This is a limitation in the use of stars and hearts. Future research could focus on how to use pictorial scales for unipolar and bipolar scales.

Smileys, another form of pictorial answer formats, produced average answer scores in line with traditional radio buttons. The smiley face scale incorporates colour in the design, with negative ratings in orange/red and positive ratings in green. This colour scheme adds another layer to the condition and might affect the findings *versus* a smiley face scale in black and white. Future research could compare colour and black-and-white pictorial answer scales to investigate the effect of colour on pictorial rating scales.

Respondents evaluated the smiley design most positively. Grid designs were evaluated the worst, with the radio button grid design evaluated even worse than the tile grid designs. Unfortunately, the data collection agency was unable to provide user agent strings; that is to say they were only collected for about 10% of the sample. It would have been interesting to investigate if, for example, tiles and pictorial designs would perform better on small smartphone screens, since they would be more user-friendly.

In the non-grid designs, we used an auto-forward function. This did not have an effect on response times, since we did not find differences in response times over formats. It could produce counteractive effects, for example by speeding up usability but slowing down cognitive processing of these designs. Future research should try to disentangle cognitive from usability effects, for example by using eye-tracking.

Since smileys are evaluated better and perform similar to traditional radio buttons, there seems to be an advantage in using smileys as a response format. More research is needed to determine the effect of the use of other emojis, such as thumps up or other pictures.

This paper demonstrates that different designs can produce different survey outcomes. Researchers wanting to compare survey outcomes, for example in benchmark studies, should pay particular attention to these design effects and only compare results when similar response formats are used. Otherwise, it is impossible to differentiate the design effect from the substantive effect. Also, cultural differences can play a role that is not however specific to pictorial scales and goes for verbal scales as well.

With more and more people accessing online surveys via smartphones, finding ways to use screen size effectively and making surveys mobile-friendly becomes one of the key tasks of survey methodologists. This paper brings us a small step forward in choosing the optimal design format for survey questions. However, there is still a lot to be learned concerning the effect of answer format design on cognitive and usability processing.

## **Acknowledgments**

We would like to thank the editors of BMS and anonymous reviewers for helpful comments on this manuscript.

## **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## References

- Arn B, Klug S and Kolodziejski J (2015) Evaluation of an Adapted Design in a Multi-Device Online Panel. *Methods, Data, Analysis* 9(2): 185-2012.
- Christian LM and Dillman DA (2004) The Influence of Graphical and Symbolic Language Manipulations to Self-Administered Questions. *Public Opinion Quarterly* 68(1): 57-80.
- Christian LM, Dillman DA and Smyth JD (2007) Helping Respondents Get it Right the First Time: The Relative Influence of Words, Symbols, and Graphics in Web and Telephone Surveys. *Public Opinion Quarterly* 71(1): 113-125.
- Couper MP (2008) *Designing Effective Web Surveys*. Cambridge: Cambridge University Press.
- Couper MP, Antoun C and Mavletova M (2017) Mobile Web Surveys. In: Biemer et al. *Total Survey Error in Practice*. New York: Wiley, 133-154.
- Couper MP, Traugott MW and Lamias MJ (2001) Web Survey Design and Administration. *Public Opinion Quarterly* 65(2): 230-253.
- De Bruijne M and Wijnant A (2014) Can Mobile Web Surveys Be Taken on Computers? A Discussion on a Multi-Device Survey Design. *Survey Practice* 6(4): 1-8.
- De Leeuw E (2001) Improving Data quality when Surveying Children and Adolescents: Cognitive and Social Development and its Role in Questionnaire Construction and Pretesting. Report prepared for the Annual Meeting of the Academy of Finland. Available at: [http://www.aka.fi/globalassets/awanhat/documents/tiedostot/lapset/presentations-of-the-annual-seminar-10-12-may-2011/surveying-children-and-adolescents\\_de-leeuw.pdf](http://www.aka.fi/globalassets/awanhat/documents/tiedostot/lapset/presentations-of-the-annual-seminar-10-12-may-2011/surveying-children-and-adolescents_de-leeuw.pdf)
- Dillman DA (2007) *Mail and Internet Surveys. The Tailored Design Method*. New York: Wiley.
- Ekman P (1999) Facial Expressions. In: Dalglish T and Power M (eds) *Handbook of Cognition and Emotion*. New York: Wiley, 301-320.
- Elfering A and Grebner S (2010) A Smile is Just a Smile: But Only for Men. Sex Differences in Meaning of Faces Scales. *Journal of Happiness Studies* 11(2): 179-191.
- Emde M and Fuchs M (2012) Exploring Animated Faces Scales in Web Surveys. Drawbacks and Prospects. *Survey Practice* (1): 1-6.
- Hall L, Hume C and Tazzyman S (2016) Five Degrees of Happiness: Effective Smiley Face Likert Scales for Evaluating with Children. DOI: 10.1145/2930674.2930719
- Johnson EP (2015) *Mobile Device and Modular Survey Design*. Available at: <http://papor.ipower.com/wp-content/uploads/2016/01/2015-PAPOR-Short-Course-1.pdf>
- Kilbride JE and Yarczower M (1980) Recognition and Imitation of Facial Expressions. A Cross-Cultural Comparison between Zambia and the United States. *Journal of Cross-Cultural Psychology* 11(3): 281-296.
- Kunin T (1955) The Construction of a New Type of Attitude Measure. *Personnel Psychology* 8(1): 65-78.
- Lugtig P and Toepoel V (2016) The Use of PCs, Smartphones, and Tablets in a Probability-Based Panel Survey - Effects on Survey Measurement Error. *Social Science Computer Review*, 34(1): 78-94.

- Masuda T, Ellsworth P, Mesquita B, Leu J, Tanida S and Van de Veerdonk E (2008) Placing the Face in Context: Cultural Differences in the Perception of Facial Emotion. *Journal of Personality and Social Psychology*, 94(3): 365-381.
- Redline CD, Dillman DA, Carley-Baxter L and Creecy R (2003) *Factors that Influence Reading and Comprehension in Self-Administered Questionnaires*. Paper presented at the Workshop on Item-Nonresponse and Data Quality, Basel Switzerland, October 10, 2003. Available at: <http://survey.sesrc.wsu.edu/dillman/papers.htm>
- Revilla M, Toninelli D and Ochoa C (2017) An Experiment Comparing Grids and Item-by-Item Formats in Web Surveys Completed Through PCs and Smartphones. *Telematics and Informatics* 34(1): 30-42.
- Revilla M, Toninelli D, Ochoa C and Loewe G (2016) Do Online Access Panels Need to Adapt Surveys for Mobile Devices? *Internet Research* 26(5): 1209-1227.
- Reynolds-Keefer L and Johnson R (2011) Is a Picture Worth a Thousand Words? Creating Effective Questionnaires with Pictures. *Practical Assessment, Research & Evaluation* 16(8): 1-7.
- Reynolds-Keefer L, Johnson R, Dickenson T and McFadden L (2009) Validity Issues in the Use of Pictorial Likert Scales. *Studies in Learning, Evaluation, Innovation and Development* 6 (3): 15-24.
- Schwarz N and Sudman S (1996) *Answering Questions*. San Francisco: Jossey-Bass.
- Stange M, Barry A, Smyth J and Olson K (2016) Effects of Smiley Face Scales on Visual Processing of Satisfaction Questions in Web Surveys. *Social Science Computer Review* 36(6): 756-766.
- Sudman S, Bradburn NM and Schwarz N (1996) *Thinking about Answers*. San Francisco: Jossey-Bass.
- SurveyMonkey (2017) Will the New Star Rating Question Type Affect your Survey Completion Rates? Available at: <https://www.surveymonkey.com/curiosity/will-new-star-rating-question-type-affect-survey-completion-rates/>
- Thomas RK and Barlas F (2017) *Thumps Up or Down? Emojis as a Response Format for Online Surveys*. Paper presented at the ARF Annual Conference 2017.
- Toepoel V, Das M and Van Soest A (2009) Design of Web Questionnaires – The Effect of Layout in Rating Scales. *Journal of Official Statistics* 25(4): 509-528.
- Toepoel V and Dillman DA (2011a) How Visual Design Affects the Interpretability of Survey Questions. *Social and Behavioral Research and the Internet. Advances in Applied Methods and Research Strategies*. Abingdon: Routledge, 165-190.
- Toepoel V and Dillman DA (2011b) Words, Numbers, and Visual Heuristics in Web Surveys - Is there a Hierarchy of Importance? *Social Science Computer Review* 29(2): 193-207.
- Toepoel V and Funke F (2018) Sliders, Visual Analogue Scales, or Buttons: Influence of Formats and Scales in Mobile and Desktop Surveys. *Mathematical Population Studies* 25(2): 112-122.
- Toepoel V and Lugtig P (2015) Online Surveys are Mixed-Device Surveys. *Methods, Data, Analysis* 9(2): 155-162.
- Tourangeau R, Couper MP and Conrad F (2004) Spacing, Position, and Order. Interpretive Heuristics for Visual Features of Survey Questions. *Public Opinion Quarterly* 68(3): 368-393.
- Tourangeau R, Couper MP and Conrad F (2007) Color, Labels, and Interpretive Heuristics for Response Scales. *Public Opinion Quarterly* 71(1): 91-112.
- Yang H (2002) Establishing the Reliability of the Smiley Face Assessment Scale: Test-Retest. *The Cyber Journal of Applied Leisure and Recreational Research*. Available at: <http://larnet.org/2002-10.html>

## Appendix I. Screenshots Designs

### Smileys



### Stars



### Hearts



### Coloured Radio Buttons



### Neutral Radio Buttons



### Coloured Tiles Grid

#### Mijn werkplezier

Hieronder worden u aspecten voorgelegd over uw werk en de organisatie waar u werkt. U kunt voor elk aspect aangeven in hoeverre u er tevreden of ontevreden over bent.

Mijn baan, alles bijeengenomen	Zeer ontevreden	<b>Ontevreden</b>	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
Mijn organisatie, alles bijeengenomen	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De inhoud van het werk	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De samenwerking met collega's	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De mate van zelfstandigheid	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De hoeveelheid werk	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De resultaatgerichtheid van mijn organisatie	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring

### Neutral Tiles Grid

#### Mijn werkplezier

Hieronder worden u aspecten voorgelegd over uw werk en de organisatie waar u werkt. U kunt voor elk aspect aangeven in hoeverre u er tevreden of ontevreden over bent.

Mijn baan, alles bijeengenomen	Zeer ontevreden	<b>Ontevreden</b>	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
Mijn organisatie, alles bijeengenomen	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De inhoud van het werk	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De samenwerking met collega's	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De mate van zelfstandigheid	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De hoeveelheid werk	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring
De resultaatgerichtheid van mijn organisatie	Zeer ontevreden	Ontevreden	Niet tevreden/ niet ontevreden	Tevreden	Zeer tevreden	Weet niet/ Geen ervaring

## Radio Button Grid

<b>Medewerkersonderzoek</b>						
Leuk dat u meedoet aan dit onderzoek! Met uw input krijgen wij waardevolle informatie over onderwerpen binnen de (semi-)publieke sector. Succes met het invullen van de vragen.						
Let op: dit is een demonstratie, de gegevens zullen niet opgeslagen worden!						
<b>Mijn werkplezier</b>						
Hieronder worden u aspecten voorgelegd over uw werk en de organisatie waar u werkt. U kunt voor elk aspect aangeven in hoeverre u er tevreden of ontevreden over bent.						
<b>In hoeverre ben je het eens met de volgende stelling(en)?</b>	Ze er on tev re den	On te v re de n	N ie t t e v r e d e n/ n ie t o n t e v r e d e n	T e v r e d e n	Z e e r t e v r e d e n	W e e t n i e t/ G e e n e r v a r i n g
• Mijn baan, alles bijeengenomen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• Mijn organisatie, alles bijeengenomen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• De inhoud van het werk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• De samenwerking met collega's	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• De mate van zelfstandigheid	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• De hoeveelheid werk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• De resultaatgerichtheid van mijn organisatie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• De wijze waarop mijn leidinggevende leiding geeft	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• De informatievoorziening binnen mijn organisatie	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
• De loopbaanontwikkelingsmogelijkheden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## Appendix 2. Questions

1. All things considered, how satisfied are you with your job?
2. All things considered, how satisfied are you with your organization?
3. How satisfied are you with the content of your work?
4. How satisfied are you with the collaboration with colleagues?
5. How satisfied are you with the degree of autonomy?
6. How satisfied are you with the amount of work?
7. How satisfied are you with the result orientation of your organization?
8. How satisfied are you with the leadership behavior of your line manager?
9. How satisfied are you with the information provision within your organization?
10. How satisfied are you with your career development opportunities?
11. How satisfied are you with your rewards?
12. How satisfied are you with the way in which you are appraised?
13. How satisfied are you with the degree of influence that you have within your organization?
14. How satisfied are you with the attention of the organization for your personal well-being?

### Appendix 3. Tukey HSD Tests for Questionnaire Experience

#### Tukey HSD Test. “The questionnaire was nice to fill in”

Designs		N	Subset for alpha = 0.05		
			1	2	3
Tukey HSD	Radio button grid	352	3.79		
	Coloured tiles grid	324	3.92	3.92	
	Neutral tiles grid	323	3.94	3.94	
	Neutral radio buttons	351		4.02	4.02
	Hearts	359		4.04	4.04
	Stars	352		4.05	4.05
	Coloured radio buttons	367		4.08	4.08
	Smileys	357			4.14

#### Tukey HSD Test. “The questions were clear”

Designs		N	Subset for alpha = 0.05		
			1	2	3
Tukey HSD	Radio button grid	349	4.17		
	Hearts	359	4.21	4.21	
	Coloured tiles grid	324	4.23	4.23	4.23
	Neutral tiles grid	323	4.23	4.23	4.23
	Stars	353	4.28	4.28	4.28
	Coloured radio buttons	369	4.29	4.29	4.29
	Neutral radio buttons	351		4.32	4.32
	Smileys	358			4.37

#### Tukey HSD Test. “I like the completion time of the questionnaire”

Designs		N	Subset for alpha = 0.05		
			1	2	3
Tukey HSD	Radio button grid	350	4.24		
	Coloured tiles grid	324	4.30	4.30	
	Neutral tiles grid	322	4.32	4.32	
	Hearts	360	4.32	4.32	
	Coloured radio buttons	367	4.35	4.35	4.35
	Stars	354	4.37	4.37	4.37
	Neutral radio buttons	349		4.38	4.38
	Smileys	357			4.46

**Tukey HSD Test.** "I find it easy to fill in the questionnaire"

Designs	N	Subset for alpha = 0.05		
		1	2	3
Tukey HSD				
Radio button grid	352	4.27		
Coloured tiles grid	324	4.35	4.35	
Neutral tiles grid	323	4.37	4.37	
Hearts	359	4.40	4.40	4.40
Coloured radio buttons	369		4.41	4.41
Stars	353		4.42	4.42
Neutral radio buttons	349		4.48	4.48
Smileys	356			4.53