Article

Social Science Computer Review
2016, Vol. 34(2) 244-254

DOI: 10.1177/0894439315575477
ssc.sagepub.com

A Web Experiment Showing Negative Effects of Slider Scales Compared to Visual Analogue Scales and Radio Button Scales

Frederik Funke[1]

Abstract

This article provides evidence that there is a substantial difference between slider scales and visual analogue scales (VAS), two types of rating scales used in web surveys that are frequently mixed up. In an experimental design, both scales were compared to standard HTML radio buttons and offered three, five, or seven response options. Slider scales negatively affect response rate (especially on mobile devices), the sample composition, the distribution of values, and also increase response times. VAS and radio buttons, however, can be used without negative side effects, even on touch screen devices like smartphones. Overall, it is recommended to avoid slider scales. As small differences in rating scales—here drag and drop versus point and click—have a huge influence on data collection, an optimal implementation of VAS is suggested. However, measurement of discrete variables with a moderate number of response options should be done with radio buttons scales unless a small screen size—for example, on smartphones—requires an economical use of space.

Keywords

survey methodology, visual analogue scale, VAS, slider scale, online research, rating scales

Online questionnaires offer numerous possibilities for data collection not available or feasible in paper-and-pencil studies. One example is visual analogue scales (VAS, see Figure 1, middle), which are frequently used in the medical sector where slight changes are of great importance (e.g., for measurement of pain) but only rarely in (large-scale) surveys. Another example is slider scales (see Figure 1, bottom)—rating scales very similar in appearance to VAS—that are frequently used in social research and especially market research. However, in web surveys, data are mostly gathered with rating scales made from HTML radio buttons (see Figure 1, top). This reluctance to use other scales may origin in the considerable additional effort of nonstandard rating scales like VAS in paper-and-pencil studies. Although in computerized research readout is done automatically, the number of papers examining the use of VAS or slider scales for survey research is very small.

[1] datagladiator.net, Staufenberg, Germany

Corresponding Author:
Frederik Funke, datagladiator.net, Kragenhofer Str. 20, 34355 Staufenberg, Germany.
Email: ff@datagladiator.net

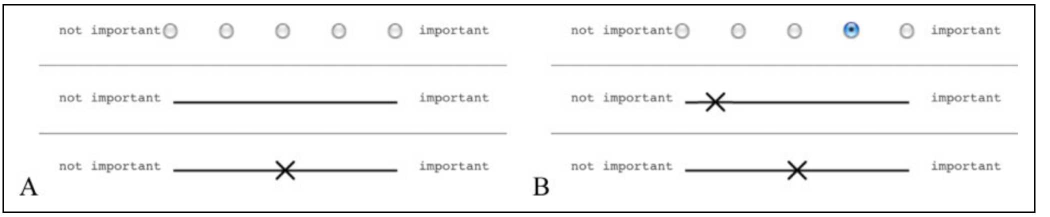

Figure 1. Radio button scale (top), VAS (middle), and slider scale (bottom) right after loading of a web page (A) and with rating (B). For working examples of all rating scales used in this study, see http://vasgenerator.net/Funke_2015_slider_vs_vas/ (VAS = visual analogue scales.)

VAS

VAS—first described by Hayes and Paterson (1921)—consist of a plain, mostly horizontal line and mostly verbal anchors on each end but also nonverbal stimuli (e.g., smileys) are used sometimes. Respondents give a rating (e.g., the extent of approval to a statement) by placing a mark on that line. In interviewer-administered or paper-based studies, VAS often have 100 mm in length so that the position of the mark measured from one side of the scale corresponds to a VAS score between 0 and 100.

VAS are greatly acknowledged in the medical sector and there are literally thousands of studies using this rating scale. Methodological studies on VAS provided evidence that the scale is linear (Hofmans & Theuns, 2008; Myles, Troedel, Boquest, & Reeves, 1999; Myles & Urquhart, 2005), that measurement with continuous scales is sensitive (Abend, Dan, Maoz, Raz, & Bar-Haim, 2014; Rausch & Zehetleitner, 2014), and that there are only few, mostly small differences in mean ratings from measurement with discrete, categorical rating scales (Averbuch & Katzper, 2004; Cork et al., 2004; Funke & Reips, 2012b; Flynn, van Schaik, & van Wersh, 2004).

VAS in web surveys. Only a few studies have been conducted on VAS in web surveys (see Couper, Traugott, & Lamias, 2001; Reips, 2002). Couper, Tourangeau, Conrad, and Singer (2006) found higher rates of missing data and higher response times. Gerich (2007) found lower mode effects—paper-and-pencil versus computer-assisted self-interviews—for VAS in comparison to 5-point scales. Reips and Funke (2008) conclude that data collected with VAS are on the level of an interval scale. Furthermore, measurement should be robust for different scale lengths and VAS are very efficient in use of space—a VAS with about 50 response options takes about the same space as a 3-point rating scale made from radio buttons and still has shown good measurement properties (Reips & Funke, 2008). Furthermore, VAS' fine gradations are especially valuable in semantic differentials or polarity profiles, where VAS can help to avoid the problem of shared ranks (see Funke & Reips, 2012b). Studer (2012) found that VAS are feasible in web surveys and suited for measurement of happiness without scaling effect. VAS can either be realized as continuous scales (each pixel in length serves as a possible rating) or be realized as discrete n-point scales.

The VAS used in this study are operated by point and click, in the very same way radio buttons are used (see Figure, middle 1 A vs. B). Respondents have to (1) move the mouse pointer to the response option and (2) click the mouse button. VAS require a client-side technology (e.g., JavaScript, Java, or Flash) running on the respondent's web browser. A low-tech rating scale made from standard HTML should be used as technological fallback.

Slider Scales

Quite similar to VAS are slider scales also made from a line but with a handle that has to be moved to give a rating. Although the idea is not new—McReynolds and Ludwig (1987) report that a precursor of slider scale has been used as early as the beginning of the 19th century—the scale is mainly used in computerized data collection. In HTML5, there is also a slider as new input type. In general, slider scales suffer from the problem of the handle's starting position. If it is initially placed at a valid rating, item nonresponse cannot be identified. An overestimation of the response category because of satisficing (Krosnick, 1991) is likely. But even if the handle is placed outside the valid range, anchoring effects (Sudman, Bradburn, & Schwarz, 1996) are likely. Funke, Reips, and Thomas (2011) found that slider scales lead to more dropout—especially for respondents with a low formal education—and that the initial position of the slider leads to a different distribution of values in comparison to radio button scales.

Despite the greater similarity in appearance to VAS, in many implementations sliders differ in the actual usage. Sliders are operated by drag and drop, which requires not two but four actions (see Figure 1, bottom A vs. B): Respondents have to (1) move the mouse pointer to the slider handle, (2) click and hold the mouse button, (3) move the handle to the response option, and (4) release the mouse button. Slider scales can be realized with HTML5 or using client-side technology like JavaScript.

Radio Button Scales

Standard closed-ended, categorical, discrete rating scales are mostly made from HTML radio buttons. The advantage is that radio buttons can be used with every web browser regardless of the technology available on the respondent's computer. Respondents immediately see which response options are available and can choose between them. Ratings (see Figure 1, top A vs. B) are given by point and click as described for VAS.

Research Questions

This experimental research focuses on the question if there are considerable differences between data collected with VAS, slider scales, and radio button scales. In recent papers , it often remains unclear if VAS or sliders have been used as rating scales. It seems that both terms are used interchangeable to a certain degree, which leads to an uncertainty about the effects of these rating scales in web surveys. As (inevitably discrete) rating scales made from radio buttons are likely to be the most frequently used response options in web surveys and in order to allow a proper comparison, this study is limited to discrete VAS and sliders (even though continuous implementations are possible; see Funke & Reips, 2012).

Method

The experiment was embedded in the evaluation of the questionnaire on study satisfaction consisting of 3 items ("How interesting was this survey?", "How important was the study‘s topic to you?", and "Do you think that your participation will affect the future situation?"). Each item was presented on a separate web page. Respondents were students at 18 different universities of applied sciences in Germany. Multiple participation was not possible.

In a fully crossed 3 $\times$ 3 factorial between-subjects design respondents were randomly assigned to different rating scales to answer the 3 items mentioned earlier. Factor 1 was the type of rating scale, that is, slider, VAS, or rating scale made from radio buttons. Factor 2 was the number of response options, that is, 3-point, 5-point, or 7-point scales. VAS were built using the free web service *VAS Generator* (see Reips & Funke, 2008); working examples of all rating scales can be found online at http://vasgenerator.net/Funke_2015_slider_vs_vas/. The differences and similarities between the three rating scales compared in this study are summarized in Table 1.

The following characteristics of the response scales were kept as similar as possible: (1) only the scales' ends had verbal labels as anchors, (2) the scales' length (i.e., horizontal line with sliders and VAS and distance between most right and most left button) was almost identical, about 200 pixels, (3) the

Table 1. Comparison of Response Scale Characteristics Between Slider Scales, VAS, and Radio Button Scales.

Dimension	Slider Scale	VAS	Radio Button Scale
Handling	1. Move pointer 2. Click and hold mouse 3. Move handle 4. Release button	1. Move pointer 2. Click	1. Move pointer 2. Click
Scale appearance on load	Marker visible	No marker (appears after first click)	No marker (appears after first click)
Technological requirements	HMTL + client-side technology (here JavaScript)	HMTL + client-side technology (here JavaScript)	HTML only

Note. VAS = visual analogue scales. There is a slider scale available with HTML5 browsers (form type *range*) that does not require client-side technologies. Although recent browsers already support HTML5, it is still a draft and HTML5 sliders still have certain restrictions (e.g., starting position not possible outside the scale) at the present.

marker (a plain cross) was the same for both graphical rating scales, (4) respondents could change ratings as often as wanted, and (5) all items were nonmandatory. As all rating scales were discrete, the marker indicating the rating with slider scales and VAS could only be placed at specified positions. If respondents tried to choose another position, the marker leaped to the closest response option.

However, there were also some inevitable differences: JavaScript had to be activated on respondents' computer to use the slider scales or VASs, but radio buttons could be used without any technological prerequisites. The number of discrete response options was quite obvious with radio buttons, but there were no markers on the discrete graphical rating scales that indicated the possible response options. Radio buttons and VAS were operated by point and click, slider scales by drag and drop. Paradata collected were response latencies and how often answers were changed (see Heerwegh, 2003).

As giving ratings on devices with touch screens operated by finger (e.g., smartphones and tablets) differs from giving ratings on desktop computers, this rapidly growing group of respondents will be analyzed separately to gather better insight.

Results

Overall, 1,807 respondents started the experiment. Sliders and VAS required JavaScript to be enabled on respondents' computers. As the availability of technology and respondent characteristics can be confounded (i.e., regarding personality traits, see Buchanan & Reips, 2001; Funke & Reips, 2012a), regardless of the experimental group—even with radio button scales—overall 17 (0.9%) respondents who had not activated this technology were excluded.

Identified nonreactively using client-side paradata, about 7.8% of respondents ($N = 140$) participated using a mobile device with a touch screen (e.g., smartphone or tablet computer), will be analyzed separately. Overall, the net sample consisted of 1,650 participants using desktop computers. Results on initial sample size, break-off, lurking, complete data sets, and response latencies are summarized in Table 2.

Break-Off

Break-off by rating scale. Overall, 51 respondents (3.1%) quit participation during the experiment. Break-off ranged from 1.5% for radio button scales to 2.8% for VAS to 4.2% for sliders, $\chi^2(2, N = 1{,}473) = 6.50$, $p = .039$. Bonferroni corrected χ^2 tests showed a statistically significant difference ($p < .05$) only for sliders versus radio button scales.

Break-off by response options. A nonsignificant ($p > .05$) tendency is that break-off rates get larger with more response options.

Table 2. Initial Sample Size, Break-Off, Lurking, Response Rate, Complete Data Sets Without Nonresponse, and Median Response Latency (Respondents Using Desktop Computers Only).

	N	Break-Off	Lurking	Response Rate (*N*)	Complete Data Sets	Response Time (Median)
3-Point slider	162	3.7%	0.0%	96% (156)	74.4%	29 s
5-Point slider	153	2.6%	0.7%	97% (148)	79.1%	22 s
7-Point slider	185	5.9%	2.3%	92% (170)	82.9%	23 s
Slider overall	500	4.2%	1.0%	95% (474)	78.9%	25 s
3-Point VAS	165	1.8%	0.0%	98% (162)	96.9%	26 s
5-Point VAS	173	3.5%	0.6%	96% (166)	98.8%	23 s
7-Point VAS	163	3.1%	0.6%	96% (157)	99.4%	23 s
VAS overall	501	2.8%	0.4%	97% (485)	98.4%	23 s
3 Radio buttons	167	0.6%	0.6%	99% (165)	98.8%	17 s
5 Radio buttons	159	1.9%	0.0%	98% (156)	98.7%	18 s
7 Radio buttons	146	2.1%	0.7%	97% (141)	100.0%	18 s
Radio buttons overall	472	1.5%	0.4%	98% (463)	99.1%	17 s
Total	1,473	2.9%	0.6%	97% (1422)	92.0%	21 s

Note. VAS = visual analogue scales. Response times (rt) for complete data sets were computed after removing outliers (Q1 − 1.5 * IQR ≥ rt ≤ Q3 + 1.5 * IQR).

Break-off and bias. Participants reported their final school grade, which can be taken as proxy indicator for cognitive abilities, conscientiousness, or need for cognition. Respondents who attended grammar school were divided into two groups (below and at least final grade 3.0). No influence was found for VAS or radio buttons. However, for slider scales, break-off was more than 5 times higher for respondents with a low final school grade, 11% versus 2.2%, $\chi^2(1, N = 347) = 8.65, p = .003$, Odds Ratio [*OR*] = 5.6. This result could be confirmed if type of school was used instead of the final grade: Respondents with A-levels (about 13 years in school) were less likely to break-off with sliders in comparison to respondents from technical secondary school (about 12 years), 3.2% versus 9.1%, $\chi^2(1, N = 436) = 5.93, p = .015$, OR = 3.1. Again, no statistically significant difference was found between VAS and radio buttons.

Break-off on touch screen devices. On smartphones and tablets, break-off rate ranged from 2.3% for radio button scales to 7.5% for VAS to 37% for sliders, $\chi^2(2,\ N = 124) = 21.78,\ p < .001$. Break-off for slider scales was statistically significantly higher in comparison to the other two scales, which did not differ from another. As break-off within sliders was that high and could be confounded with certain respondent characteristics, all further analyses within the subsample using touch screen devices will only be done for VAS and radio button scales.

All participants who broke off were excluded just as all lurkers—that is, participants who did not give a single rating for all 3 items (see Bosnjak, 2001)—ahead of the following analyses.

Response Latencies

Furthermore, 51 outliers—participants with a response time below the first quartile minus 2.5 interquartile ranges or above the third quartile plus 2.5 interquartile ranges within each response scale—were temporarily excluded from analysis of response times.

Response latencies by rating scale. Giving ratings on radio buttons was fastest (*Median* = 17 s), and there was only little difference between VAS (*Median* = 23) and sliders (*Median* = 26 s). As Kruskal–Wallis tests show, the only significant differences are between VAS and radio buttons, $H(1, N = 903) = 336.41$, $p < .001$, and between sliders and radio buttons, $H(1, N = 805) = 363.78$, $p < .001$.

Also on touch screen devices, response times were higher with VAS (*Median* = 32 s) than with radio button scales (*Median* = 20 s), $H(1, N = 66) = 18.90$, $p < .001$.

Response latencies by response options. The number of response options had no influence ($p > .05$) on median response times, that is, 22 s for 3-point scales, 20 s for 5-point scales, and 21 s for 7-point scales.

Item Nonresponse

Although item nonresponse is easy to determine with radio button scales and VAS—as the marker appears only after a click on the rating scale—it is more demanding with sliders because the handle was initially positioned at a valid rating. In this case, only if the slider's handle was not touched at all, it was counted as item nonresponse. All other events (i.e., moving to another position, clicking without moving, or moving back and forth) were counted as valid ratings.

Item nonresponse by rating scale. Only few data sets contained at least one missing answer with radio buttons (0.9%) and VAS (1.6%, $p > .05$). In contrast, 21.1% of the data sets collected with slider scales were incomplete. These differences are statistically significant in comparison to radio button scales, $\chi^2(1, N = 937) = 97.17$, $p < .001$, $OR = 30.7$, and in comparison to VAS, $\chi^2(1, N = 959) = 90.72$, $p < .001$, $OR = 16.0$. See interim discussion for consequences on further analyses. On touch screen devices, item nonresponse was higher, but differences in incomplete data sets between radio buttons (4.8%) and VAS (2.9%) did not reach statistical significance ($p > .05$).

Item nonresponse by response options. The number of response options had no statistically significant ($p > .05$) effect on the number of data sets with at least one missing value and ranged from 9.7% for 3-point scales to 7.4% for 5-point scales to 6.4% for 7-point scales.

Interim Discussion

With slider scales—at least with regard to the implementation used in this study—it is difficult to differentiate between nonresponse and responses of a substantial response category. If all sliders that have not been moved or at least clicked on were counted as item nonresponse, this would lead—as seen earlier—to an inconsiderably high amount of incomplete data sets. There are two ways to deal with this. Either all cases of nonresponse are excluded from analyses—which would substantially decrease the number of cases—or all cases of nonresponse are treated as ratings of middle intensity. Although neither way seems satisfying, for the following analyses, all sliders that have not been moved are counted as valid ratings. This approach ought to lead to an overestimation of the middle response option.

Results (Continued)

Use of Middle Response Option

For all rating scales, Pearson χ^2 tests with Bonferroni correction were used to compare the frequency of the middle response option.

Use of middle category by rating scale. Respondents using VAS chose the middle response option for any of the 3 items at least once in 56% and with radio buttons in 52%. But even though all sliders that had not been moved from the central position were counted as valid ratings, only 42% chose the middle response option at least once with sliders, $\chi^2(2, N = 1{,}422) = 24.82$, $p < .001$. Post hoc tests

Table 3. Mean Ratings (Standard Deviation) by Response Scale (Respondents Using Desktop Computers Only).

	Item 1	Item 2	Item 3
3-Point slider	$M = 75$ ($SD = 32$)	$M = 91$ ($SD = 23$)	$M = 53$ ($SD = 47$)
5-Point slider	$M = 66$ ($SD = 23$)	$M = 80$ ($SD = 22$)	$M = 49$ ($SD = 30$)
7-Point slider	$M = 66$ ($SD = 22$)	$M = 79$ ($SD = 21$)	$M = 50$ ($SD = 29$)
Slider overall	$M = 69$ ($SD = 24$)	$M = 83$ ($SD = 23$)	$M = 51$ ($SD = 36$)
3-Point VAS	$M = 71$ ($SD = 30$)	$M = 88$ ($SD = 23$)	$M = 46$ ($SD = 38$)
5-Point VAS	$M = 66$ ($SD = 20$)	$M = 79$ ($SD = 20$)	$M = 49$ ($SD = 26$)
7-Point VAS	$M = 68$ ($SD = 21$)	$M = 81$ ($SD = 20$)	$M = 48$ ($SD = 27$)
VAS overall	$M = 68$ ($SD = 24$)	$M = 82$ ($SD = 22$)	$M = 48$ ($SD = 31$)
3 Radio buttons	$M = 71$ ($SD = 30$)	$M = 87$ ($SD = 23$)	$M = 47$ ($SD = 38$)
5 Radio buttons	$M = 71$ ($SD = 23$)	$M = 84$ ($SD = 20$)	$M = 48$ ($SD = 28$)
7 Radio buttons	$M = 70$ ($SD = 22$)	$M = 82$ ($SD = 22$)	$M = 48$ ($SD = 28$)
Radio buttons overall	$M = 71$ ($SD = 25$)	$M = 83$ ($SD = 22$)	$M = 47$ ($SD = 32$)
Total	$M = 69$ ($SD = 25$)	$M = 83$ ($SD = 22$)	$M = 48$ ($SD = 33$)

Note. VAS = visual analogue scales. Ratings were recoded so that regardless of the number of response options the lowest value (leftmost response option) was 0 and the highest value (rightmost response option) was 100.

show that only slider scales statistically significantly differ from VAS and from radio buttons, which did not differ from one another.

Use of middle category by response options. At no surprise—because there are more options to choose from—the number of respondents who choose the middle response at least once increases with the number of response options, that is, 72% with 3-point scales, 44% with 5-point scales, and 33% with 7-point scales, $\chi^2(2, N = 1{,}422) = 193.19, p < .001$ (all pairwise comparisons being also statistically significant, $p < .05$).

Mean Ratings

To facilitate comparison, all ratings have been recoded so that 0 represents the lowest, leftmost response option and 100 represents the highest, rightmost response option (e.g., with 5-point scale the scores 0, 25, 50, 75, and 100 were assigned). Mean ratings are illustrated in Table 3.

Overall, only very few differences in means reach statistical significance. Only 3-point scales lead to slightly higher ratings for Item 1 and Item 2 than 5-point and 7-point scales. And 3-point VAS lead to slightly higher ratings for Item 2 in comparison to the other VAS scales.

Changing Responses

Taking advantage of client-side paradata, it was recorded how often respondents changed their ratings for each item (e.g., to correct an answer). Analyses are based on all respondents who rated all 3 items. Due to the problems with slider scales mentioned earlier, these analyses are restricted to VAS and radio button scales.

Changing responses by rating scale. With VAS, 81% of the respondents changed at least one rating, while ratings with radio buttons were adjusted by only 43%, $\chi^2(1, N = 936) = 137.42, p < .001$, OR = 5.4. The mean number of changes with VAS was 3.8 ($SD = 3.4$), with radio button scales just 0.7 ($SD = 0.9$) times, $t(548.79) = 19.00, p < .001$.

Changing responses by response options. For radio button scales, the number of response options did not influence (p > .05) the number of clicks for giving ratings: M(3-point) = 0.6 (SD = 0.9), M(5-point) = 0.8 (SD = 1.0), and M(7-point) = 0.6 (SD = 0.9). Within VAS, the number of clicks needed for all ratings decreases with the number of response options: M(3-point VAS) = 5.2 (SD = 3.4), M(5-point VAS) = 3.2 (SD = 3.1), and M(7-point VAS) = 2.9 (SD = 3.4), $F(2, 474) = 22.97$, $p < .001$, $\eta^2 = .09$. Bonferroni post hoc tests show that all differences are statistically significant (p < .05) except for the difference between 5-point VAS and 7-point VAS.

Discussion

Three types of rating scales—slider scale, VAS, and rating scales made from radio button—were tested in a web experiment in three different implementations as discrete 3-point, 5-point, and 7-point rating scales. The six dependent variables were break-off, response times, item nonresponse, use of middle response option, mean ratings, and how often ratings were changed.

Impact of Number of Response Options

Overall, the number of response options only affected the use of the middle response option, which is likely to be an artifact. As fewer response options available equal fewer choices, thus increasing the likelihood to choose the middle response option.

Impact of Type of Rating Scale

Slider scales. Sliders lead to about 3 times higher break-off rates in comparison to rating scales made from radio buttons. This is of particular concern, as break-off is associated with lower school grades—a proxy indicator either for cognitive abilities or personality traits—which leads to a biased sample composition. Funke et al. (2011) found in a very heterogeneous sample that respondents with a lower formal education are more likely to break-off with sliders. This study replicates the finding even in a quite homogeneous sample consisting of students with a high formal education and broad computer experience. One explanation is that the operation of slider scale—here drag and drop requiring four steps versus point and click with just two steps with the other scales—was more demanding.

Furthermore, slider scales performed especially bad on mobile devices like smartphones or tablet computers with touch screens, where more than one of three respondents quit participation. One explanation is that the sliding gesture on these devices is already reserved for scrolling so that the operation was either unclear or technical problems occurred.

Additionally, on desktop browsers, respondents needed more time to complete the items with sliders than with radio button scales. Although sliders did not affect mean ratings, they affected the distribution of values. The middle response option—where the slider was initially positioned—was less frequently chosen, even though all cases of item nonresponse were treated as middle response option. One explanation for this finding is that respondents do not see the initial response position of the slider as a valid rating. It can be concluded that if the handle is initially placed at a valid response option, a considerable share of respondents uses a slider scale with n response option as an $n - 1$ point scale.

Overall, sliders seem to be confusing for respondents—as illustrated by the high break-off rate and longer response times—but also for the researcher because of the difficult interpretation of sliders that have not been moved. One could argue that a redesigned slider scale (e.g., with the handle outside the actual rating scale) could tackle these problems. However, it is likely that some problems are inherent characteristics of sliders and cannot be solved as any initial position of the slider could produce an anchoring effect leading to biased estimates.

VAS versus radio buttons. Overall, measurement with VAS produced results that were very similar to measurement with radio button scales regarding break-off, complete data sets, use of middle response option, and mean ratings. Differences were only found regarding higher response times with VAS (in line with Couper, Tourangeau, Conrad, & Singer, 2006, and Funke & Reips, 2012b) and for the frequency of changing ratings (in line with Funke & Reips, 2012b). A simple—still to be tested—explanation for higher response times is that respondents changed ratings for the 3 items about 4 times and with radio buttons only once. However, changing answers might also be an indicator for problems with the rating scale.

Couper and colleagues' (2006) findings that VAS produce higher break-off rates were not replicated. However, this could be explained by different conceptions of VAS. Considering the implementation presented in Table 1, Couper et al. probably did not test VAS but slider scales or an in between, which could account for the different results.

As results obtained from VAS and radio button scales are quite similar, it should be possible to tackle the problem of respondents who had not activated JavaScript in their web browsers. In those cases, low-tech radio buttons could be displayed automatically without running the risk of loosing respondents or biasing data.

VAS on touch screen devices. On devices with touch screens, no difference to radio button scales could be observed. Thus, VAS should be especially useful in mobile web surveys on small screen devices like smartphones as the efficient use of space—even small VAS have good measurement properties (see Reips & Funke, 2008)—helps to prevent the problem of scrolling, which is one source for biased results (see Couper, Tourangeau, Conrad, & Crawford, 2004).

Suggested Implementation for VAS

The summary of the results suggests, VAS should be implemented in web surveys in the following way: (1) no initial marker, (2) giving ratings by point and click only (not additionally by sliding or drag and drop), (3) changing ratings should be possible, and (4) radio buttons should be programmed as fallback if JavaScript is not available. Furthermore, continuous implementation (i.e., every pixel in length serves as possible response option) could be considered.

Further Research

The results presented here provides further evidence that VAS do not negatively affect data collection—at least for six out of seven dependent variables tested. In addition to the inherent positive characteristics of (especially continuous) VAS—that is, many possibilities for recoding with a single data set, more powerful tests for data distributions, detection of very small differences, fewer problems with shared ranks, and very efficient use of space—more studies could be done on possible positive effects. As this study was limited to discrete rating scales, more research should be done on continuous scales. Further studies could also focus on the importance of VAS for measurement error, especially formatting error (see Groves et al., 2009). Furthermore, in a panel design, test–retest reliability could be determined in comparison to discrete rating scales. Other research questions could examine how respondents actually change their answers with VAS, if measurement is valid even for very small differences, if higher response times with VAS are either an indicator for problems or deeper cognitive processing, and if the use of VAS even influences question understanding. Finally, disentangling the effect of a marker initially present (here visibility of the slider) and the handling (drag and drop vs. point and click) could be studied.

Conclusion

This study provides further evidence that VAS are feasible in web surveys (e.g., Funke & Reips, 2012b; Studer, 2012). Although sliders and VAS look similar, they affect the data collection process differently. The differences in use—drag and drop versus point and click—as well as the initial presence of a sliding handle substantially influence data quality. The suggested differentiation between these two graphical ratings scales could not only be helpful when designing questionnaires but also when interpreting results from other studies. Overall, the effects of slider scales are negative and it is clearly not advisable to use this scale in web surveys.

In sum, the author would recommend to use rating scales made from radio buttons for measurement of discrete latent variables as no technological-dependent dropout can happen. But whenever continuous latent variables are to be measured, where small differences matter, or the efficient use of space is important, VAS could be the rating scale of choice.

Authors' Note

Frederik Funke is an independent consultant, lecturer, trainer, data gladiator, and freelance scientist in Germany. His website is datagladiator.net. He would like to acknowledge the contribution of the COST Action IS1004. www.webdatanet.eu.

Declaration of Conflicting Interests

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

Funding

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

References

Abend, R., Dan, O., Maoz, K., Raz, S., & Bar-Haim, Y. (2014). Reliability, validity and sensitivity of a computerized visual analog scale measuring state anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*, 447–453.

Averbuch, M., & Katzper, M. (2004). Assessment of visual analog versus categorical scale for measurement of osteoarthritis pain. *Journal of Clinical Pharmacology*, *44*, 368–372.

Bosnjak, M. (2001). Participation in non-restricted web-surveys: A typology and explanatory model for item-nonresponse. In U.-D. Reips & M. Bosnjak (Eds.), *Dimensions of Internet science* (pp. 193–207). Lengerich, Germany: Pabst Science.

Buchanan, T., & Reips, U.-D. (2001). Platform-dependent biases in online research: Do Mac users really think different? In K. J. Jonas, P. Breuer, B. Schauenburg, & M. Boos (Eds.), *Perspectives on Internet research: Concepts and methods*. Retrieved from http://www.psych.uni-goettingen.de/congress/gor-2001/contrib/buchanan-tom

Cork, R. C., Isaac, I., Elsharydah, A., Saleemi, S., Zavisca, F., & Alexander, L. (2004). A comparison of the verbal rating scale and the visual analog scale for pain assessment. *The Internet Journal of Anesthesiology*, 8. doi:10.5580/1a73

Couper, M. P., Tourangeau, R., Conrad, F. G., & Crawford, S. D. (2004). What they see is what we get: Response options for web surveys. *Social Science Computer Review*, *22*, 111–127.

Couper, M. P., Tourangeau, R., Conrad, F. G., & Singer, E. (2006). Evaluating the effectiveness of visual analog scales: A web experiment. *Social Science Computer Review*, *24*, 227–245.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly*, *65*, 230–253.

Flynn, D., van Schaik, P., & van Wersh, A. (2004). A comparison of multi-item Likert and visual analogue scales for the assessment of transactionally defined coping function. *European Journal of Psychological Assessment*, *20*, 49–58.

Funke, F., & Reips, U.-D. (2012a, March). *Better low-tech than sorry: How technophile questionnaires may affect psychological representativeness*. Poster presented at the 14th annual General Online Research (G.O.R.) conference of the German Society for Online Research (D.G.O.F.), Mannheim, Germany.

Funke, F., & Reips, U.-D. (2012b). Why semantic differentials in web-based research should be made from visual analogue scales and not from 5-point scales. *Field Methods*, *24*, 310–327. doi:10.1177/1525822X12444061

Funke, F., Reips, U.-D., & Thomas, R. K. (2011). Sliders for the smart: Type of rating scale on the web interacts with educational level. *Social Science Computer Review*, *29*, 221–231. doi:10.1177/0894439310376896

Gerich, J. (2007). Visual analogue scales for mode-independent measurement in self-administered questionnaires. *Behavior Research Methods*, *39*, 985–992.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley.

Hayes, M. H. S., & Patterson, D. G. (1921). Experimental development of the graphic rating method. *Psychological Bulletin*, *18*, 98–99.

Heerwegh, D. (2003). Explaining response latencies and changing answers using client side paradata from a web survey. *Social Science Computer Review*, *21*, 360–373.

Hofmans, J., & Theuns, P. (2008). On the linearity of predefined and self-anchoring visual analogue scales. *British Journal of Mathematical and Statistical Psychology*, *61*, 401–413.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applies Cognitive Psychology*, *5*, 213–236.

McReynold, P., & Ludwig, K. (1987). On the history of rating scales. *Personality and Individual Differences*, *8*, 281–283.

Myles, P. S., Troedel, S., Boquest, M., & Reeves, M. (1999). The pain visual analog scale: Is it linear or nonlinear? *Anesthesia & Analgesia*, *89*, 1517–1520.

Myles, P. S., & Urquhart, N. (2005). The linearity of the visual analogue scale in patients with severe acute pain. *Anaesthesia and Intensive Care*, *33*, 54–58.

Rausch, M., & Zehetleitner, M. (2014). A comparison between a visual analogue scale and a four point scale as measures of conscious experience of motion. *Consciousness and Cognition*, *28*, 126–140.

Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, *49*, 243–256.

Reips, U.-D., & Funke, F. (2008). Interval level measurement with visual analogue scales in Internet-based research: VAS generator. *Behavior Research Methods*, *40*, 699–704. doi:10.3758/BRM.40.3.699

Studer, R. (2012). *Does it matter how happiness is measured? Evidence from a randomized controlled experiment* (Working paper). Department of Economics, University of Zurich. Retrieved from www.econ.uzh.ch/faculty/studer/vas.pdf

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

Author Biography

Frederik Funke is a freelance survey methodologist, trainer, and consultant for data analysis and web surveys. He also works as a lecturer in academia and senior project manager for online research at the LINK Institut, Frankfurt. He received a master's degree in sociology and a PhD in psychology, both in the area of online research methods. His research focuses on rating scales in web surveys and he published on the topic in *Behavior Research Methods*, *Field Methods*, and *Social Science Computer Review*. For more see http://research.frederikfunke.net. Email: ff@datagladiator.net