Social Science Computer Review
http://ssc.sagepub.com/

Sliders for the Smart: Type of Rating Scale on the Web Interacts With Educational Level

Frederik Funke, Ulf-Dietrich Reips and Randall K. Thomas
Social Science Computer Review 2011 29: 221 originally published online 16 August 2010
DOI: 10.1177/0894439310376896

The online version of this article can be found at:
http://ssc.sagepub.com/content/29/2/221

Published by:
SAGE
http://www.sagepublications.com

Social Science Computer Review
29(2) 221-231

DOI: 10.1177/0894439310376896
http://ssc.sagepub.com
SAGE

Sliders for the Smart: Type of Rating Scale on the Web Interacts With Educational Level

Frederik Funke[1], Ulf-Dietrich Reips[2], and Randall K. Thomas[3]

Abstract

Slider scales and radio buttons scales were experimentally compared in horizontal and vertical orientation. Slider scales lead to statistically significantly higher break-off rates (odds ratio = 6.9) and substantially higher response times. Problems with slider scales were especially prevalent in participants with less than average education, suggesting the slider scale format is more challenging in terms of previous knowledge needed or cognitive load. An alternative explanation, *technology-dependent sampling* (Buchanan & Reips, 2001), cannot fully account for the current results. The authors clearly advise against the use of Java-based slider scales and advocate low-tech solutions for the design of Web-based data collection. Orientation on screen had no observable effect on data quality or usability of rating scales. Implications of item format for Web-based surveys are discussed.

Keywords

web experiment, online questionnaire, web survey, slider scales, radio button scales, interaction of scale type with education, horizontal orientation, vertical orientation

Introduction

Participation in a survey is a special kind of indirect communication between respondent and researcher. As in self-administered studies, no interviewer is present to provide situational clarification aids, respondents use various elements in a questionnaire as cues to infer further information from. In turn, questionnaire design elements and rating scales can be used to help respondents to understand a question in a certain desired way, reducing the variance of perceived question meaning and increasing the data quality. In this article, we focus on *rating scales* and *spatial orientation* as a means of design in a Web-based questionnaire (e.g., Couper, 2008; Dillman, Smyth, & Christian, 2009; Reips, 2000) and how the design may interact with respondent variables such as education (Buchanan & Reips, 2001).

[1]University of Tübingen, Germany
[2]Universidad de Deusto and IKERBASQUE, Basque Foundation for Science, Spain
[3]ICF International, United States of America

Corresponding Author:
Frederik Funke, Steinbruchweg 10, Kassel 34123, Germany.
Email: email@frederikfunke.de

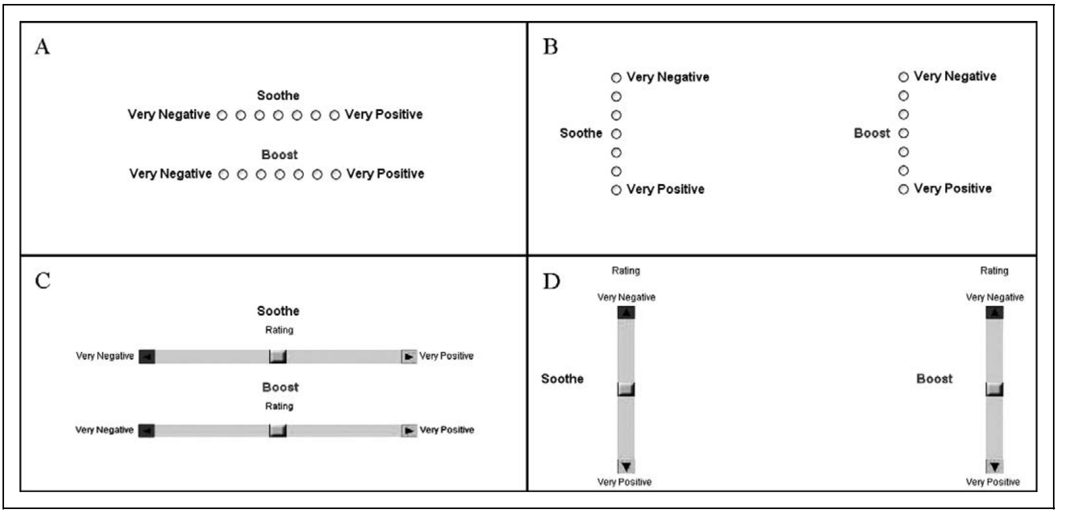

Figure 1. Rating scales used in the experimental conditions in Study 1: radio buttons (A and B) and slider scales (C and D) in horizontal (A and C) and vertical (B and D) orientation.

Design Effects and Web-Based Research

Questionnaire layout and graphical design (for an overview, see Dillman et al., 2009; Lyberg et al., 1997) can have a considerable impact on data quality and substantive answers in both paper-based questionnaires and web surveys (e.g., Christian, Dillman, & Smyth, 2005; Couper, Tourangeau, Conrad, & Singer, 2006; Couper, Traugott, & Lamias, 2001; Reips, 2010; Tourangeau, Couper, & Conrad, 2004, 2007). Rating scales can affect data quality regarding mean ratings, distribution of answers, response time, or item nonresponse (e.g., Couper, Conrad, & Tourangeau, 2007; Healey, 2007; Heerwegh & Loosveldt, 2002; Krosnick, 1999; Krosnick & Fabrigar, 1997; but see Reips, 2002a). Especially Web-based questionnaires are prone to involuntary changes in design (e.g., leading to a nonlinear format, see Christian, Parsons, & Dillman, 2009). These changes mainly occur because of a considerable amount of variation of different hardware and software on the developers' and the respondents' side (Schmidt, 2007). Consequently, a robust low-tech implementation—where serious differences are less frequent—is the method of choice (Reips, 2002b, 2006).

Experimental Manipulation

To gain more information about how rating scales and their design influence data and data quality, we conducted a 2 × 2 web experiment (see Reips, 2007). First, we manipulated the *spatial orientation* of the rating scale on the screen. Second, we tested two different *rating scales*, a slider scale and a common radio button scale (see Figure 1). Furthermore, we looked at high versus low level of education.

Factor 1: Horizontal Versus Vertical Orientation on Screen

Graham (1956) compared horizontal and vertical rating scales and found "shape of the visual field and the relative ease of moving the eyes from side to side, rather than up and down, are thought to account for the greater accuracy on the horizontal scale" (p. 157). Changes in the orientation of graphical rating scales were examined by Downie et al. (1978) as well as by Scott and Huskisson (1979) in paper-based questionnaires. Both compared horizontal and vertical visual analogue scales—

graphical rating scales that are somewhat similar to slider scales—in within-subjects designs and found high correlations between both scales.

There is only little research in Web-based studies dealing with vertical or horizontal design, but it is mostly limited to close-ended questions (e.g., Smyth, Dillman, Christian, & Stern, 2006) or fully anchored rating scales rather than end-anchored scales. Vertical rating scales may be especially interesting for mobile web surveys (see Fuchs, 2008), as most mobile Internet devices (e.g., smart phones) have an upright display that may not display horizontal scales without scrolling. Vertical rating scales on displays in horizontal orientation (and vice versa) could bias results, as response options that are directly visible are preferred to response options only visible after scrolling (Couper, Tourangeau, Conrad, & Crawford, 2004).

Factor 2: Slider Scale Versus Radio Button Scale

Levels were either a simple ordinal rating scale, realized with HTML radio buttons or a slider scale programmed in Java. Both scales provided seven discrete values.

Slider scales are graphical rating scales not available for self-administered paper-and-pencil interviews because a slider has flexible parts that cannot be printed in a questionnaire. In our case, the slider handle was initially positioned in the middle of the rating scale. In contrast to radio buttons scales—on which the appropriate answer is chosen by a single click—ratings with slider scales may be more demanding and require more hand–eye coordination: Respondents have to move the mouse pointer to the slider, then click and hold the mouse button, move the slider to the desired position, and release the mouse button.

An inherent problem with slider scales is the initial positioning of the handle, which could influence the probability of choosing the preselected value, resulting in biased estimates (e.g., if respondents interpret it as a typical judgment or the most desired answer). Even if the handle is initially positioned outside the scale, anchoring effects could occur. Additionally, if the handle is placed at the position of a valid answer, intentional response and non-response cannot be distinguished.

Slider scales cannot be programmed in plain HMTL, more sophisticated programming (e.g., Java, JavaScript, and Flash) is necessary. Client-side availability of the respective technology determines whether a slider scale can be used or not. If a technology is not available in the respondent's browser, it can result in item nonresponse, transmission of a default or invalid value, or even in respondents dropping out of a study (Buchanan & Reips, 2001).

Arnau, Thompson, and Cook (2001) compared an unnumbered slider scale with a 9-point radio button scale. The slider scale was programmed in Java like the slider used in this study, but in contrast, it provided 100 discrete values. They found that the rating scale did not change the latent structure of responses and recommended the use of low-tech radio buttons.

Method

Participants

The study reported was embedded in a survey on health-related products and fielded from December 6 to December 31, 2001. Of the 2,596 respondents who started the survey, 2,340 (90.1%) reached the experimental section. Respondents' software was scanned for technological requirements (Microsoft Windows as operating system, Internet Explorer version 5.0 or higher as the actual browser, and Java installed). Of the 2,340, 1,926 participants (82.3%) met the required technological criteria and were randomly assigned to two independent experiments. In this article, we only refer to the experiment described above. Overall, we obtained a net sample of 779 cases (see Table 1) and a response rate within the experiment of 97%. Slightly more participants were male (50.8%) and the mean reported age was $M = 43.0$ ($SD = 13.6$) years.

Table 1. Assignment to Experimental Conditions

	Alignment		
Rating scale	Horizontal	Vertical	Overall
Slider scale	25.3% (n = 197)	24.9% (n = 194)	50.2% (n = 391)
Radio button scale	23.7% (n = 185)	26.1% (n = 203)	49.8% (n = 388)
Overall	49.0% (n = 382)	51.0% (n = 397)	100% (N = 779)

Formal education. All but seven respondents (99.1%) provided information on their education. We recoded the reported education according to the International Standard Classification of Education (ISCED; see United Nations Educational, Scientific and Cultural Organization, 1997) into two groups: Below ISCED 5A versus ISCED 5A or higher (ISCED 5A is considered as higher education and roughly starts with college degree, for example, BA or BS).

Procedure

The experimental treatment followed the presentation and evaluation of two abstract product concepts—*boost* (“increase both mental and physical energy for up to 3 hrs”) and *soothe* (“relax a person for up to 3 hrs”)—and appeared late in the questionnaire (mean duration of participation before the experiment $M = 9.7$ min, $SD = 6.8$). Both concepts were presented on a single webpage, counterbalanced for order. Ratings with radio button scales were mandatory. With slider scales, the value of the middle category was submitted, if respondents proceeded to the next page without changing the position of the handle.

Results

Break-Off

Overall, 3.0% ($n = 23$) of the respondents quit participation during the experiment (see Table 2). Most of the break-off occurred in the slider scale condition, $\chi^2(1, N = 779) = 12.81, p < .001$, odds ratio = 6.92.

Break-off and formal education. Within the group of respondents with a low formal education, the probability for dropping out was higher with slider scales than with radio button scales, $\chi^2(1, N = 451) = 5.89, p = .018$, odds ratio = 5.45. When only looking at the group of respondents with a high formal education, Fisher’s exact test did not detect any difference in break-off between slider scales and radio button scales, $\chi^2(1, N = 321) = 1.66, p = 1.000$. Spatial orientation did not statistically significantly affect break-off rate in any condition (see Table 2).

Further analyses of response times and responses are based on the remaining 756 complete cases.

Task Duration

Before analysis of response time (e.g., Heerwegh, 2003), we removed all outliers within each group according to Tukey’s (1977) definition: Response times lower than the first quartile minus 1.5 interquartile ranges or higher than the third quartile plus 1.5 interquartile ranges were excluded. Overall, we disregarded 6.7% of response times for these analyses resulting in 704 cases. Neither type of rating scale nor spatial orientation had a statistically significant association with respondent exclusion, $\chi^2(3, N = 756) = 1.79, p = .616$.

Table 2. Break-Off Within Each Condition, Overall, and per Formal Education

Rating scale	Dropout	No dropout
	Overall	
Slider scale—horizontal	5.6% (n = 11)[a]	94.4% (n = 186)
Slider scale—vertical	4.6% (n = 9)[b]	95.4% (n = 185)
Radio button scale—horizontal	1.6% (n = 3)[a]	98.4% (n = 182)
Radio button scale—vertical	0.0% (n = 0)[b]	100.0% (n = 203)
Total	3.0% (n = 23)	97.0% (n = 779)
	Education < ISCED 5A	
Slider scale—horizontal	4.9% (n = 5)	95.1% (n = 98)
Slider scale—vertical	4.3% (n = 5)[c]	95.7% (n = 112)
Radio button scale—horizontal	1.8% (n = 2)	98.2% (n = 107)
Radio button scale—vertical	0.0% (n = 0)[c]	100.0% (n = 122)
Total	2.7% (n = 11)	97.3% (n = 439)
	Education $\geq$ ISCED 5A	
Slider scale—horizontal	2.2% (n = 2)	97.8% (n = 88)
Slider scale—vertical	2.7% (n = 2)	97.3% (n = 73)
Radio button scale—horizontal	1.3% (n = 1)	98.7% (n = 75)
Radio button scale—vertical	0.0% (n = 0)	100.0% (n = 80)
Total	1.6% (n = 5)	98.4% (n = 5)

Note. Means with the same superscript differ significantly according to chi-square tests/Fisher's exact tests (2 issues), only comparisons between same type of scale or same alignment were computed.
[a,c] $p < .05$, [b] $p \leq .001$.

Table 3. Mean Task Duration (*SD*) in Seconds by Condition, Outlier Excluded

	Alignment		
Rating scale	Horizontal	Vertical	Overall
Slider scale	33.1 (11.8)	37.8 (14.7)	35.4 (13.5)***
Radio button scale	14.2 (5.5)	17.3 (5.9)	15.8 (5.9)***
Overall	23.7 (13.2)***	26.9 (15.0)***	25.3 (14.2)***

*** $p \leq .001$

Task duration (see Table 3) was considerably higher with slider scales than with radio buttons, $F(1, 703) = 638{,}23$, $p < .001$, $\eta^2 = .48$. Ratings on horizontal scales took significantly less time to complete than on vertical scales, but the effect size was small, $F(1, 703) = 8.81$, $p = .003$, $\eta^2 = .01$. No significant interaction between scale and spatial orientation occurred, $F(1, 703) = 1.02$, $p = .313$.

We analyzed the influence of spatial orientation separately within each condition because the bulky Java technology needed for the slider scales was likely to increase response times. In comparison to vertical alignment completion took less time with horizontal slider scales, $F(1, 341) = 10.41$, $p = .001$, $\eta^2 = .03$, as well as with horizontal radio button scales, $F(1, 362) = 26.70$, $p < .001$, $\eta^2 = .07$.

Content of Responses

Mean ratings. Table 4 shows the mean ratings of both concepts. Slider scales lead to higher ratings for measurement of both concepts. The difference was statistically significant for *boost,* $F(1, 754) = 7.48$, $p = .006$, $\eta^2 = .01$, but not for *soothe,* $F(1, 754) = 3.17$, $p = .076$. Spatial orientation did not

Table 4. Mean Ratings (Standard Deviations) per Condition

	Alignment		
Rating scale	Horizontal	Vertical	Overall
	Concept *boost*		
Slider scale	4.2 (2.0)	4.2 (2.0)[a]	4.2 (2.0)[b]
Radio button scale	3.8 (1.9)	3.8 (1.9)[a]	3.8 (1.9)[b]
Total	4.0 (2.0)	4.0 (2.0)	4.0 (2.0)
	Concept *soothe*		
Slider scale	4.4 (2.0)	4.2 (2.1)	4.3 (2.0)
Radio button scale	4.2 (1.7)	4.0 (1.8)	4.1 (1.8)
Overall	4.3 (1.9)	4.1 (1.9)	4.2 (1.9)

Note. 1 = *very negative*, 7 = *very positive*. Means with the same superscript differ statistically significantly.
[a] $p < .05$.
[b] $p < .01$.

have a statistically significant influence on ratings, neither for *boost,* $F(1, 754) < 1$, *ns*, nor for *soothe,* $F(1, 754) = 2.85$, $p = .092$.

Mode. We found only small differences in mode values between slider scales and radio button scales. For measurement of the concept *boost*, the mode is shifted one category to the positive end (right respectively bottom) of the scale with slider scales (Category 6 instead of 5). Within each scale, spatial orientation does not matter. Measurement of *soothe* leads to the same mode (Category 6) for all rating scales but radio button scales' mode is shifted two categories to the negative (left respectively top).

Distribution of values. Overall, rating scale has an impact on the distribution of values (see Figure 2) for *boost*, $\chi^2(6, N = 756) = 17.58$, $p = .007$, as well as for *soothe*, $\chi^2(6, N = 756) = 23.62$, $p = .001$. Alignment had neither an influence for the distribution of *boost*, $\chi^2(6, N = 756) = 3.71$, $p = .716$, nor for *soothe*, $\chi^2(6, N = 756) = 10.38$, $p = .110$. As standardized residuals suggest, especially differences in the middle response option make a major contribution to the overall difference between rating scales (see Figure 2).

Use of middle category. For concept *boost*, 10.5% selected the middle category with slider scales, whereas 15.8% gave a middle rating with radio button scales, $\chi^2(1, N = 756) = 4.68$, $p = .031$, odds ratio = 1.60. Results for measurement of the concept *soothe* are comparable: 11.6% chose the middle category with slider scales and 21.0% with radio button scales, $\chi^2(1, N = 756) = 12.30$, $p < .001$, odds ratio = 2.03. Spatial orientation did not matter regarding the use of the middle category for both concepts, neither within nor between scales.

Summary

We compared Java-based slider scales and HTML radio buttons in horizontal versus vertical orientation in a Web-based questionnaire. Analyses were conducted both within the factor rating scales and within the factor spatial orientation. No interactions were observed at all.

Slider Scales Versus Radio Button Scales

Break-off and education. The number of participants breaking off a study is a main indicator for task difficulty or technical problems and can be a major source of bias. Overall, slider scales

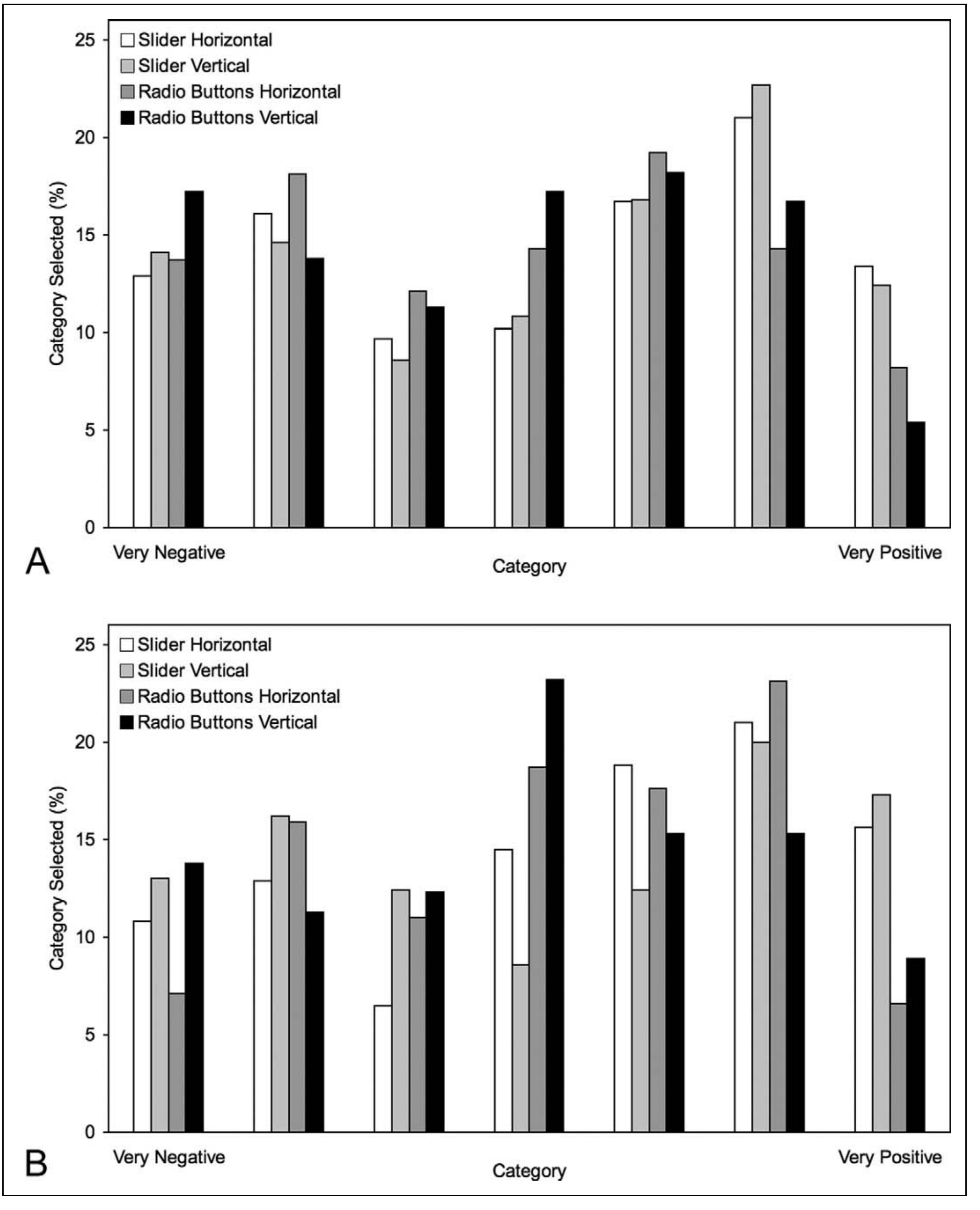

Figure 2. Distribution of values for measurement of concept *boost* (A) and concept *soothe* (B).

produced about seven times more break-offs than radio button scales. We observed a strong relation between low formal education—a simplistic proxy for cognitive abilities—and break-off. This difference resonates well with the study by Buchanan and Reips (2001), where education and personality traits also interacted with technology to explain nonresponse behavior. The current results suggest that certain formats are more challenging in terms of previous knowledge needed or cognitive load.

Response time. In the current study, we take a longer duration not as an indicator for deeper cognitive processing but for problems on the stage of formatting the answer on the available rating scale (see Sudman, Bradburn, & Schwarz, 1996). Slider scales lead to a significant increase of response time with a very large effect size, regardless of respondent's level of formal education. This finding is in line with Couper et al. (2006), who found substantially higher response times with Java-based visual analogue scales (see Reips & Funke, 2008). It remains inconclusive if technical reasons (e.g., loading time of the Java applet) or if the different and more demanding handling of the slider account for the main share of variance in response times.

Distribution of values. In comparison to measurement with radio button scales, *fewer* respondents chose the middle category with slider scales for both evaluated concepts. This is especially remarkable as the center category was preselected with slider scales and if respondents proceeded with the questionnaire without changing the position of the slider, this was treated as an answer of middle intensity. There are two contrasting post hoc explanations for this. Either the respondents thought that the middle position was no valid choice and treated the scale not like a 7-point scale but like a 6-point scale. Or they felt forced to communicate an attitude tending toward one extreme. With the current data, this issue remains inconclusive but it illustrates that making ratings on slider scales substantially differs from making ratings with radio button scales.

Central tendency. We observed small upward shifts of mean ratings for measurement with slider scales in comparison to radio button scales. The difference, however, reached statistical significance only for 1 of the 2 items and the effect size was low, that is, roughly half of a point on the 7-point scale. The consistently higher ratings with slider scales could be the effect of respondents being reluctant to use the middle category and instead choosing a higher category.

Horizontal Versus Vertical Alignment

The only observable effect of changing alignment of rating scales was that judgments on vertical scales took slightly longer on average than ratings on horizontal scales.

Conclusions

Keep Scales Simple

Possible anchoring effects, more demanding usage in terms of cognitive effort for scale administration, as well as problems with identifying intentional versus non-intentional responses and the initial position of the slider are serious negative scale characteristics of slider scales. These restrictions likely cannot be compensated for by a better technological implementation. Additionally, we found empirical evidence for higher response times and a higher break-off rate with respondents with a low formal education, which disqualifies slider scales for being used as rating scale. As these disadvantages outweight the potential advantages by far, the authors clearly advice against the use of Java-based slider scales in general. Less demanding techniques following a low-tech paradigm should be used whenever possible (see Reips, 2002b). HTML radio buttons should be used for to measure ordinal concepts and visual analogue scales for interval concepts (Reips & Funke, 2008).

Twist Without Risk

Overall, it seems that horizontal and vertical layout can be substituted mutually broadening the possibilities for the layout of web surveys without running the risk of negatively affecting the measurement process. This may be especially valuable for mobile web surveys on handheld devices such as cellular phones, where small screens require a careful use of space.

Declaration of Conflicting Interests

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

Funding

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

References

Arnau, R. C., Thompson, R. L., & Cook, C. (2001). Do different response formats change the latent structure of responses? An empirical investigation using taxometric analysis. *Educational and Psychological Measurement*, *61*, 23-44.

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* [WWW document]. Retrieved from http://www.psych.uni-goettingen.de/congress/gor-2001/contrib/buchanan-tom

Christian, L. M., Dillman, D. A., & Smyth, J. D. (2005). *Instructing Web and telephone respondent to report date answers in format desired by surveyor* (Technical Report 05-067). Retrieved July 16, 2008, from http://www.sesrc.wsu.edu/dillman/papers.htm

Christian, L. M., Parsons, N. L., & Dillman, D. A. (2009). Designing scalar questions for Web surveys. *Sociological Methods & Research*, *37*, 393-425.

Couper, M. P. (2008). *Designing effective Web surveys.* Cambridge, UK: Cambridge University Press.

Couper, M. P., Conrad, F. G., & Tourangeau, R. (2007). Visual context effects in Web surveys. *Public Opinion Quarterly*, *71*, 623-634.

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.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: Wiley.

Downie, W. W., Leatham, P. A., Rhind, V. M., Wright, V., Branco, J. A., & Anderson, J. A. (1978). Studies with pain rating scales. *Annals of the Rheumatic Diseases*, *37*, 378-381.

Fuchs, M. (2008). Mobile Web survey: A preliminary discussion of methodological implications. In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 77-94). New York, NY: Wiley.

Graham, N. E. (1956). The speed and accuracy of reading horizontal, vertical, and circular scales. *Journal of Applied Psychology*, *40*, 228-232.

Healey, B. (2007). Drop downs and scroll mice: The effect of response option format and input mechanism employed on data quality in Web surveys. *Social Science Computer Review*, *25*, 111-128.

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

Heerwegh, D., & Loosveldt, G. (2002). An evaluation of the effect of response formats on data quality in Web surveys. *Social Science Computer Review*, *20*, 471-484.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537-567.

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. deLeeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 141-164). New York, NY: Wiley.

Lyberg, L., Biemer, P., Collins, M., deLeeuw, E., Dippo, C., Schwarz, N., & Trewin, D. (Eds.). (1997). *Survey measurement and process quality.* New York, NY: Wiley.

Reips, U.-D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-118). San Diego, CA: Academic Press.

Reips, U.-D. (2002a). Context effects in Web surveys. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences* (pp. 69-80). Seattle, WA: Hogrefe & Huber.

Reips, U.-D. (2002b). Internet-based psychological experimenting: Five dos and five don'ts. *Social Science Computer Review*, *20*, 241-249.

Reips, U.-D. (2006). Web-based methods. In M. Eid & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 73-85). Washington, DC: American Psychological Association.

Reips, U.-D. (2007). The methodology of Internet-based experiments. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.), *Oxford handbook of Internet psychology* (pp. 373-390). Oxford, UK: Oxford University Press.

Reips, U.-D. (2010). Design and formatting in Internet-based research. In S. Gosling & J. Johnson (Ed.), *Advanced Internet methods in the behavioral sciences* (pp. 29-43). Washington, DC: American Psychological Association.

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.

Schmidt, W. C. (2007). Technical considerations when implementing online research. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.), *Oxford handbook of Internet psychology* (pp. 459-470). Oxford, UK: Oxford University Press.

Scott, J., & Huskisson, E. C. (1979). Vertical or horizontal visual analogue scales. *Annals of the Rheumatic Diseases*, *38*, 560.

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006). Effects of using visual design principles to group response options in Web surveys. *International Journal of Internet Science*, *1*, 5-15.

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

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2004). Spacing, position, and order: Interpretative heuristics for visual features of survey questions. *Public Opinion Quarterly*, *68*, 368-393.

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2007). Color, labels, and interpretative heuristics for response scales. *Public Opinion Quarterly*, *71*, 91-112.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

United Nations Educational, Scientific and Cultural Organization (1997). *International standard classification of education – 1997 version*. Retrieved July 16, 2008, from http://www.uis.unesco.org/TEMPLATE/pdf/isced/ISCED_A.pdf

Bios

Frederik Funke (email@frederikfunke.de) studied sociology, psychology, and philosophy. His research centers on methodological problems of Web-based surveying, especially on the impact of ratings scales on data quality. Much of his recent research and application development has focused on visual analogue scales (see http://vasgenerator.net).

Ulf-Dietrich Reips (reips@deusto.es; http://iscience.deusto.es/) is an IKERBASQUE research professor at University of Deusto in Bilbao, Spain. He is working on Internet-based research methodologies, in particular Internet-based experimenting, the psychology of the Internet, and the cognition of causality. In 1994, he

founded the Web Experimental Psychology Lab, the first laboratory for conducting real experiments on the World Wide Web, and provides many web services for researchers via his iScience Server at http://iscience.eu.

Randall K. Thomas (randall.k.thomas@gmail.com) is the Chief Survey Methodologist for ICF International in Fairfax, Virginia and has more than 25 years experience in designing questionnaires and conducting survey projects across multiple modes and across countries. One of his primary areas of specialty is in research on the effects of questionnaire design effects in Web-based survey.