Frederik Funke[1] & Ulf-Dietrich Reips[2]

Difference and Correspondences Between Visual Analogue Scales, Slider Scales and Radio Button Scales in Web Surveys

Research interest

We conducted a Web experiment to compare data and paradata from three scales (see Figure 1) for answering closed-ended questions in self-administered Web questionnaires: visual analogue scales (VAS), slider scales (SLS) and radio button scales (RBS).

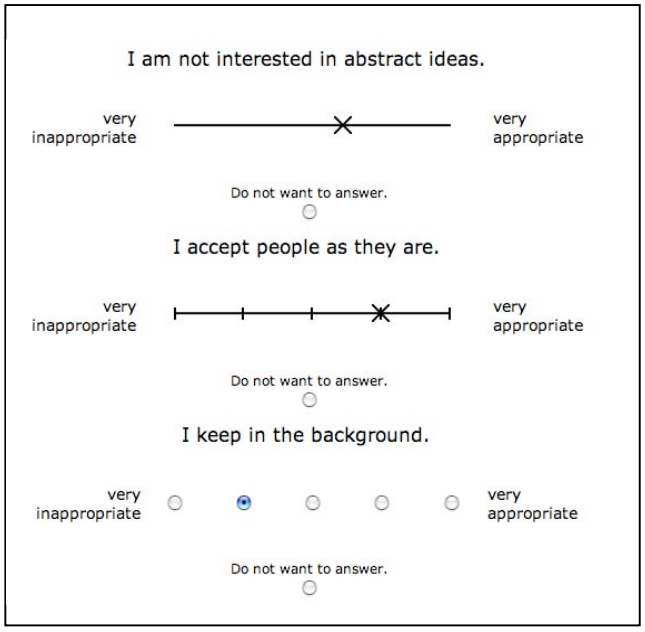


Figure 1. Scales compared in this study: VAS, SLS and RBS (from top to bottom)

VAS are nearly continuous measurement instruments, each pixel is clickable and results in a raw value. In contrast, RBS only provide a limited number of categories. As we know from previous research, data from VAS reach the desired level of an interval scale (Reips & Funke, in press) and there is a systematic difference between the scales (especially concerning the extremes; see Funke & Reips, 2007). SLS are an in-between answer format. They are visually similar to VAS, but functionally more similar to the RBS. Do the scales have a different influence on data collection and data quality? Does usage follow function or appearance?

[1] University of Tübingen, Germany; http://www.frederikfunke.de; email@frederikfunke.de
[2] University of Zurich, Switzerland; u.reips@psychologie.uzh.ch

Design

Respondents of a 40 item personality inventory (validated for Internet usage by Hartig, Jude, & Rauch, 2003) were randomly assigned to either a VAS with 250 possible values, an SLS with 5 discrete values, or a 5 point RBS. Directly after load neither scale provided an initial marker (with VAS and SLS the marker appeared after the first click on the scale). The VAS used in this survey could only be clicked (just as the RBS) whereas the SLS' marker could be clicked or slid but its final position was limited to the indicated discrete values.

Empirical data

As shown in Table 1, completion rate and missing data were not statistically significant different for the three answer formats, but there was a tendency that VAS performed better on these indicators than SLS and RBS (in contrast to findings by Couper, Tourangeau, Conrad, & Singer, 2006). As expected (see Couper et al., 2006) response time was higher with VAS than with RBS.

Table 1. Indicators for data quality

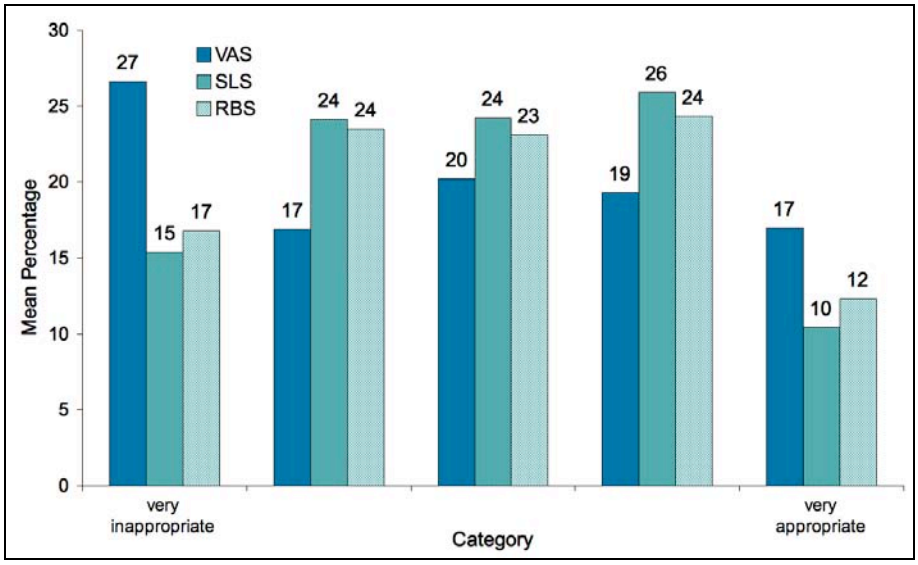
Indicator	VAS	SLS	RBS	Total	
		Paradata			
Completion rate[a]	97%	95%	94%	96%	n.s.
n (net sample)[b]	107	87	88	282	
Mean missing data rate[b]	0.6%	0.9%	0.9%	0.8%	n.s.
Mean response time per item in seconds (*SD*)[b, c]	7.3 (2.0)	6.9 (1.9)	6.6 (1.5)	6.9 (1.8)*	
		Central tendency[d]			
M (*SD*) for all 40 items	2.84 (0.17)	2.92 (0.16)	2.92 (0.20)	2.89 (0.18)**	
		Test-retest reliability[e]			
n (net sample)	32	23	27	82	
Mean reliability for all 40 items	.88	.82	.83	.84***	

[a]Serious respondents only. [b]Serious and complete respondents only. [c]Adjustment: unreasonably high response times (> 60 sec) and outlier (i.e. not within *M* +/-2.5 interquartile ranges) were removed. [d]Scales running from1 to 5. [e]Two waves; complete and serious respondents only.
*p < .05. **p < .01. ***p < .001.

Means were different, $F(2, 279) = 5.67$, $p = .004$, with VAS measuring lower mean scores. We were able to determine test-retest reliability with a smaller sample and found a statistically significant difference between the scales, $V(2, 3280) = 35.7$, $p < .001$, with VAS producing the highest reliability scores.

Finally, we compared the distributions of values. Therefore we transformed data from VAS into 5 categories consisting of equal data intervals (see Figure 2). Our finding: SLS and RBS are used in a similar way, the distribution of VAS' frequencies differs substantially.

Figure 2. Comparison of frequencies



Conclusion

It makes a difference if VAS, SLS or RBS are used. Function is more important than appearance: even though SLS look more like VAS, they are used like RBS regarding data quality and distribution of values and we found no benefit in employing SLS. Confirming our previous finding on VAS' superior data level (Reips & Funke, in press), VAS again turn out to be the better scales. Present evidence shows them to be advantageous regarding dropout, missing data, and reliability of measurement. Drawing a line between discrete categories (as with SLS) is not enough. Superficial changes in appearance do not substitute for the power of a continuous measurement device like the VAS.

References

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*(2), 227-245.

Funke, F., & Reips, U.-D. (2007, June). *Improving data quality in Web surveys with visual analogue scales.* Paper presented at the second conference of the European Research Association, Prague (CZ).

Hartig, J., Jude, N., & Rauch, W. (2003). Entwicklung und Erprobung eines deutschen Big-Five-Fragebogens auf Basis des International Personality Item Pools (IPIP40) [Development and test of a German big five questionnaire based on the International Personality Item Pool (IPIP40)]]. *Arbeiten aus dem Institut für Psychologie der Johann Wolfgang Goethe Universität*, 2003/1. Frankfurt/Main.

Reips, U.-D., & Funke, F. (in press). Interval level measurement with visual analogue scales in Internet based research: VAS Generator. *Behavior Research Methods*.