Investigating Cognitive Effort and Response Quality of Question Formats in Web Surveys Using Paradata

Jan Karem Höhne
University of Göttingen (Germany)

Stephan Schlosser
University of Göttingen (Germany)

Dagmar Krebs
University of Gießen (Germany)

Abstract

Measuring attitudes and opinions employing agree/disagree (A/D) questions is a common method in social research, because it appears to be possible to measure different constructs with identical response scales. However, theoretical considerations suggest that A/D questions require a considerable cognitive processing. Item-specific (IS) questions, in contrast, offer content-related response categories, implying less cognitive processing. To investigate the respective cognitive effort and response quality associated with A/D and IS questions we conducted a web-based experiment with n = 1,005 students. Cognitive effort was assessed by response times and answer changes. Response quality, in contrast, was assessed by different indicators such as dropouts. According to our results single IS questions require higher cognitive effort than single A/D questions in terms of response times. Moreover, our findings show substantial differences in processing single and grid questions.

Keywords: agree/disagree questions, asking manner, item-specific questions, paradata, survey focus

Introduction

Since Likert's (1932) well-known article "A Technique for the Measurement of Attitudes" was published, the use of agree/disagree (A/D) questions (i.e. response categories are based on an agreement continuum) have become increasingly popular in attitude measurement. For instance, A/D questions are frequently used in the German General Social Survey, the Eurobarometer, and the International Social Survey Program. The question topics can range anywhere from political and societal attitudes to religion and belief systems or health and demographics. The reasons for the popularity are twofold (Saris et al. 2010): first, it seems to be possible to measure different contents with identical response scales. Second, they streamline questionnaires – especially if grids are used.

This document is a preprint and thus it may differ from the final version: Höhne, J.K., Schlosser, S., & Krebs, D. (2017). Investigating cognitive effort and response quality of question formats in web surveys using Paradata. Field Methods, 29, 365–382. DOI: 10.1177/1525822X17710640.

Although the merits of A/D questions are obvious, this question format has several disadvantages (Fowler 1995): first, question stems are frequently not clearly located at one end of the content continuum. Second, the cognitive processing is complex, because third, respondents must translate their mental decision into a dimension of "agree" vs. "disagree". Finally, they are prone to response bias such as acquiescence or social desirability (Converse and Presser 1986; Lelkes and Weiss 2015). Therefore, Fowler (1995) asserts that item-specific (IS) questions (i.e. response categories directly express the attitude or opinion of interest) represent a much simpler, more direct, and more informative method than A/D questions.

To investigate how respondent's process survey questions the use of paradata is a suitable strategy. In our experiment, we use a (comprehensive) paradata system that facilitates, among others, the measurement of response times, the position of the computer mouse and mouse clicks. For cognitive effort, we use the time a respondent needs to come up with an answer and select a response category and the number of answer changes measured by mouse clicks. For instance, Lenzner, Kaczmirek, and Lenzner (2010) investigated cognitive burden of different question comprehensibility determinants by means of response times. Heerwegh (2011) and Stern (2008) examined the occurrence of answer changes due to mapping difficulties using mouse clicks. Regarding response quality, we use different indicators such as dropouts (see Chang and Krosnick 2009; Galesic and Bosnjak 2009; Lenzner, Kaczmirek, and Lenzner 2010).

Theoretical Overview

A special technique to measure different attitudes and opinions was proposed by Likert (1932) that contains statements being assessed on a five-point, fully labeled A/D response scale; arraying respondents along an agreement continuum. However, the empirical study of response bias, such as acquiescence, incited methodological research on the adequacy of A/D questions (Höhne and Lenzner 2015; Krebs and Hoffmeyer-Zlotnik 2010; Krosnick 1991; Schuman and Presser 1996). Therefore, locating or arraying respondents along an agreement dimension rather than on a content dimension is receiving increasing attention in survey methodology. As mentioned, A/D questions frequently cause problems, since response categories cannot be interpreted clearly. Consider the following example for an illustration of the A/D question format:

To what extent do you agree or disagree with the following statement? Before an election, I inform myself thoroughly about the suitability of the candidates – *agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly*

From a psychological perspective, answering A/D questions requires that respondents accomplish specific mental tasks. Carpenter and Just (1975), among others, developed a model of the response process focusing on the characteristics of A/D questions. According to this model, respondents must first, discover the literal meaning of words or phrases (*semantic meaning*). Then they must identify the dimension of interest – i.e. "What does the researcher actually mean?" (*pragmatic meaning*). In this example, the dimension of interest could be "thoroughness of informing" – identified by the words "inform myself" and "thoroughly". After having deciphered the pragmatic meaning, respondents must perform a mental positioning task

(*placement on the scale of interest*), considering how thoroughly they inform themselves before an election. Finally, they must translate their mental decision into the A/D response categories (*mapping*). Implicitly, this means that ordering respondents on an A/D continuum correlates monotonically to ordering them on the scale of interest (Krosnick and Presser 2010) – in the example above, "thoroughness of informing". However, a monotonic equivalence between the scale of interest and the A/D continuum has not yet been empirically proved. For this reason the following example of an IS question seems to be a (methodologically) more promising way to ask survey questions:

How thoroughly do you inform yourself about the suitability of the candidates before an election – *very thoroughly, thoroughly, neither nor, not thoroughly, or not at all thoroughly.*

As stated by Fowler (1995) response categories of A/D questions have a low discriminatory power since they are considered in terms of "agree" vs. "disagree". Consequently, they require an intricate mapping process and render little information about the question topic. The response scales of IS questions, in contrast, express the underlying content dimension directly; providing an opportunity of a better discrimination between response categories. For this reason, IS questions are highly recommended for rendering more information (Fowler 1995; Krosnick and Presser 2010; Saris et al. 2010), and may yield better response quality.

Further, while the occurrence of response bias in A/D questions is well documented in the survey literature (Höhne and Lenzner 2015; Krebs and Hoffmeyer-Zlotnik 2010; Krosnick 1991; Schuman and Presser 1996), the reasons as to why this question format produces such bias are quite uncertain. Altogether, it can be presumed that A/D questions foster boredom and weariness, since the "asking manner" does not change. Respondents must repeat the same answering task over and over again, especially if grids are used. Thus, A/D questions seemingly dismay respondents and discourage them from diligently reading and responding (Krosnick 1991). In contrast, IS questions vary the "asking manner" from question to question so that they presumably do not decrease respondents' attention and/or motivation, thereby rendering a more appropriate response quality.

Although A/D questions presumably are associated with difficult cognitive processing respondents do not need to read the response categories repeatedly since the "asking manner" does not change. Respondents can (mentally) extrapolate the response continuum of A/D questions (Höhne and Lenzner 2015); fostering a superficial cognitive response process. IS questions, while presumably simpler to process, require constant reconsideration of the dimension of interest encouraging respondents to perform an active and intensive cognitive processing of the individual question. It is to assume that IS questions are cognitively more demanding, and thus, require more cognitive effort. Accordingly, it is further to assume that response quality is better for IS than for A/D questions.

Hypotheses

The web-based experiment provides the opportunity to operationalize cognitive effort and response quality associated with A/D and IS questions by means of paradata. While cognitive

effort is indicated by response times and answer changes, speeding, dropouts, and non-differentiation are taken as indicators of (primarily bad) response quality.

Before presenting our hypotheses we introduce the notion of "presentation mode" of survey questions because it is directly related to our concept of "asking manner", and thus, to cognitive effort. We presented A/D and IS questions in single and grid presentation modes. While for single questions, response categories are arranged below each question, for grids they are arranged alongside each question. Organizing IS questions in grids is a rather uncommon procedure, but it can be applied if all IS questions address the same dimension of interest. While differences between A/D and IS single questions are discussed in the survey literature, they are (mostly) neglected within grid presentation mode.

Response Times

Measuring response times enjoys a long tradition in social psychology and survey research (Couper and Kreuter 2013; Yan and Tourangeau 2008) and has been proven as a useful strategy to investigate cognitive effort (Bassili 1996; Bassili and Fletscher 1991; Bassili and Scott 1996; Fazio 1990; Yan and Olson 2013). Generally, it is assumed that the time of processing corresponds (directly) to the cognitive effort required to answer a question. This, in turn, suggests that the longer a respondent needs to respond, the higher the cognitive effort must be. In the context of our previously described concept of "asking manner", we assume that single IS questions cause significantly higher response times than their single A/D counterparts (*hypothesis 1a*). Since the "asking manner" of A/D and IS questions in grids does not change, we additionally hypothesize that the cognitive effort in this presentation mode is the same for both question formats. Hence, no significant differences in response times are expected between A/D and IS grid questions (*hypothesis 1b*).

Answer Changes

As suggested by Heerwegh (2011) and Stern (2008) answer changes can be caused by different aspects of response formats, such as low discriminatory power of response categories. In general, answer changes can result from the uncertainty of respondents to decide between response categories; especially if an intensive response process is supposed. Hence, the number of answer changes can be used as an (additional) indicator of cognitive effort. Due to the changing "asking manner" this linkage applies particularly to single IS questions. We therefore assume that single IS questions produce significantly more answer changes than their single A/D counterparts (*hypothesis 2a*). Additionally, we expect no significant differences in answer changes between A/D and IS grid questions since both do not change "asking manner" (*hypothesis 2b*).

Response Quality

Response quality depends largely on the difficulty of a task itself (Krosnick 1991); implying that the higher the task difficulty the higher the probability of poor response quality. As indicators of poor response quality we use speeding (extremely fast responding without the probability of careful processing), dropouts (break-off the survey), and non-differentiation (rating several issues identically) (Chang and Krosnick 2009; Galesic and Bosnjak 2009;

Lenzner, Kaczmirek, and Lenzner 2010). Because A/D questions are tiring and tend to attenuate respondents' motivation we expect to observe a lower response quality for single A/D than for single IS questions (*hypothesis 3a*). In line with the previous hypotheses, we expect no differences in response quality between A/D and IS grid questions (*hypothesis 3b*).

Research Design

To control for carry-over effects from grid to single questions and vice versa, we varied the presentation sequence within each question format. Table 1 describes our 2-by-2 research design.

Table 1: Experimental design defined by question format and presentation sequence

Experimental Group	Question Format	Presentation Sequence	Sample Size
1	A/D	GS	255
2	IS	GS	237
3	A/D	SG	278
4	IS	SG	235

Notes: A/D = agree/disagree and IS = item-specific; GS = grid/single questions and SG = single/grid questions.

Survey Questions

The single questions used were taken from the *Cross Cultural Survey for Work and Gender Attitudes* (2010). The grid questions, in contrast, were partially taken from the *German General Social Survey* (2006). Taking questions from established social surveys includes the advantage of using implicitly valid questions for our study.

The web survey contained 24 experimental questions; eight single questions that dealt with work and competition (achievement motivation) and sixteen grid questions that dealt with the importance of different job motivations. For each question taken from the surveys, we developed a counterpart preserving question content. All questions were attitudinal questions with a 5-point end-labeled response scale (see Appendix for examples of single and grid questions in both question formats).

One key requirement to investigate the cognitive effort associated with different question formats using response times is to keep the questions identical in the number of syllables, because these influence the processing time (Baddeley 1992). Therefore, we developed all single and grid questions (including response categories) so that they did not differ in more than two syllables from one another.

Procedure

This study was conducted at two German Universities in May 2015. Participants were invited by email (in total we sent out n = 58,829 invitation emails). The invitation email included an introduction to the topics of the survey and a URL link directing respondents to the survey. Once there, an introductory page informed them about the procedure of the survey and instructed them to read the questions carefully and in the given order. Respondents were additionally informed that they were participating in an experiment and that different types of paradata (e.g. response times) are collected during the survey. While each single question

appeared on an extra screen, grid questions were presented on two screens, with eight questions per grid, respectively.

We used "Embedded Client Side Paradata (ECSP)" (Schlosser 2016), a JavaScript based system to observe respondents' activities and behavior during the web survey. Recorded and stored paradata include response times in milliseconds (the time elapsing between question presentation on the screen and the time the page was submitted), mouse movements and clicks, scrolling, keystrokes, the screen size in pixels, and information about the activity of the web survey browser while processing.

Sample

In total, n = 2,884 students participated in the web survey, which corresponds to a response rate of 4.9%. Due to technical difficulties, we excluded all participants with mobile devices, such as smartphones and tablets (n = 709), versions of Microsoft Internet Explorer earlier than 11 (n = 24), and those who had deactivated JavaScript (n = 28). Moreover, some participants were ineligible because they only visited the title page (n = 122), dropped out of the web survey before being asked any experimental questions (n = 357), German was not their mother tongue (n = 107), and two experimental conditions being subject of an additional article on adverbial intensifiers (n = 532). Altogether, n = 1,005 participants remained for statistical analyses. To deal with response time outliers, we used, among others, "Survey Focus", a new outlier definition procedure. Accordingly, we first exclude as outliers all respondents who left the web-survey page for a certain time. For the remaining respondents, we applied an outlier definition based on the response time distributions (Hoaglin, Mosteller, and Tukey 2000): excluding as outliers all respondents with response times below or above the median plus/minus the upper and lower quartile range multiplied by three. The amount of outliers is evenly distributed over the experimental groups.

Participants were between 17 and 52 years old with a mean age of M = 24.8 (SD = 4.2). 51.6% of the participants were female. All participants were graduates of a college preparatory school or university and 93% had participated in a web survey once before. Sample composition over the experimental groups did not reveal any statistically significant differences regarding age (χ^2 = 2.44; df = 3, p = .49), gender (χ^2 = 2.35; df = 3, p = .50), and survey experience (χ^2 = .60; df = 3, p = .89).

Results

Depending on the specific research question, the analysis of web survey paradata can be classified into different levels of aggregation (Heerwegh 2011; Yan and Olson 2013). To investigate the cognitive effort associated with A/D and IS questions, we analyze paradata on a survey level (i.e. paradata are aggregated across respondents and variables for each experimental group).

Response Times

Response times as an indicator of cognitive effort are the primary dependent variable of the following analyses. Due to the fact that response times typically are right-skewed (Fazio 1990; Ratcliff 1993), we applied a logarithmic transformation to decrease the skewness of response

time distributions. Moreover, the following statistical analyses are based on unadjusted response time measurements without checking for baseline reading speed (see Couper and Kreuter 2013).

To investigate whether the question format (A/D vs. IS) and/or the presentation sequence (grid/single vs. single/grid questions) have an effect on response times we calculated two-way analyses of variance for single and grid questions, separately. While question format has a highly significant effect on response times for single questions ($F(1,915) = 37.64$, $p < .001$), there is no main effect of the presentation sequence and no interaction effect between the two factors. For grid questions, however, neither significant main effects nor an interaction effect were observable.

Next, to investigate whether IS questions produce significantly higher response times than A/D questions, we calculated two one-way analyses of variance with the factor question format (A/D vs. IS) for single and grid questions separately and Cohen's *d* as a measure of effect size. Table 2 displays the statistical results for single questions and reveals (highly) significant differences in average response times between the four experimental groups. In line with our expectations, single IS questions require consistently higher response times than their A/D counterparts. This result is independent of the presentation sequence. Moreover, there are no significant differences between the two groups (1 and 3) receiving A/D questions and the two groups (2 and 4) receiving IS questions. This result is confirmed by Cohen's *d* (Cohen 1969) because the effect sizes of those groups getting the same response format are very small ($d = .12$ and $d = .06$). Hence, presentation sequence does not matter within question format thereby strengthening the result of the two-way analysis of variance for single questions.

Table 2: Mean differences of response times between the four experimental groups for the eight aggregated single questions.

	Differences of means for log-transformed response time data					
Presentation Form	Agree/ Disagree (GS) *Group 1*	Item-Specific (GS) *Group 2*	Agree/ Disagree (SG) *Group 3*	*F* value ($df_1 = 3$)	df_2	*p* value
Item-Specific (GS) *Group 2*	-.108** (S.E. = .023) d = .44			13.75	915	$p < .001$
Agree/ Disagree (SG) *Group 3*	-.030 (S.E. = .022) d = .12	*.078* (S.E. = .022) d = .33*				
Item-Specific (SG) *Group 4*	*-.123** (S.E. = .023) d = .50*	-.015 (S.E. = .024) d = .06	-.093** (S.E. = .022) d = .40			

Notes: *p < .01; **p < .001. Coefficients in italics represent differing experimental conditions and cannot be compared directly. Mean differences: means of column conditions minus means of row conditions. Cohen's d states the effect sizes. GS = grid/single sequence and SG = single/grid sequence.

In contrast to single questions, we did not expect longer response times for the sixteen IS grid questions than for their sixteen A/D counterparts. The statistical results displayed in table 3 with an empirical significance level of $p = .197$ corroborate that there are no differences in average response times between the four experimental groups, as suggested by *hypothesis 1b*. Accordingly, Cohen's *d* shows consistently very small effect sizes ($d < .2$). Hence, the cognitive effort associated with A/D and IS questions seems to be the same as long as they are presented in grids. To annotate the observed small tendency of longer response times for the "grid/single" than the "single/grid" sequence in table 3, we additionally investigated the time to respond to the first grid question. The results reveal that the "grid/single" sequence causes significantly longer response times than the "single/grid" sequence for the first grid question.

Table 3: Mean differences of response times between the four experimental groups for the sixteen aggregated grid questions.

	Differences of means for log-transformed response time data					
Presentation Form	Agree/ Disagree (GS) *Group 1*	Item-Specific (GS) *Group 2*	Agree/ Disagree (SG) *Group 3*	*F* value ($df_1 = 3$)	df_2	*p* value
Item-Specific (GS) *Group 2*	.018 (S.E. = .024) d = .07			1.56	918	$p = .197$
Agree/ Disagree (SG) *Group 3*	.049 (S.E. = .023) d = .19	*.032 (S.E. = .023) d = .13*				
Item-Specific (SG) *Group 4*	*.020 (S.E. = .024) d = .08*	.003 (S.E. = .024) d = .01	-.029 (S.E. = .023) d = .12			

Notes: Coefficients in italics represent differing experimental conditions and cannot be compared directly. Mean differences: means of column conditions minus means of row conditions. Cohen's d states the effect sizes. GS = grid/single sequence and SG = single/grid sequence.

Answer Changes

We used answer changes measured by mouse clicks as a second indicator of cognitive effort (Heerwegh 2011; Stern 2008). Particularly, we expected, that IS single questions would produce a higher number of answer changes than A/D single questions. Furthermore, we expected to observe no differences between A/D and IS grid questions. Again, we analyzed the amount of answer changes separately for single and grid questions. Regarding single questions, there are no significant differences between the four experimental groups ($F(3,915) = .15$, $p = .928$), implying that answer changes in single questions are neither affected by question format

nor by presentation sequence. For IS grid questions, however, there is a (slightly) significant difference ($F(3,918) = 2.72$, $p = .044$) in average answer changes between the groups 2 and 4, revealing a higher number of answer changes for the "single/grid" sequence. Thus, presentation sequence seems to have an impact on answer changes within IS grid questions.

Response Quality

We used the following indicators to measure response quality: speeding, dropouts, and non-differentiation. Since for these analyses the presentation sequence is irrelevant, groups 1 and 3 (having received A/D questions) as well as groups 2 and 4 (having received IS questions) were pooled. The analyses were still conducted separately for single and grid questions.

As an indicator of speeding, we considered the lower 10th percentile of all response times to be extremely fast in responding. For single questions, 6.6% ($n = 61$) of respondents producing the fastest response times did so in response to an A/D question and 2.8% ($n = 26$) in response to an IS question. With a chi-square value of $\chi^2 = 11.45$; $df = 1$, $p < .001$ this indicates a highly significant difference in speeding between the two question formats. For grid questions, 6.3% ($n = 58$) of the respondents with the fastest response times produced them in response to an A/D question and 4.8% ($n = 44$) in response to an IS question. With a chi-square value of $\chi^2 = .97$; $df = 1$, $p = .33$, the response quality of grid questions in terms of extremely fast responding does not significantly differ between the two question formats. Again, as postulated the observed differences occur only between single A/D and IS questions.

The second response quality indicator are dropout rates; implying a premature break-off of the survey. For A/D single questions 2.4% ($n = 28$) and for IS single questions 2.7% ($n = 31$) of the respondents dropped prematurely from the web survey, $\chi^2 = .77$; $df = 1$, $p = .38$. Similarly, for A/D grid questions 1.8% ($n = 20$) and for IS grid questions 1.6% ($n = 18$) of the respondents have left the web survey sooner, $\chi^2 = .01$; $df = 1$, $p = .92$. Apparently, the two question formats do not differ in terms of dropouts irrespective of the presentation mode.

Finally, identical ratings of different issues or objects are an indicator of non-differentiation. To investigate this response quality indicator, the proportion of respondents who answered several survey questions equally was considered. Comparing the number of these respondents for single questions results in marginal differences between the A/D condition (9.6%; $n = 88$) and the IS condition (8.9%; $n = 82$), $\chi^2 = .11$; $df = 1$, $p = .75$. For grid questions, a similar pattern is observable since selection of the same response category over all questions occurred for 15.5% ($n = 143$) of the A/D questions and for 13.6% ($n = 125$) of the IS questions ($\chi^2 = .18$; $df = 1$, $p = .67$). Accordingly, A/D and IS single and grid questions yield a comparable response quality in terms of non-differentiation.

Discussion and Conclusion

The aim of our study was to investigate the cognitive effort and response quality associated with A/D and IS questions by means of paradata. According to the results we found that IS single questions require a higher cognitive effort, in terms of response times, than A/D single questions. Regarding the theoretical complexity of cognitive processing and cognitive effort, we therefore conclude: while cognitive processing of A/D questions is theoretically seen as more demanding than that of IS questions the actually expended cognitive effort is lower.

This discrepancy between the complexity of cognitive processing and cognitive effort can be explained using our concept of "asking manner". For A/D questions the response continuum runs continuously from "agree" to "disagree" or vice versa, so that respondents have to read response categories only once and repeat the same answering task over and over again; requiring little cognitive effort, and thus, resulting in shorter response times than IS questions. In contrast, IS questions change the "asking manner" from question to question since the underlying response categories address the dimension of interest directly. Therefore, reading of IS questions including the "item specific" response categories requires not only more time but challenge respondents to engage in an active and intensive response process for each question in the light of its specific content. The results regarding the concept of "asking manner" confirm the theoretical considerations on cognitive effort (in terms of response times) required to respond to A/D and IS questions.

Besides question format the presentation mode (single or grid questions) might have an impact on cognitive effort as well. As expected in the context of our "asking manner" concept A/D and IS grid questions reveal no differences in response times, and thus, no differences with respect to cognitive effort. Hence, in this respect IS questions organized in grids seem to adjust to A/D questions. Again, this result supports our concept of "asking manner" as an important link between response time and cognitive effort.

Next, we investigated the number of answer changes in A/D and IS questions presented in single and grid mode as further indicator of cognitive effort. Contrary to recent research indicating systematic differences in answer changes between response formats (Heerwegh 2011; Stern 2008) we could not find major differences between A/D and IS questions irrespective of presentation mode. Hence, the difficulty of mapping answers into response categories seems to be equal for both question formats in the two presentation modes. However, it must be noted that we tested end instead of fully labeled response scales eventually rendering the offered response categories virtually equivalent.

Regarding low response quality, we found only minor differences between the two question formats, except for speeding. In line with our concept of "asking manner", we observed that speeding occurs more often for single A/D questions than for their IS counterparts. This can be seen as additional evidence that A/D questions mislead respondents to perform superficial cognitive information processing.

Using a student sample to conduct our research on A/D and IS questions might be seen as a limitation to this study. However, the use of student samples does not limit the external validity of empirical results *per se*. Moreover, since the participants of our study are university students presumably with above-average (cognitive) abilities participating voluntarily without any incentives, we tested our research question under harsh conditions. Due to self-selection, however, it is possible that mainly participants with high motivation or interest started the survey (see Couper et al. 2004). Since these factors are important regarding response quality it would be desirable for further research to control for the influence of motivation and/or interest.

Due to the fact that our research is related to PC devices we suggest that further research investigates the cognitive effort and response quality associated with A/D and IS questions regarding different device types. Moreover, the influence of presentation sequence on grids might be interesting because our results indicate that the time respondents need to answer the

first grid question is longer for the "grid/single" than for the "single/grid" sequence. Therefore, attention should be given to the processing time for the first and subsequent grid questions.

In sum, our findings have both theoretical and practical implications. From a theoretical perspective, we can show that A/D questions do not require a higher cognitive effort than IS questions. Indeed, theoretically, A/D questions imply a quite difficult and sophisticated response process, as suggested by Carpenter and Just (1975). In the context of our findings, however, it seems to be more likely that they tempt respondents to engage in a perfunctory (cognitive) response process. We consider the unchanging "asking manner" a major reason for this. This objection is especially applicable to grid questions. Therefore, we argue that primarily the continuity of the "asking manner" is a very important key determinant of the cognitive effort and response quality associated with different question formats. Moreover, our empirical findings and especially our theoretical reasoning illustrate apparent differences between the presumed complexity of cognitive processing and the expended cognitive effort: demanding cognitive processing can be attenuated by an unchanging "asking manner" decreasing respondents' motivation to diligent question processing. Given that IS questions require more cognitive effort we – in line with Saris et al. (2010) – recommend to use this format because the higher the cognitive effort the more diligent the response process is supposed to be.

References

Baddeley, A. 1992. Working Memory. *Science* 255:556–559.

Bassili, J. N. 1996. How and Why of Response Time Latency Measurement in Telephone Surveys. In *Answering Questions. Methodology for Determining Cognitive and Communicative Process in Survey Research*, ed. N. Schwarz and S. Sudman, 319–346. San Francisco, CA: Jossey-Bass

Bassili, J. N., and J. F. Fletscher. 1991. Response Time Measurement in Survey Research. A Method for CATI and a new Look at Nonattitudes. *Public Opinion Quarterly* 55:331–346.

Bassili, J. N., and B. S. Scott. 1996. Response Latency as a Signal to Question Problems in Survey Research. *Public Opinion Quarterly* 60:390–399.

Carpenter, P. A., and M. A. Just. 1975. Sentence Comprehension: A Psycholinguistic Processing Model of Verification. *Psychological Review* 82:45–73.

Chang, L., and J. A. Krosnick. 2009. National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality. *Public Opinion Quarterly* 73:641–678.

Cohen, J. 1969. *Statistical Power Analysis for the Behavioral Science.* New York, NY: Academic Press.

Converse, J. M., and S. Presser. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire.* Beverly Hills, CA: Sage.

Couper, M. P., and F. Kreuter. 2013. Using Paradata to Explore Item Level Response Times in Surveys. *Journal of the Royal Statistical Society* 176:271–286.

Couper, M. P., R. Tourangeau, F. G. Conrad, and S. D. Crawford. 2004. What They See Is What We Get: Response Options for Web Surveys. *Social Science Computer Review* 22:111–127.

Fazio, R. H. 1990. A Practical Guide to the Use of Response Latency in Social Psychological Research. In *Research Methods in Personality and Social Research,* ed. C. Hendrick and M. S. Clark, 74–97. Newbury, UK: Sage

Fowler, F. 1995. *Improving Survey Questions. Design and Evaluation.* Thousand Oaks, CA: Sage.

Galesic, M., and M. Bosnjak. 2009. Effects of Questionnaire Length on Participation and Indicators of Response Quality in a Web Survey. *Public Opinion Quarterly* 73:349–360.

Heergwegh, D. 2011. Internet Survey Paradata. In *Social and Behavioral Research and the Internet. Advances in Applied Methods and Research Strategies,* ed. M. Das, P. Ester, and L. Kaczmirek, 325–348. New York, NY: Routledge.

Hoaglin, D. C., F. Mosteller, and J. W. Tukey. 2000. *Understanding Robust and Exploratory Data Analysis.* New York, NY: John Wiley & Sons.

Höhne, J. K., and T. Lenzner. 2015. Investigating Response Order Effects in Web Surveys using Eye Tracking. *Psihologija* 48:361–377.

Krebs, D., and J. H. P. Hoffmeyer-Zlotnik. 2010. Positive First or Negative First? Effects of the Order of Answering Categories on Response Behavior. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 6:118–127.

Krosnick, J. A. 1991. Response Strategies for Coping with the Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology* 5:213–236.

Krosnick, J. A., and S. Presser. 2010. Question and Questionnaire Design. In *Handbook of survey research,* ed. P. V. Marsden, and J. D. Wright, 263–313. Bingley, UK: Emerald.

Lelkes, Y., and R. Weiss. 2015. Much Ado about Acquiescence: The relative Validity and Reliability of Construct-Specific and Agree-Disagree Questions. *Research and Politics*.

Lenzner, T., L. Kaczmirek, and A. Lenzner. 2010. Cognitive Burden of Survey Questions and Response Times: A Psycholinguistic Experiment. *Applied Cognitive Psychology* 24:1003–1020.

Likert, R. 1932. A Technique for the Measurement of Attitudes. *Archives of Psychology* 140:1–55.

Ratcliff, R. 1993. Methods for Dealing with Reaction Time Outliers. *Psychological Bulletin* 114:510–532.

Saris, W., M. Revilla, J. A. Krosnick, and E. M. Shaeffer. 2010. Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options. *Survey Research Methods* 4:61–79.

Schlosser, S. (2016). *Embedded Client Side Paradata (ECSP).* https://zenodo.org/record/55169#.V1gBT-RSE4E (accessed June 08, 2016). DOI: 10.5281/zenodo.55169

Schuman, H., and S. Presser. 1996. *Questions and Answers in Attitude Surveys. Experiments on Question Form, Wording, and Context.* Thousand Oaks, CA: Sage.

Stern, M. J. 2008. The Use of Client-side Paradata in Analyzing the Effects of Visual Layout on Changing Responses in Web Surveys. *Field Methods* 20:377–398.

Yan, T., and K. Olson. 2013. Analyzing Paradata to Investigate Measurement Error. In *Improving Surveys with Paradata. Analytic Uses of Process Information,* ed. F. Kreuter, 73–96. Hoboken, N.J.: John Wiley & Sons.

Yan, T., and R. Tourangeau. 2008. Fast Times and Easy Questions. The Effects of Age, Experience, and Question Complexity on Web Survey Response Times. *Applied Cognitive Psychology* 22:51–68.

Appendix

Example of single and grid questions with an A/D and IS question format.

Single Questions

Agree/Disagree (A/D)					***Item-Specific (IS)***				
I try harder when I'm in competition with other people.					When I'm in competition with other people I try …				
agree strongly				*disagree strongly*	*very much harder*				*not at all harder*
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Grid Questions

Agree/Disagree (A/D)

	agree strongly				*disagree strongly*
A high income is important to me.	☐	☐	☐	☐	☐
A flextime is important to me.	☐	☐	☐	☐	☐

Item-Specific (IS)

	very important				*not at all important*
To me a high income is ...	☐	☐	☐	☐	☐
To me a flextime is ...	☐	☐	☐	☐	☐

Notes: The original questions (including response categories) did not differ in more than two syllables from one another and were formulated in German. All questions (including screenshots) are available from the first author.