

## The use and positioning of clarification features in web surveys

Anke Metzler, Tanja Kunz, and Marek Fuchs

*Darmstadt University of Technology, Institute of Sociology, Germany*

Clarification features are used in Web surveys to improve the quality of responses. It is generally advised to place clarification features after the question stem. However, based on initial findings of an eye-tracking study (Kunz & Fuchs, 2012), we expected that the optimal position depends on the respective stage of the question-answer process as it is referred to by the clarification feature. In three Web surveys, the use and positioning of clarification features were tested in open-ended questions, with three different positions being experimentally varied: before the question stem, after the question stem, and after the answer box. Results indicated that contrary to expectations, the optimal position of clarification features did not differ depending on the respective stage of the question-answer process. Clarification features were principally most effective when they were positioned after the question stem, whereas clarification features placed before the question stem were least effective in improving the quality of responses.

*Keywords:* open-ended questions, questionnaire design, definitions, instructions, data quality

Web surveys became a popular mode of data collection. However, besides various advantages such as cost benefits, lack of interviewer effects and fast data collection, Web surveys also face several challenges. Just like self-administered surveys in general, a core characteristic of Web surveys is that respondents have to read and interpret the survey questions and then format their answers by themselves. There is no interviewer who provides additional support in order to understand the question meaning, who motivates respondents to thoughtfully search their memories for all relevant information which is needed to answer the questions, and who encourages respondents to provide their responses in the desired format. Due to a lack of interviewer assistance as well as due to the fact that all information is usually presented visually, question wording and visual features of questionnaire design play a particularly important role in Web surveys, in common with self-administered surveys in general (Christian & Dillman, 2004; Jenkins & Dillman, 1997). Respondents often draw on verbal and visual questionnaire features such as information provided by preceding questions, or in the case of closed questions, information provided by the predefined response categories and

their verbal, numeric, or symbolic labels in order to better understand the meaning of a survey question and to get an idea of how to format their responses (Schwarz, Knauper, Hippler, Noelle-Neumann, & Clark, 1991; Strack, 1992; Toepoel & Dillman, 2011). In the case of open-ended questions, however, no predefined response categories are provided which is why open-ended questions are generally more likely to be misinterpreted by the respondents, are more burdensome for respondents, and suffer from lower data quality compared to closed questions in all kinds of surveys (Dillman, Smith, & Christian, 2014, pp. 128-134; Friberg & Rosenvinge, 2013; Peytchev, 2009). Thus, whether open-ended questions yield high-quality responses mainly depends on the respondents' motivation to think carefully about the issue in question and their willingness to provide detailed responses (Dillman et al., 2014, p. 131). Nevertheless, open-ended questions are quite often used due to the fact that respondents can provide their answers spontaneously and without being influenced by predefined response categories, resulting in a larger variety of responses to open-ended questions compared to closed questions (Reja, Manfreda, Hlebec, & Vehovar, 2003).

In order to improve the quality of responses to open-ended questions in Web surveys, respondents need to develop a consistent understanding of the question meaning in accordance with the survey researcher's intention, they need to be sufficiently motivated to expend the effort to think carefully about the question, and they need to provide a sufficiently detailed answer without additional assistance or probing by the interviewer. As previous studies have shown, additional information such as definitions of key terms and concepts, examples, or instructions provided jointly with the question stem can improve the quality of the respondents' answers to open-ended questions in Web surveys (Couper, Kennedy, Conrad, & Tourangeau, 2011; Smyth, Dillman, Christian, & McBride, 2009). However, although providing additional clarification features can help enhance data quality, respondents do not always recognize their need for clarification, and even if they admit their need for clarification, additional information is often not considered in the question-answer process (Conrad, Couper, Tourangeau, & Peytchev, 2006; Conrad, Schober, & Coiner, 2007; Redline, 2013). Thus, although the provision of clarification features and the way in which they are presented is crucial to receiving optimal responses to open-ended questions in Web surveys, little research has been devoted to this issue. Therefore, in the present paper, the use and positioning of clarification features in open-ended questions was examined in several Web surveys.

## **Background and research questions**

**The design and position of clarification features.** Previous research on clarification features in Web surveys considered two different aspects. First, it has been assessed whether clarification features such as definitions, retrieval cues, motivating statements, or formatting instructions should be presented along with the question by default, irrespective of whether respondents actually need the additional information or not, or whether clarification features should be displayed only if actively requested by respondents. For instance, Conrad et al. (2006) examined several respondent-initiated methods of providing clarification

features. Their findings suggested that respondents are not willing to make the extra effort of obtaining the clarifying information. They showed that additional information was requested more frequently when respondents could access the clarification features via a mouse rollover, whereas even one mouse click on a hyperlink seemed to be a disproportionate effort for the respondents. In this regard, several studies proved that presenting clarification features such as definitions of key terms by default was more effective than solely providing them on the respondents' request, since the additional information gains more attention and is therefore more effective in affecting survey responses if provided, irrespective of the respondents' willingness to request the information (Conrad et al., 2006; Conrad et al., 2007; Galesic, Tourangeau, Couper, & Conrad, 2008; Peytchev, Conrad, Couper, & Tourangeau, 2010).

In addition to the problem that respondents are often not willing to make the extra effort to obtain clarification features, they also often do not realize their need for clarification. Respondents tend to rather rely on their everyday understanding of key terms and concepts, increasing the risk that their interpretations do not match the meaning intended by the researcher (Conrad et al., 2006; Conrad et al., 2007; Tourangeau et al., 2006). Thus, clarification features that are always visible to respondents seem to be the best way to convey additional information and to increase the likelihood that they are read and integrated in the processing of survey questions.

Even though there is little dispute that clarification features should be presented by default, the optimal position of clarification features is still being discussed. Concerning the position of definitions of ambiguous terms or concepts in Web surveys, Couper (2008) noted that "by placing it [a definition] between the question and response options, the respondent's eyes will likely move over the definition in the normal course of reading" (p. 289). Peytchev et al. (2010) also suggested presenting definitions before the response options in order to increase the likelihood that respondents would recognize a definition and also to increase the likelihood that they would read it completely. By contrast, the findings of Redline (2013) indicated that placing clarification features in terms of definitions before the question stem was more effective than placing them after the question stem. However, Redline (2013) did not visually separate the definitions from the core question text, but presented them jointly as one continuous text. Thus, taking the conventional "left to right" and "top to bottom" reading order of respondents into account, respondents had to read or at least scan through the definition that precedes the question stem in order to reach the core question text, whereas definitions following the core question text could easily be skipped, since respondents are likely to stop reading once they have reached the end of the core question text. Christian et al. (2005, 2007) analyzed the positioning of formatting instructions referring to date answers. They tested the effect of verbal labels and symbols in conveying the desired response format. The results were in line with their expectations that providing symbols instead of verbal labels increased the percentage of respondents that were reporting their answers in the desired format, with symbols being the most effective when placed to the left of the answer boxes. Again, this finding can be explained by the

conventional reading order increasing the probability that formatting instructions were more likely to be considered by the respondents, when placing them within the respondents' navigational path (Conrad et al., 2006; Dillman et al., 2014, pp. 187-189).

**Clarification features within the question-answer process.** The general wisdom according to which it is advisable to place clarification features within the respondents' navigational path raises the question of whether clarification features of all kinds should actually be placed in the same position relative to the remaining components of a survey question. Answering a survey question requires respondents to go through various cognitive stages. According to one of the most prominent models of the question-answer process, respondents have to go through four stages in order to arrive at a thorough answer: comprehension, retrieval, judgment and estimation, and reporting. These stages are not necessarily processed in a strict sequence. Instead, respondents are likely to go back and forth between these stages (Cannell, Miller, & Oksenberg, 1981; Tourangeau, Rips, & Rasinski, 2000, pp. 165–167; for a similar model see Sudman, Bradburn, & Schwarz, 1996, pp. 56–58). Taking into account the various types of clarification features such as definitions, retrieval cues, motivating statements, and formatting instructions, each of these refers to a different stage of the question-answer process. Hence, when respondents read and process the different components of a survey question in a certain sequence, various types of clarification features may also differ with respect to the position within the respondents' navigational path at which they are considered by the respondents and actually integrated into the question-answer process.

On the first stage of the question-answer process, i.e., question comprehension, respondents need to understand the question and then interpret its meaning. It is important that respondents interpret the question consistently and in line with the meaning intended by the researcher which, however, is often difficult to achieve (Fowler, 1995, pp. 2-3; Schwarz & Oyserman, 2001). Thus, in order to clarify the meaning of questions and thereby help respondents understand the questions correctly and consistently, clarifying information in terms of definitions of unclear or ambiguous terms and concepts can be provided (Conrad et al., 2006; Peytchev et al., 2010).

Once respondents have understood the question and derived its meaning, they have to retrieve relevant information from memory. On this second stage of the question-answer process, i.e., information retrieval, respondents often experience difficulty recalling all relevant information, because they are either not able or not willing to expend the cognitive effort that is necessary to thoughtfully search their memory (Krosnick, 1991). At this stage, clarification features in terms of retrieval cues activating the memory search process or motivating statements requesting, for example, to think carefully and to recall all relevant information are assumed to improve the likelihood of exhaustive retrieval (Tourangeau, Conrad, Couper, & Ye, 2014).

Processes on this second stage of information retrieval and the third stage of the question-answer process, i.e., judgment and estimation, are highly

integrated, which is why it is difficult to clearly distinguish between these two cognitive processes. Generally, respondents use the information retrieved to make a judgment or estimation. In the case of attitude questions, respondents “may either retrieve a previously formed opinion from memory, or they may ‘compute’ an opinion on the spot” (Schwarz, 1997, p. 32). Similarly for behavioral questions, respondents may either recall and count “relevant instances of this behavior from memory” (Schwarz, 1997, p. 32), or they may estimate the frequency based on some rates with or without correcting for exemptions (Schwarz, 1997; Schwarz & Oyserman, 2001; Sudman et al., 1996, pp. 253-257). Because the second and third stages are highly integrated, clarification features addressing the third stage only are rarely being used.

At the final stage of the question-answer process, i.e., reporting, respondents are expected to format and edit their answers. Closed questions request respondents to map their answer onto the response options provided, whereas in open-ended questions respondents have to formulate a response in their own words, which requires more cognitive effort than just selecting one of the response options offered in the questionnaire (Dillman et al., 2014, p. 131; Peytchev, 2009). Before communicating their answers respondents usually edit the answers, taking into account facets of social desirability and self-presentation (Schwarz & Oyserman, 2001). Besides the visual design of an answer box in terms of its size as well as the absence or presence of verbal or symbolic labels (Fuchs, 2009a, 2009b), clarification features can specify how to format the answer. In numeric open-ended questions, clarification features in terms of formatting instructions provide detailed information concerning the desired format of dates, durations, or amounts, and can help respondents format their responses as requested (Couper et al., 2011; Fuchs, 2007). Further, although Web surveys enable automated response validations in terms of presetting the type of data, format, and range of acceptable answers to avert formatting errors in numeric open-ended questions, respondents “may have tolerance thresholds for acceptable amounts of prompting” (Peytchev & Crawford, 2005, p. 456). Thus, because such types of interactive edit checks may be annoying to respondents and involve the risk that respondents break off from the survey, they should be used carefully and not too often (Bethlehem & Biffignandi, 2012, p. 173; Peytchev & Crawford, 2005). In narrative open-ended questions, formatting instructions can ask respondents, for example, to provide an answer which is as detailed as possible or to take sufficient time in answering a question (Oudejans & Christian, 2011; Smyth et al., 2009). Furthermore, “informing respondents that their answers are important, and clarifying why they are important, gives them a reason to expend the time and energy needed to produce good open-ended responses” (Dillman et al., 2014, p. 131).

Kunz and Fuchs (2012) examined the use of various types of clarification features for open-ended questions and the extent of any attention which respondents might actually have paid to them at differing positions by recording the respondents’ (n = 108) eye movements during Web survey completion. In a lab-experimental between-subjects design, the position of clarification features was varied by presenting them either before the question stem, after the question stem,

or after the answer box. Furthermore, they tested different types of clarification features which addressed various stages of the question-answer process. Findings indicated that depending on the stage addressed by the clarification features, the optimal position of clarification features varied: Clarification features supporting the comprehension stage were best positioned before the question stem. Clarification features influencing the retrieval process were most effective before or after the question stem. However, clarification features concerning the formatting process received more attention and thus were more effective when presented after the question stem or after the answer box. Thus, initial findings suggested that the effectiveness of clarification features varies as a function of the stage of the cognitive question-answer process they refer to, in combination with the position at which they are presented (Kunz & Fuchs, 2012).

### **Hypotheses**

Clarification features such as definitions, retrieval cues, motivating statements, or formatting instructions can principally help improve data quality in open-ended questions. At the same time, however, they are often not sufficiently considered by respondents. This raises the question of whether visual design in Web surveys can help increase the extent of any attention which respondents pay to clarification features which, in turn, may be decisively influenced by the respective position of a clarification feature relative to the remaining components of a survey question.

Generally, it was assumed that survey responses to open-ended questions would differ depending on the presence or absence of clarification features. We assumed that the use of clarification features would improve the quality of survey responses to open-ended questions by clarifying the question meaning and explaining the favored response format, or by motivating the respondents to think carefully about the question and then provide a detailed response.

Furthermore, the effectiveness of clarification features was assumed to vary as a function of their position within the established navigational path as well as depending on the respective stage of the question-answer process they referred to. According to the findings of Kunz and Fuchs (2012), it was assumed that the order in which different components of a survey question were presented could specifically be used to promote the respondents' attention to the clarification features. Consequently, instead of taking a single optimal position for all types of clarification features for granted, the optimal position of a clarification feature was supposed to rather depend on the respective cognitive stage it refers to.

Three different positions of clarification features were conceivable, namely before the question stem, after the question stem, and after the answer box. Taking into account the established navigational path and considering the respective stage addressed by the clarification features we expected, in line with the findings of Kunz and Fuchs (2012), that (H1) definitions promoting the comprehension of a question would be best positioned before the question stem, since a proper understanding of key terms and concepts is a basic prerequisite for a correct understanding when respondents start reading the survey question;

(H2) retrieval cues and motivating statements supporting the retrieval of relevant information would be most effective after the question stem, because at this position they will be considered by respondents after reading the question stem, which is expected to motivate respondents to not prematurely abandon the memory search process; whereas, (H3) formatting instructions specifying the desired format of an answer would be most effective when they were placed after the answer box, because at this position instructions will be noticed by respondents when they turn their attention to the answer field to provide their response, and thus, a clarification feature appears in the navigational path when it is needed by respondents.

## Method

### Experimental design

In order to test the use and optimal position of clarification features, a series of experimental questions were included in three Web surveys conducted among university applicants at Darmstadt University of Technology (Germany) in 2012 ( $n = 5,977$ ), 2013 ( $n = 7,395$ ), and 2014 ( $n = 5,996$ ) (see Table 1). The three Web surveys asked for “qualifications and expectations of university applicants” and comprised about 40 survey questions. Response rates amounted to 32 percent in 2012, 40 percent in 2013, and 35 percent in 2014 (AAPOR RR6). The three Web surveys were designed as census surveys among all applicants for a university place in the respective year. Since the overwhelming majority of applicants applied to more than one university at the same time, statistical hypothesis tests were conducted. In order to increase response rates, nonrespondents and breakoffs received up to two reminders during fieldwork. Females constituted 48 percent of the respondents in the 2012 Survey, 44 percent in the 2013 Survey, and 45 percent in the 2014 Survey. The respondents’ mean age was 21 years in 2012 and 2014, respectively, and 20 years in 2013. Participants showed on average good German language skills with a mean German grade of 2.4 in all three surveys (on a scale ranging from 1 = very good to 6 = insufficient). Finally, respondents’ prior survey experience measured by the number of Web surveys they have taken part in within the last twelve months was limited in our samples with an average of two Web surveys in 2012 and 2014, respectively, and three Web surveys in 2013.

Table 1  
*Overview of the three Web surveys*

	2012 Survey	2013 Survey	2014 Survey
Sample	University applicants		
Topic	Qualifications and expectations of university applicants		
Field phase	July/August 2012	July/August 2013	July/August 2014
Net sample size (n)	5,977	7,395	5,996
Response rate (%)	32	40	35
Number of questions	42	37	42
Gender (% female)	48	44	45
Age (mean)	21	20	21
Secondary school grade (subject German, mean)	2.4	2.4	2.4
Number of prior Web surveys (mean)	1.8	2.3	2.6

Using a between-subjects design, the presence or absence of clarification features as well as three different positions of clarification features were tested. Respondents were randomly assigned either to the control group (CG), where no clarification feature was presented, or to one of the three experimental groups (EGs) providing the clarification features at varying positions: Clarification features were displayed either before the question stem (EGa), after the question stem (EGb), or after the answer box (EGc) (see Figure 1). The clarification features were visually separated from the core question stem by providing the additional information in a separate paragraph and in normal instead of bold typeface. Thus, respondents were easily able to distinguish between the core question stem and the additional information conveyed by the clarification features (Dillman et al., 2014, p. 189).



*Figure 1.* Exemplary illustration of the experimental conditions – from the top left to the bottom right position: no clarification feature (CG), before the question stem (EGa), after the question stem (EGb), and after the answer box (EGc). Original questionnaire in German; translation by authors.

## Experimental questions and dependent variables

The effectiveness of using clarification features in open-ended questions and variations in their positioning relative to the remaining question components was tested for two different kinds of open-ended questions: numeric open-ended questions asking for short answers such as dates, times, frequencies and counts, and narrative open-ended questions seeking longer answers in the respondents' own words. Besides numeric and narrative open-ended questions, Couper et al. (2011) distinguish open-ended questions that require short verbal responses or single-word responses which, however, were not considered.

The content of the clarification features referred to the comprehension, retrieval, and formatting stage of the question-answer process. For each of these stages, different versions of clarification feature types were distinguished which are described in greater detail below (see Table 2). As aforementioned, the process of retrieving information (second stage) and



the process of judging and estimating (third stage) can hardly be separated from each other. By implication, we did not explicitly distinguish between these two stages, but included clarification features which primarily referred to information retrieval.

To address question comprehension (stage one of the question-answer process), a definition was provided that would affect the respondents' understanding of a question in terms of either extending or restricting the perceived meaning of the key concept measured in numeric open-ended questions. The frequency reported by a respondent is deemed an indicator of the effectiveness of definitions. Reported frequencies were expected to increase (decrease) with greater effectiveness of extending (restricting) definitions. For example, respondents assigned to one of the experimental groups with a definition extending the scope of the question meaning were expected to report, on average, higher frequencies, numbers, or amounts as compared to respondents in the control group where no definition was provided. Conversely, respondents assigned to one of the experimental groups with a restricting definition were expected to report, on average, lower numbers, frequencies, or amounts as compared to respondents in the control group. In addition, differences between the three experimental groups were expected due to the varying positions of clarification features. According to results reported by Kunz and Fuchs (2012), definitions supporting the question comprehension were expected to be most effective when positioned before the question stem and should yield, on average, significantly higher frequencies in the case of an extending definition, and significantly lower frequencies in the case of a restricting definition as compared to presenting the definitions after the question stem or after the answer box.

The stage of information retrieval (stage two of the question-answer process) was addressed by implementing either retrieval cues or motivating statements. Retrieval cues comprised examples for activating the memory search process, while motivating statements requested respondents to exactly remember all relevant incidences which was necessary to answer the question. Special attention was paid to ensure a broad range of examples to support an exhaustive retrieval of all relevant information and to avoid limited answers due to a selective choice of retrieval cues. Both types of clarification features aimed at enhancing the recall of relevant information in numeric and narrative open-ended questions. Respondents assigned to retrieval cues as well as motivating statements were expected to report, on average, a higher number of incidences compared to respondents in the control group. Retrieval cues were used in numeric as well as narrative open-ended questions and motivating statements solely in narrative open-ended questions. Since clarification features enhancing the retrieval of relevant information were expected to be best positioned after the question stem (Kunz & Fuchs 2012), respondents receiving retrieval cues or motivating statements after the question stem should report, on average, a significantly higher number of incidences compared to respondents receiving retrieval cues or motivating statements before the question stem or after the answer box.

Formatting instructions for numeric and narrative open-ended questions were implemented to address response formatting (stage four of the question-answer process). Formatting instructions for numeric open-ended questions requested a certain format for dates or durations. Respondents receiving such a formatting instruction in one of the experimental groups should be more likely to format their answers in the desired format as compared to respondents in the control group. Formatting instructions for narrative open-ended questions requested the respondents to provide their answer in as much detail as possible. In comparison to the control group, respondents in the experimental groups were expected to elaborate on their answers and thus produce longer responses. In addition, the position after the answer box was assumed to be more effective for formatting instructions, resulting in a larger share of correctly-formatted responses and in a higher average number of characters as if formatting instructions had been presented before or after the question stem.

Table 2

*Types of clarification features addressing different stages of the question-answer process and dependent variables*

Stage	Clarification feature type	Open-ended question type	Clarification feature wording	Dependent variable and expected direction of effect
Comprehension	Extending definition	numeric	(«Please include...»)	Increased reported frequency
	Restricting definition	numeric	(«Please exclude...»)	Reduced reported frequency
Retrieval	Retrieval cues	numeric & narrative	(«Please also consider the study counseling office, the student council,...»)	Increased reported incidences
	Motivating statements	narrative	(«Please try to remember exactly...»)	Increased reported incidences
Formatting	Formatting instructions	numeric	(«Please provide the answer in the following format hh:mm.»)	More-correctly formatted answers
	Formatting instructions	narrative	(«Please report your answer as detailed as possible.»)	Higher number of characters

As aforementioned, the experimental questions for testing the use and positioning of clarification features were included in three Web surveys. In the 2012 Survey, each clarification feature type was implemented twice, based on two distinct experimental questions, respectively. Accordingly, a distinction is made between Study 1a and 1b, hereinafter, although both studies were included in the same survey. In the 2013 Survey and 2014 Survey, each clarification feature type was tested using only one experimental question, respectively. In the following, Study 2 refers to the set of experimental questions implemented in the 2013 Survey, and Study 3 refers to the questions implemented in the 2014 Survey. The exact wording of the experimental questions and clarification features is provided in the appendix.

## Results

### Descriptive analyses

**Stage I: Comprehension.** Clarification features in terms of definitions of key terms and concepts were provided in order to improve the respondents' understanding of the question meaning. Extending definitions were implemented in numeric open-ended questions asking for the amount of time (in hours) spent on school activities or on communicating with schoolmates. In general, the mean number of hours reported by respondents being assigned to one of the experimental groups was assumed to be higher than in the control group which, in fact, could be shown in all four studies (Study 1a:  $F(1, 3,294) = 420.82, p < .001, \eta^2 = .113$ ; Study 1b:  $F(1, 3,164) = 48.17, p < .001, \eta^2 = .015$ ; Study 2:  $F(1, 3,833) = 485.78, p < .001, \eta^2 = .113$ ; Study 3:  $F(1, 2,773) = 285.55, p < .001, \eta^2 = .093$ ).

Concerning the various positions of an extending definition, either provided before the question stem (EGa), after the question stem (EGb), or after the answer box (EGc), an extending definition was expected to be most effective when positioned before the question stem. However, contrary to prior expectations, results of Bonferroni post-hoc tests showed that placing the extending definition before the question stem (EGa) was least effective, since in all four studies the mean number of hours reported by respondents was significantly lower as if the extending definition was placed after the question stem (EGb) or after the answer box (EGc). No significant differences were found between the two positions after the question stem (EGb) and after the answer box (EGc) (see Table 3).

Furthermore, item nonresponse was analyzed as a common indicator of the extent of respondent burden. In the present surveys, respondents could easily skip questions without being prompted to provide an answer. In this regard, item nonresponse referred to the proportion of respondents who provided no answer to a respective open-ended question. Higher item nonresponse rates in the experimental groups compared to the control group might indicate a higher respondent burden due to the necessity to read and process the additional clarifying information. However, findings indicated that item nonresponse rates in the three experimental groups receiving an extending definition were significantly lower compared to the control group where no definition was provided (Study 1a:  $\chi^2(1, 3,593) = 41.84$ ,  $p < .001$ ; Study 1b:  $\chi^2(1, 3,592) = 23.97$ ,  $p < .001$ ; Study 2:  $\chi^2(1, 4,169) = 24.38$ ,  $p < .001$ ; Study 3:  $\chi^2(1, 3,078) = 39.92$ ,  $p < .001$ ). No significant differences were found between the three experimental groups.

Table 3

*Mean number of hours reported by respondents, depending on the position of an extending definition in numeric open-ended questions*

Study	Position of extending definition			
	no definition (CG)	before question (EGa)	after question (EGb)	after answer box (EGc)
1a	22 <sup>a</sup>	33 <sup>b</sup>	42 <sup>c</sup>	41 <sup>c</sup>
1b	21 <sup>a</sup>	25 <sup>b</sup>	29 <sup>c</sup>	30 <sup>c</sup>
2	24 <sup>a</sup>	32 <sup>b</sup>	40 <sup>c</sup>	40 <sup>c</sup>
3	25 <sup>a</sup>	33 <sup>b</sup>	39 <sup>c</sup>	39 <sup>c</sup>

*Note.* Pairwise comparisons between the experimental conditions using the Bonferroni correction: if a pair of values is significantly different at the .05 level, the values have different superscript letters assigned to them. Results of overall *F*-tests are presented in the text, comparing the control group with the three experimental groups taken all together. Cases with unusually long response times equal to or above 7,200 seconds (session timeout exceeded on the target page comprising the experimental questions) were excluded from the analyses. Outliers were excluded at two standard deviations above the group mean.

Restricting definitions were implemented in numeric open-ended questions asking for the number of friends or the amount of time (in hours) spent on computer and Internet usage. In general, the frequencies reported by respondents were assumed to decrease on average when restricting definitions were provided in the experimental groups compared to the control group which could actually be shown

in Studies 1a, 1b, and 2 (Study 1a:  $F(1, 5,406) = 47.01, p < .001, \eta^2 = .009$ ; Study 1b:  $F(1, 3,297) = 211.21, p < .001, \eta^2 = .060$ ; Study 2:  $F(1, 6,436) = 28.755, p < .001, \eta^2 = .004$ ). In Study 3, however, no significant differences were found, depending on whether a restricting definition was provided or not.

The optimal position of a restricting definition was expected to be before the question stem. Contrary to expectations, results of the Bonferroni post-hoc tests indicated that a restricting definition provided after the question stem (EGb) was more effective with respect to a significant decrease of the mean number of friends or hours reported by the respondents than in the case of providing it before the question stem (EGa) in Studies 1a, 1b, and 2. Whereas, in Study 2, a restricting definition placed after the question stem (EGb) was more effective than when placed after the answer box (EGc), no significant differences between these two positions were found in Studies 1a and 1b. Furthermore, placing a restricting definition after the answer box (EGc) or before the question stem (EGa) yielded no significant differences in Studies 1a, 1b, and 2. In Study 3, no significant differences between the three experimental conditions were found, depending on the respective position of a restricting definition (see Table 4).

Findings on item nonresponse showed no significant differences between the control group and the experimental groups for the majority of experimental questions, and the one significant difference in Study 1a showed a lower item nonresponse rate for the three experimental groups than for the control group (Study 1a:  $\chi^2(1, 5,972) = 17.14, p < .001$ ). Furthermore, no significant differences were found between the three experimental groups.

Table 4

*Mean number of hours or friends reported by respondents, depending on the position of a restricting definition in numeric open-ended questions*

Study	Position of restricting definition			
	no definition (CG)	before question (EGa)	after question (EGb)	after answer box (EGc)
1a	19 <sup>a</sup>	15 <sup>b</sup>	13 <sup>c</sup>	13 <sup>b,c</sup>
1b	16 <sup>a</sup>	10 <sup>b</sup>	9 <sup>c</sup>	10 <sup>b,c</sup>
2	17 <sup>a</sup>	16 <sup>b</sup>	15 <sup>c</sup>	16 <sup>b</sup>
3	18 <sup>a</sup>	18 <sup>a</sup>	17 <sup>a</sup>	18 <sup>a</sup>

*Note.* Pairwise comparisons between the experimental conditions using the Bonferroni correction: if a pair of values is significantly different at the .05 level, the values have different superscript letters assigned to them. Results of overall  $F$ -tests are presented in the text, comparing the control group with the three experimental groups taken all together. Cases with unusually long response times equal to or above 7,200 seconds (session timeout exceeded on the target page comprising the experimental questions) were excluded from the analyses. Outliers were excluded at two standard deviations above the group mean.

**Stage II: Retrieval.** Clarification features in terms of retrieval cues and motivating statements were implemented to facilitate and improve the retrieval of relevant information needed to answer survey questions. Retrieval cues were provided in numeric open-ended questions asking respondents for the frequency of physical impairments or the number of information sources on studying.

Retrieval cues in narrative open-ended questions asked respondents for the kind of information sources on studying or the kind of current challenges, with the number of indications being re-coded in the final analysis to numeric responses in order to enable direct comparison with the results of Studies 1a and 1b. Prior expectations concerning the use of retrieval cues could be confirmed, since the mean number of incidences reported by respondents being faced with retrieval cues in one of the experimental groups was significantly higher in all four studies compared to the control group, where no retrieval cues were provided (Study 1a:  $F(1, 2,651) = 70.70, p < .001, \eta^2 = .026$ ; Study 1b:  $F(1, 5,317) = 47.51, p < .001, \eta^2 = .009$ ; Study 2:  $F(1, 5,590) = 181.10, p < .001, \eta^2 = .031$ ; Study 3:  $F(1, 2,321) = 77.80, p < .001, \eta^2 = .032$ ).

Retrieval cues were expected to be most effective when they were placed after the question stem. In each of the four studies, results of the Bonferroni post-hoc tests indicated that respondents actually reported a significantly higher mean number of incidences when retrieval cues were presented after the question stem (EGb) compared to respondents receiving the retrieval cues before the question stem (EGa). Placing retrieval cues after the answer box (EGc) was also more effective than placing them before the question stem (EGa), and just as effective as presenting them after the question stem (EGb), as shown in Studies 1a, 1b, and 3 (see Table 5).

Findings on item nonresponse showed no significant differences between the control group and the experimental groups, except for Study 2. In Study 2 the item nonresponse rate was significantly higher for the experimental groups providing retrieval cues than for the control group, which did not show any retrieval cues (Study 2:  $\chi^2(1, 7,391) = 6.96, p < .01$ ). With regard to the three experimental groups, only a few significant differences were found, but which revealed no consistent pattern.

Table 5

*Mean number of incidences reported by respondents, depending on the position of retrieval cues in numeric and narrative open-ended questions*

Study	Position of retrieval cues			
	no retrieval cues (CG)	before question (EGa)	after question (EGb)	after answer box (EGc)
1a	0.4 <sup>a</sup>	1.2 <sup>b</sup>	2.2 <sup>c</sup>	2.4 <sup>c</sup>
1b	2.6 <sup>a</sup>	2.8 <sup>b</sup>	3.1 <sup>c</sup>	3.0 <sup>c</sup>
2	1.4 <sup>a</sup>	1.7 <sup>b</sup>	1.9 <sup>c</sup>	1.7 <sup>b</sup>
3	1.9 <sup>a</sup>	2.2 <sup>b</sup>	2.7 <sup>c</sup>	2.7 <sup>c</sup>

*Note.* Pairwise comparisons between the experimental conditions using the Bonferroni correction: if a pair of values is significantly different at the .05 level, the values have different superscript letters assigned to them. Results of overall *F*-tests are presented in the text, comparing the control group with the three experimental groups taken all together. Cases with unusually long response times equal to or above 7,200 seconds (session timeout exceeded on the target page comprising the experimental questions) were excluded from the analyses. Outliers were excluded at two standard deviations above the group mean.

Motivating statements requesting respondents to exactly remember all relevant incidences and take them into account when answering a survey question were implemented in narrative open-ended questions asking respondents to name reasons for their choice of study program or situations of

stress during the application process. It was assumed that asking respondents to carefully consider all relevant incidences will enhance the retrieval of relevant incidences which is why respondents receiving a motivating statement in one of the experimental groups were expected to list, on average, more incidences. In fact, in all four studies, respondents in the experimental groups providing motivating statements reported, on average, a significantly higher number of incidences than respondents in the control group with no motivating statement presented (Study 1a:  $F(1, 4,679) = 29.38, p < .001, \eta^2 = .006$ ; Study 1b:  $F(1, 4,012) = 4.64, p < .05, \eta^2 = .001$ ; Study 2:  $F(1, 4,002) = 48.23, p < .001, \eta^2 = .012$ ; Study 3:  $F(1, 3,144) = 52.38, p < .001, \eta^2 = .016$ ).

Motivating statements were expected to be best positioned after the question stem. In line with prior expectations, results of Bonferroni post-hoc tests indicated that a motivating statement yielded a stronger effect in terms of a significantly higher mean number of incidences when being presented after the question stem (EGb) compared to before the question stem (EGa) in Studies 1a, 1b and 2. Study 3 did not show any significant differences between the experimental groups. In Study 1a, the position after the answer box (EGc) was just as effective as after the question stem (EGb). In Study 1b, the effect of the position after the answer box (EGc) did not differ significantly from the other two positions and in Study 2 the position after the answer box (EGc) was just as least-effective as before the question (EGa) (see Table 6).

With one exception, findings on item nonresponse showed no significant differences between the control group and the experimental groups. In Study 1a, the item nonresponse rate was significantly lower for the experimental groups than for the control group (Study 1a:  $\chi^2(1, 5,631) = 5.99, p < .05$ ). With regard to the three experimental groups, only a few significant differences were found, indicating that item nonresponse rates were lower when the motivating statement was provided before the question stem (EGa) compared to after the question stem (EGb).

Table 6  
*Mean number of incidences reported by respondents, depending on the position of a motivating statement in narrative open-ended question*

Study	Position of motivating statements			
	no statement (CG)	before question (EGa)	after question (EGb)	after answer box (EGc)
1a	1.6 <sup>a</sup>	1.7 <sup>a</sup>	1.8 <sup>b</sup>	1.8 <sup>b</sup>
1b	0.8 <sup>a</sup>	0.8 <sup>a</sup>	0.8 <sup>b</sup>	0.8 <sup>a,b</sup>
2	0.5 <sup>a</sup>	0.5 <sup>a,c</sup>	0.8 <sup>b</sup>	0.6 <sup>c</sup>
3	0.6 <sup>a</sup>	0.8 <sup>b</sup>	0.8 <sup>b</sup>	0.8 <sup>b</sup>

*Note.* Pairwise comparisons between the experimental conditions using the Bonferroni correction: if a pair of values is significantly different at the .05 level, the values have different superscript letters assigned to them. Results of overall  $F$ -tests are presented in the text, comparing the control group with the three experimental groups taken all together. Cases with unusually long response times equal to or above 7,200 seconds (session timeout exceeded on the target page comprising the experimental questions) were excluded from the analyses. Outliers were excluded at two standard deviations above the group mean.

**Stage IV: Formatting.** Clarification features in terms of formatting instructions were implemented in numeric open-ended questions that were asking for the amount of time (hh:mm) spent on extracurricular activities or the date of the respondents' decision to study (mm.yyyy). Specifying the response format should help respondents report their answers in the desired format. In all four studies, a significantly higher percentage of correctly formatted responses was actually found when respondents received a formatting instruction in one of the experimental groups compared to the control group where no formatting instruction was used (Study 1a:  $\chi^2(1, 5,329) = 921.99, p < .001$ ; Study 1b:  $\chi^2(1, 5,650) = 2,141.78, p < .001$ ; Study 2:  $\chi^2(1, 7,098) = 1,633.33, p < .001$ ; Study 3:  $\chi^2(1, 5,730) = 837.07, p < .001$ ).

Concerning formatting instructions which specified the desired response format in numeric open-ended questions, the position after the answer box was expected to be the optimal position. In fact, presenting the formatting instruction after the answer box (EGc) resulted in a significantly higher percentage of answers that were provided in the desired response format, compared to presenting the formatting instruction before the question stem (EGa) in all four studies; as well as compared to presenting the formatting instruction after the question stem (EGb) in Study 2. By contrast, formatting instructions were least effective in terms of a significantly lower percentage of correctly formatted answers when presented before the question stem (EGa) compared to the other two experimental groups in all four studies (see Table 7).

Findings on item nonresponse showed no significant differences between the control group and the experimental groups for the majority of experimental questions. In Study 1b the item nonresponse rate was significantly lower for the experimental groups than for the control group (Study 1b:  $\chi^2(1, 5,972) = 8.27, p < .01$ ). With regard to the three experimental groups, no significant differences were found.

Table 7

*Percentage of correctly formatted answers reported by respondents, depending on the position of a formatting instruction in numeric open-ended questions*

Study	Position of formatting instruction			
	no instruction (CG)	before question (EGa)	after question (EGb)	after answer box (EGc)
1a	0 <sup>a</sup>	32 <sup>b</sup>	55 <sup>c</sup>	49 <sup>d</sup>
1b	3 <sup>a</sup>	69 <sup>b</sup>	77 <sup>c</sup>	77 <sup>c</sup>
2	24 <sup>a</sup>	65 <sup>b</sup>	81 <sup>c</sup>	86 <sup>d</sup>
3	29 <sup>a</sup>	67 <sup>b</sup>	82 <sup>c</sup>	85 <sup>c</sup>

*Note.* Pairwise comparisons between the experimental conditions using the Bonferroni correction: if a pair of values is significantly different at the .05 level, the values have different superscript letters assigned to them. Results of overall  $\chi^2$ -tests are presented in the text, comparing the control group with the three experimental groups taken all together. Cases with unusually long response times equal to or above 7,200 seconds (session timeout exceeded on the target page comprising the experimental questions) were excluded from the analyses.

Formatting instructions which asked respondents to report their study expectations, previous achievements in school, or reasons for studying in as much detail as possible should lead respondents to elaborate on their answers and increase the amount of information they reported to narrative open-ended questions. In all four studies, respondents in the experimental groups who received a formatting instruction actually provided more detailed responses with a significantly higher mean number of characters compared to respondents in the control group (Study 1a:  $F(1, 4,229) = 222.76, p < .001, \eta^2 = .050$ ; Study 1b:  $F(1, 3,967) = 91.92, p < .001, \eta^2 = .023$ ; Study 2:  $F(1, 5,769) = 357.05, p < .001, \eta^2 = .058$ ; Study 3:  $F(1, 4,136) = 160.77, p < .001, \eta^2 = .037$ ).

Formatting instructions that requested respondents to answer a narrative open-ended question in as much detail as possible were expected to be best positioned after the answer box. Contrary to prior expectations, the position after the question stem (EGb) consistently resulted in more detailed responses with a significantly higher mean number of characters in all four studies than the position after the answer box (EGc) or before the question stem (EGa). Solely in Study 3, presenting formatting instructions after the answer box (EGc) was more effective than placing them before the question stem (EGa). In Study 2, however, presenting formatting instructions after the answer box (EGc) was less effective than presenting formatting instructions before the question stem (EGa), whereas no significant difference between these two positions was found in Studies 1a and 1b (see Table 8).

Findings revealed significantly lower item nonresponse rates in the experimental groups compared to the control group (Study 1a:  $\chi^2(1, 5,616) = 78.52, p < .001$ ; Study 1b:  $\chi^2(1, 5,608) = 30.54, p < .001$ ; Study 2:  $\chi^2(1, 7,217) = 47.41, p < .001$ ; Study 3:  $\chi^2(1, 5,839) = 45.46, p < .001$ ). Only few significant differences were found between the three experimental groups in terms of respondents being more likely to leave an open-ended narrative question unanswered when formatting instructions were provided before the question stem (EGa) compared to after the question stem (EGb).

Table 8

*Mean number of characters reported by respondents, depending on the position of a formatting instruction in narrative open-ended questions*

Study	Position of formatting instruction			
	no instruction (CG)	before question (EGa)	after question (EGb)	after answer box (EGc)
1a	114 <sup>a</sup>	164 <sup>b</sup>	189 <sup>c</sup>	176 <sup>b</sup>
1b	113 <sup>a</sup>	150 <sup>b</sup>	186 <sup>c</sup>	148 <sup>b</sup>
2	111 <sup>a</sup>	184 <sup>b</sup>	213 <sup>c</sup>	157 <sup>d</sup>
3	106 <sup>a</sup>	160 <sup>b</sup>	212 <sup>c</sup>	185 <sup>d</sup>

*Note.* Pairwise comparisons between the experimental conditions using the Bonferroni correction: if a pair of values is significantly different at the .05 level, the values have different superscript letters assigned to them. Results of overall  $F$ -tests are presented in the text, comparing the control group with the three experimental groups taken all together. Cases with unusually long response times equal to or above 7,200 seconds (session timeout exceeded on the target page comprising the experimental questions) were excluded from the analyses. Outliers were excluded at two standard deviations above the group mean.



## Multilevel analyses

Multilevel analyses were conducted in order to examine the effects of using and varying the positions of different types of clarification features on the respondents' answers, irrespective of the specific content of an experimental question. Separate multilevel analyses using the hierarchical linear model were conducted for each clarification feature type (except for formatting instructions in numeric open-ended questions), with "experimental question" being used as a level-2 identifier to demonstrate that the effects of clarification features found in the descriptive analyses were independent of the question content. In order to examine formatting instructions in numeric open-ended questions, a hierarchical logistic regression model was applied with the binary dependent variable of correctly or incorrectly formatted answers. Findings on the six different forms of clarifications features are depicted in Table 9 and are discussed in the following.

Results regarding the use of extending definitions indicated that the frequencies reported in numeric open-ended questions were significantly higher in each of the three experimental groups presenting an extending definition in various positions compared to the control group where no definition was provided (Model ED\_0). Extending definitions placed after the question stem (EGb) or after the answer box (EGc) yielded significantly higher frequencies than extending definitions before the question stem (EGa) (see Model ED\_1 and Model ED\_2). No significant differences were found between presenting extending definitions either after the question stem (EGb) or after the answer box (EGc) (see Model ED\_1).

The use of restricting definitions resulted in significantly lower frequencies reported in numeric open-ended questions in each of the experimental groups compared to the control group (Model RD\_0). Restricting definitions placed after the question stem (EGb) yielded significantly lower frequencies than restricting definitions before the question stem (EGa) or after the answer box (EGc) (Model RD\_1). No significant differences were found between placing restricting definitions before the question stem (EGa) or after the answer box (EGc) (Model RD\_2).

Findings concerning the use of retrieval cues in numeric and narrative open-ended questions indicated that the number of incidences reported by respondents was significantly increased in each of the experimental groups compared to the control group (Model RC\_0). Retrieval cues placed after the question stem (EGb) or after the answer box (EGc) yielded a significantly higher number of incidences than retrieval cues placed before the question stem (EGa) (Model RC\_1 and Model RC\_2). No significant differences were found between placing retrieval cues after the question stem (EGb) or after the answer box (EGc) (Model RC\_1).

Providing motivating statements that asked respondents to name as many incidences as possible in a narrative open-ended question significantly increased the number of reported incidences in each of the three experimental

groups compared to the control group (Model MS\_0). Motivating statements placed after the question stem (EGb) yielded a significantly higher number of incidences than motivating statements placed before the question stem (EGa) or placed after the answer box (EGc) (Model MS\_1), while offering respondents motivating statements which were placed after the answer box (EGc) resulted in a significantly higher number of incidences than placing motivating statements before the question stem (EGa) (Model MS\_2).

Concerning the use of formatting instructions in numeric open-ended questions, a hierarchical logistic regression model revealed a significantly higher probability of correctly formatted answers among respondents receiving a formatting instruction in one of the experimental groups than among respondents in the control group (Model FNU\_0). Presenting formatting instructions after the question stem (EGb) or after the answer box (EGc) significantly increased the probability of correctly formatted answers compared to placing formatting instructions before the question stem (EGa) (Model FNU\_1 and Model FNU\_2). No significant differences were found between placing formatting instructions after the question stem (EGb) or after the answer box (EGc) (Model FNU\_1).

Results on using formatting instructions in narrative open-ended questions revealed that respondents reported longer answers with a significantly higher number of characters in each of the three experimental groups compared to the control group (Model FNA\_0). Formatting instructions in narrative open-ended questions placed after the question stem (EGb) yielded longer responses with a significantly higher number of characters than formatting instructions presented before the question stem (EGa) or after the answer box (EGc) (Model FNA\_1). No significant differences were found between the positioning of formatting instructions before the question stem (EGa), and after the answer box (EGc) (Model FNA\_2).

Table 9  
*Multilevel analyses using hierarchical linear<sup>1</sup> and hierarchical logistic regression<sup>2</sup> models*

	Extending Definition <sup>1</sup>		Restricting Definition <sup>1</sup>		Retrieval Cues <sup>1</sup>	
	Model ED_0	Model ED_1	Model RD_0	Model RD_1	Model RC_0	Model RC_1
<b>Position:</b> before the question stem (CG) after the question stem (CG) after the answer box (CG)	7.761*** 14.746*** 14.454***		-2.067*** -3.211*** -2.182***		.291*** .701*** .641***	
<b>Position:</b> before the question stem (EGb) after the answer box (EGb)		-6.992*** -.265		1.150*** 1.032***		-410*** -.062
<b>Position:</b> after the answer box (EGa)		6.703***				.349***
Log likelihood	115004.593	88385.566	119992.141	94380.265	66485.519	53705.809
Variance (experimental q.)	16.431	24.127	19.254	25.881	.421	.328
Residual variance	387.218	411.854	170.436	167.028	3.693	4.496
ICC	4.1%	5.5%	10.2%	13.4%	10.2%	6.8%
		5.1%	12.6%			7.6%
	Motivating statements <sup>1</sup>		Formatting numeric <sup>2</sup>		Formatting narrative <sup>1</sup>	
	Model MS_0	Model MS_1	Model FNU_0	Model FNU_1	Model FNA_0	Model FNA_1
<b>Position:</b> before the question stem (CG) after the question stem (CG) after the answer box (CG)	.070*** .186*** .124***		2.326*** 3.123*** 3.175***		52.747*** 88.302*** 53.232***	
<b>Position:</b> before the question stem (EGb) after the answer box (EGb)		-116*** -.062***		-784*** .049		-35.596*** -35.030***
<b>Position:</b> after the answer box (EGa)		.052***				.562
Log likelihood	29529.079	23366.122	-12220.834	-10353.005	223668.055	180039.176
Variance (experimental q.)	.247	.249	.461	.386	52.878	90.589
Residual variance	.377	.394	3.290	3.290	14882.118	17001.810
ICC	39.6%	41.5%	12.3%	10.5%	0.4%	0.5%
		%	11.8%	11.8%	0.4%	0.7%

Note. <sup>1</sup> Hierarchical linear models were computed for all clarification feature types, except for the formatting instruction in numeric open-ended questions for which a hierarchical logistic regression model<sup>2</sup> was computed (0= answer is not formatted in the desired format; 1=answer is formatted in the desired format). The table shows standardized coefficients with \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ , reference category in parentheses. The variable “experimental question” was used as level-2 identifier.

### **Response times**

Within the context of Web surveys, paradata (i.e., in terms of the amount of time respondents spent on answering a survey question) are commonly used to gain deeper insights into the cognitive processing of the verbal and visual features of questionnaire design. Response times can be considered an indicator of the respondents' cognitive effort expended on processing a survey question and their susceptibility to cognitive shortcutting within the question-answer process. Thus, response times are often used as an indirect indicator of data quality (Stieger & Reips, 2010).

In favor of a better understanding of the amount of time respondents spent on answering a survey question, reading time and response time need to be distinguished (Stieger & Reips, 2010). With respect to numeric open-ended questions requiring solely short keyboard entries, a potentially longer response time for questions comprising clarification features in one of the experimental groups compared to respective questions, without any clarification features in the control group, was deemed to be due to the additional text, and thus, longer reading times when clarification features were presented to and actually processed by the respondents. In this regard, differences in the mean response times for a numeric open-ended question in one of the experimental groups versus the control group was used as an indicator of whether respondents have actually read the clarification features and incorporated their content within the question-answer process or not. With respect to narrative open-ended questions, however, considerable variations in the amount of time that respondents actually needed for typing a longer response were expected to occur due to individual differences in the respondents' ability to type, the length of a response, and the speed of their thought processes. Since no paradata were available in the present studies, enabling an approximate differentiation between the amount of time spent on reading the question and thinking about the answer on the one hand, and the amount of time spent on typing the response on the other hand, response times were not analyzed for narrative open-ended questions.

Findings on response times in numeric open-ended questions suggested that respondents assigned to one of the experimental groups comprising clarification features spent on average significantly more time on answering the questions than respondents in the control group without any clarification features (see Figure 2). No significant differences in response times were found between the three experimental groups.

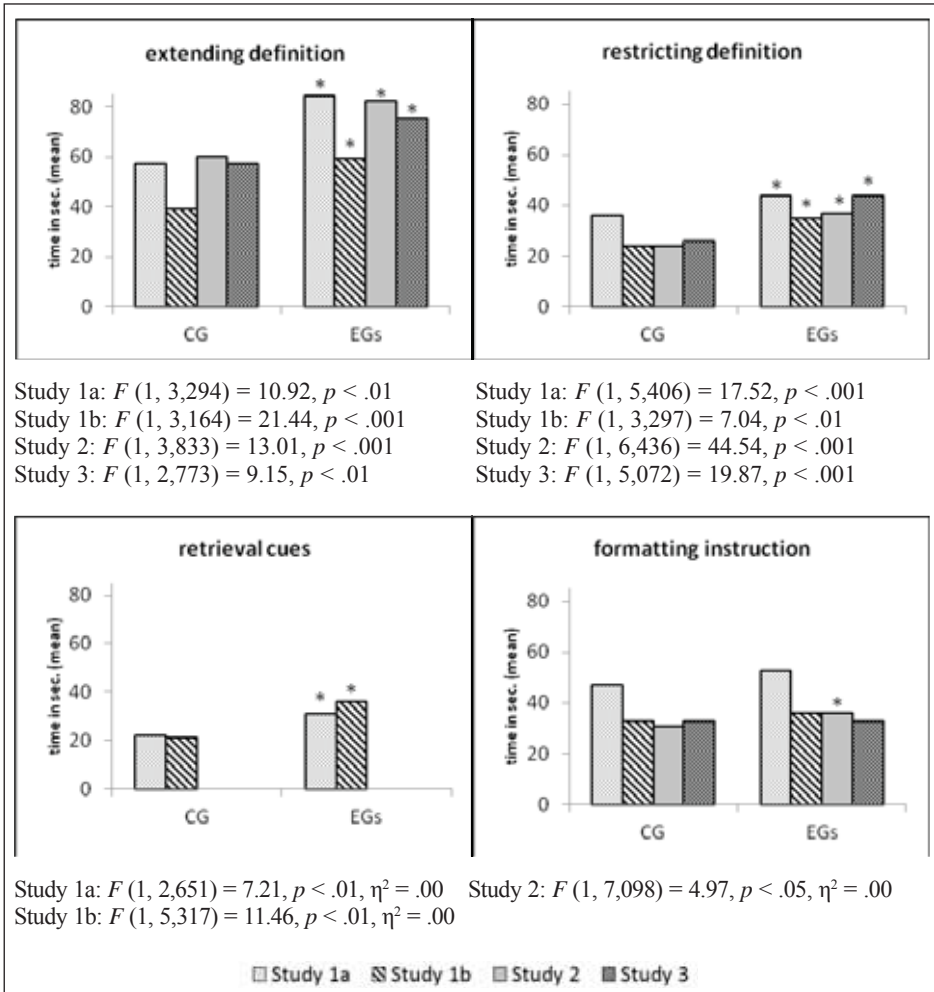


Figure 2. Mean response times (in seconds) in numeric open-ended questions, depending on the use of clarification features.

Note. \*Overall  $F$ -tests revealed significant ( $p < .05$  or less) differences between the control group as compared to the experimental groups. Bonferroni post-hoc tests revealed no significant differences between the experimental groups, which is why the three experimental groups were not presented separately. Concerning the use of retrieval cues, analyses were restricted to Study 1a and 1b, since the respective experimental questions in Studies 2 and 3 were originally asked in the form of narrative open-ended questions.

### Summary and conclusions

Clarification features in terms of definitions, retrieval cues, motivating statements, and formatting instructions are commonly provided in self-administered questionnaires in general and in Web surveys in particular in order to enhance the respondents' processing of survey questions. More precisely,

clarification features are used to prevent respondents from misinterpreting the question meaning, from prematurely abandoning the retrieval of relevant information, and from providing the answer in a format which would be undesirable to the survey researcher. Although clarification features have been proven to increase the quality of the respondents' answers to closed questions and particularly to open-ended questions, it has also been demonstrated that respondents are likely to overlook or even ignore definitions, instructions, or motivating statements (Conrad et al., 2006; Conrad et al., 2007; Redline, 2013). In the present studies, the use of clarification features and their optimal positioning in numeric and narrative open-ended questions were tested. We aimed to determine the optimal positioning of clarification features in order to maximize the respondents' attention to and thorough processing of clarification features, and thus, increase the quality of the respondents' answers to open-ended questions in Web surveys.

The findings presented in this article indicated a positive effect of clarification features on data quality in numeric and narrative open-ended questions regardless of their positioning. By providing clarification features, the respondents' understanding of the question meaning can actually be enhanced; respondents retrieve relevant information more exhaustively and answer survey questions in a more detailed manner. Thus, providing clarification features seems to foster optimizing response behavior, since respondents are better able and more motivated to go through the question-answer process thoughtfully. As indicated by largely non-significant differences in item nonresponse or by a significantly lower item nonresponse rate in the experimental groups compared to the control group, there seems to be no additional response burden due to the necessity to read and process the additional clarifying information.

The results further indicate that the effectiveness of clarification features depends on their optimal positioning relative to the other question components. Three different positions of clarification features were tested in the presented studies: before the question stem, after the question stem, and after the answer box. Definitions supporting the comprehension of the question meaning, retrieval cues and motivating statements supporting the retrieval of relevant information, and formatting instructions promoting a desired response format were most effective when they were presented after the question stem. By contrast, clarification features placed before the question stem consistently yielded the lowest effect on survey responses. Findings concerning clarification features placed after the answer box remained inconclusive. In some experimental questions, placing clarification features after the answer box yielded effects that were similar to placing them after the question stem; in other experimental questions, placing them after the answer box was less effective. According to these findings, our first and third hypotheses (H1 & H3) have to be rejected, whereas the second hypothesis (H2) can be confirmed. In summary, these findings are consistent with the established convention of presenting clarification features immediately after the question stem and before the answer box (Couper, 2008; Peytchev et al., 2010). This applies irrespective of the cognitive stage of the question-answer

process addressed by the respective clarification feature. Analyses considering multiple experiments at a time confirmed these results and further indicated that the effects of using and varying the position of different types of clarification features on the respondents' answers were irrespective of the specific content of an experimental question.

Analyses of response times showed that, irrespective of whether there are any effects of clarification features on substantive answers or not, providing clarification features in numeric open-ended questions generally increased the overall time that respondents spent on answering a respective question. These findings suggested that respondents read or at least scanned through the additional text of the clarification features regardless of their position. Thus, the lower effectiveness of clarification features positioned before the question stem or after the answer box was not due to respondents simply ignoring them at these positions. Rather, it was assumed that respondents were more likely to not only read, but actually incorporate the content of the clarification features within the question-answer process when they were placed after the question stem, since at this position, additional information was obviously provided exactly where it was needed (Dillman et al., 2014, pp. 187-189).

In a more general context, the present findings indicate that visual design in terms of the spatial arrangement of key components of a survey question can decisively influence the respondents' answers to numeric and narrative open-ended questions in Web surveys. Not surprisingly, in the present studies, significant differences were found in survey responses depending on whether clarifying information was provided or not. However, significant differences were also found depending on the positioning of clarification features. Hence, the effectiveness of clarification features can vary depending on whether they are actually presented within the respondents' navigational path, taking the conventional reading order from "left to right" and from "top to bottom" into account. Obviously, when clarification features are presented within the respondents' focus of attention and exactly when this additional information fits in the question-answer process, clarification features are more likely to be recognized by respondents as a relevant part of the survey question which, in turn, can significantly increase their effectiveness. This also suggests that with respect to designing numeric and narrative open-ended questions in Web surveys, questionnaire developers should not only ask whether or not additional clarifying information needs to be provided, but also in which position in order to achieve the best possible effects on the respondents' answers.

Generally speaking, the present results indicate that respondents are more likely to integrate clarification features into the question-answer process, when they are presented after the question stem. However, results concerning numeric open-ended questions also suggest that the position after the answer box can be similarly effective as the position after the question stem (with the exemption of restricting definitions), whereas with respect to narrative open-ended questions, findings on the position after the answer box remain inconclusive. Although further research is needed regarding this assumption, the present findings suggest

that the optimal positioning of clarification features may also depend on the type of question, i.e., whether it is a numeric or narrative open-ended question and, at least regarding the first stage of the question-answer process, on the purpose implied by a clarification feature, i.e., whether it contains a definition that restricts or extends the meaning of the survey question as compared to everyday understanding.

Finally, a key shortcoming of the reported experiments is noteworthy: The samples of the three Web surveys used in these studies were composed of university applicants. Since university applicants can be assumed to be highly motivated, they are presumably more willing to show optimizing response behaviors than respondents of a general population survey. In particular, it can be assumed that university applicants are more likely to cycle back and forth between the various stages of the question-answer process, which might be a possible explanation of why, contrary to expectations, the optimal position of clarification features did not vary depending on the respective stage of the question-answer process they refer to. In addition, university applicants are typically younger than the general population and are also highly educated. Both factors suggest that they are also more computer-literate and more experienced in dealing with Websites and Internet forms than the general population, which in turn may limit the generalizability of the results. Thus, in order to make reliable statements about the general population, the present findings on the use and positioning of clarification features need to be replicated in a sample of the general population.

## References

- Bethlehem, J. G., & Biffignandi, S. (2012). *Handbook of Web Surveys*. Hoboken: Wiley.
- Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on Interviewing Techniques. *Sociological Methodology*, 12, 389-437.
- Christian, L. M., & Dillman, D. A. (2004). The Influence of Graphical and Symbolic Language Manipulations on Responses to Self-Administered Questions. *Public Opinion Quarterly*, 68(1), 57-80.
- Christian, L. M., Dillman, D. A., & Smyth, J. D. (2005). *Instructing Web and Telephone Respondents to Report Date Answers in a Format Desired by the Surveyor. Technical Report 05-067*. Pullman, WA: University of Washington, Social and Economic Sciences Research Center.
- Christian, L. M., Dillman, D. A., & Smyth, J. D. (2007). Helping Respondents Get It Right the First Time: The Influence of Words, Symbols, and Graphics in Web Surveys. *Public Opinion Quarterly*, 71(1), 113-125.
- Conrad, F. G., Couper, M. P., Tourangeau, R., & Peytchev, A. (2006). Use and Non-Use of Clarification Features in Web Surveys. *Journal of Official Statistics*, 22(2), 245-269.
- Conrad, F. G., Schober, M. F., & Coiner, T. (2007). Bringing Features of Human Dialogue to Web Surveys. *Applied Cognitive Psychology*, 21(2), 165-187.
- Couper, M. P. (2008). *Designing Effective Web Surveys*. New York: Cambridge University Press.
- Couper, M. P., Kennedy, C., Conrad, F. G., & Tourangeau, R. (2011). Designing Input Fields for Non-Narrative Open-Ended Responses in Web Surveys. *Journal of Official Statistics*, 27(1), 65-85.



- Dillman, D. A., Smith, J. D., & Christian, L. M. (2014). *Internet, Phone, Mail and Mixed-Mode Surveys. The Tailored Design Method*. Hoboken, NJ: Wiley.
- Fowler, F. J. (1995). *Improving Survey Questions. Design and Evaluation*. Thousand Oaks, CA: Sage.
- Friborg, O., & Rosenvinge, J. H. (2013). A Comparison of Open-Ended and Closed Questions in the Prediction of Mental Health. *Quality & Quantity*, 47(3), 1397-1411.
- Fuchs, M. (2007). Asking for Numbers and Quantities: Visual Design Effects in Web Surveys. *JSM Proceedings of the Survey Research Methods Section* (pp.3786-3792). Alexandria, VA: American Statistical Association. Retrieved from <http://www.amstat.org/sections/srms/Proceedings/>
- Fuchs, M. (2009a). Asking for Numbers and Quantities: Visual Design Effects in Paper&Pencil Surveys. *International Journal of Public Opinion Research*, 21(1), 65-84.
- Fuchs, M. (2009b). Differences in the Visual Design Language of Paper-and-Pencil Surveys Versus Web Surveys. A Field Experimental Study on the Length of Response Fields in Open-Ended Frequency Questions. *Social Science Computer Review*, 27(2), 213-227.
- Galesic, M., Tourangeau, R., Couper, M. P., & Conrad, F. G. (2008). Eye-Tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding. *Public Opinion Quarterly*, 72(5), 892-913.
- Jenkins, C. R., & Dillman, D. A. (1997). Towards a Theory of Self-Administered Questionnaire Design. In L. E. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 165-198). New York: Wiley.
- Krosnick, J. A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5(3), 213-236.
- Kunz, T., & Fuchs, M. (2012). Positioning of Clarification Features in Web Surveys: Evidence from Eye Tracking Data. *JSM Proceedings of the Survey Research Methods Section*. Alexandria, VA: American Statistical Association. Retrieved from <http://www.amstat.org/sections/srms/Proceedings/>
- Oudejans, M., & Christian, L. M. (2011). Using Interactive Features to Motivate and Probe Responses to Open-Ended Questions. In M. Das, P. Ester & L. Kaczmarek (Eds.), *Social and Behavioral Research and the Internet* (pp. 215-244). New York: Routledge.
- Peytchev, A. (2009). Survey Breakoff. *Public Opinion Quarterly*, 73(1), 74-97.
- Peytchev, A., Conrad, F. G., Couper, M. P., & Tourangeau, R. (2010). Increasing Respondents' Use of Definitions in Web Surveys. *Journal of Official Statistics*, 26(4), 633-650.
- Peytchev, A., & Crawford, S. (2005). A Typology of Real-Time Validations in Web-Based Surveys. *Social Science Computer Review*, 23(2), 235-249.
- Redline, C. D. (2013). Clarifying Categorical Concepts in a Web Survey. *Public Opinion Quarterly*, 77(S1), 89-105.
- Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-Ended vs. Close-Ended Questions in Web Questionnaires. *Metodoloski zvezki*, 19, 159-177.
- Schwarz, N. (1997). Questionnaire Design: The Rocky Road from Concepts to Answers. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz & D. Trewin (Eds.), *Survey Measurement and Process Quality* (pp. 29-45). New York: Wiley.
- Schwarz, N., Knauper, B., Hippler, H.-J., Noelle-Neumann, E., & Clark, L. (1991). Rating Scales: Numeric Values May Change the Meaning of Scale Labels. *Public Opinion Quarterly*, 55(4), 570-582.
- Schwarz, N., & Oyserman, D. (2001). Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction. *American Journal of Evaluation*, 22(2), 127-160.
- Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-Ended Questions in Web Surveys. Can Increasing the Size of Answer Boxes and Providing Extra Verbal Instructions Improve Response Quality? *Public Opinion Quarterly*, 73(2), 325-337.

- Stieger, S., & Reips, U.-D. (2010). What are Participants Doing While Filling in an Online Questionnaire: A Paradata Collection Tool and an Empirical Study. *Computers in Human Behavior*, 26(6), 1488-1495.
- Strack, F. (1992). "Order Effects" in Survey Research: Activation and Information Functions of Preceding Questions. In N. Schwarz & S. Sudman (Eds.), *Context Effects in Social and Psychological Research* (pp. 23-34). New York: Springer Verlag.
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Josey-Bass Publishers.
- Toepoel, V., & Dillman, D. A. (2011). Words, Numbers, and Visual Heuristics in Web Surveys: Is there a Hierarchy of Importance? *Social Science Computer Review*, 29(2), 193-207.
- Tourangeau, R., Conrad, F. G., Arens, Z., Fricker, S., Lee, S., & Smith, E. (2006). Everyday Concepts and Classification Errors: Judgments of Disability and Residence. *Journal of Official Statistics*, 22(3), 385-418.
- Tourangeau, R., Conrad, F. G., Couper, M. P., & Ye, C. (2014). The Effects of Providing Examples in Survey Questions. *Public Opinion Quarterly*, 78(1), 100-125.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

## Appendix

Table A1

*Wording of experimental questions and clarification features of study 1a (English translations).*

Stage	Clarification feature type	Experimental question wording	Clarification feature wording
Comprehension	Extending definition	During the last school year of 2011/2012, how many hours did you spent on school-related activities in a typically school week?	By school-related activities we mean presence in class, doing homework, preparing classwork and tests, project work, preparing presentations and talks, written papers and assignments as well as the specific exchange of information with classmates.
	Restricting definition	How many people do you count as friends?	By friends we mean people, with whom you maintain personal contact and who you meet regularly (either in private or public places). Do not include people, with whom you only keep in contact over the Web (e.g., via email or social networks such as Facebook and StudiVZ).
Retrieval	Retrieval cues	How often did you suffer from symptoms of physical impairments during the time of your final exams within the school year of 2011/2012?	Please think about e.g., headaches, tummy bugs, back complaints, sleep disorders, circulatory problems, angina pectoris, sustained tiredness, concentration problems, lack of appetite, sickness, strong sweating.
	Motivating statements	On a personal level, which reasons were crucial to you, in applying for studies at a university instead of pursuing a vocational education or profession?	Please try to remember exactly, and also consider (if possible) all reasons in your answer.
Formatting	Formatting instructions (narrative)	Which specific expectations do you have on your intended studies at the Darmstadt University of Technology?	This question is important. Therefore, please report your answer in as much detail as possible.
	Formatting instructions (numeric)	During the last six months how much time did you spend in a typical week on extracurricular activities such as sports, other hobbies and voluntary work?	Please report the time in this format: hh:mm.

Table A2

*Wording of experimental questions and clarification features of study 1b (English translations).*

Stage	Clarification feature type	Experimental question wording	Clarification feature wording
Comprehension	Extending definition	During the last school year of 2011/2012, how many hours did you spent on communicating with your classmates in a typical school week?	By communication with your classmates we mean the personal conversations inside and outside of school, the contact by email or text message, the communication via social networks such as Facebook and StudiVZ, Twitter, blogs or chats. Please refer to the time spent on conversations about personal topics as well as the specific exchange about educational teaching contents.
	Restricting definition	During the last school year 2011/12, how many hours did you spend on your computer and on Internet usage in a typical school week?	By computer and Internet usage we only refer to the usage because of school-related purposes. Do not include the time exposure of computer and Internet usage for exclusively private purposes.
Retrieval	Retrieval cues	How many different sources did you use prior to your application to get information about studying at the Darmstadt University of Technology?	Please consider external offers such as student career counseling at school or at the agency for labor, university rankings, study guides and the visit of an educational exhibition or of special homepages as well as internal offers of the Darmstadt University of Technology, such as the central or specialized student counseling, the homepage or special information events of the Darmstadt University of Technology, and conversations with students and professors at the Darmstadt University of Technology.
	Motivating statements	In which situations did you feel heavily burdened or even overchallenged during the last six months?	Please try to remember each situation and the approximate duration of the respective situation and also consider them in your reported answer.
Formatting	Formatting instructions (narrative)	Which personal experiences of success have you already achieved within your educational and professional career?	This question is important. Therefore, please answer it in as much detail as possible.
	Formatting instructions (numeric)	When die you decide to apply for studies at the Darmstadt University of Technology?	Please report the date in the format: mm.yyyy.

Table A3  
*Wording of experimental questions and clarification features of Study 2 (English translations)*

Stage	Clarification feature type	Experimental question wording	Clarification feature wording
Comprehension	Extending definition	How many hours did you spend on school-related activities in a typical school week during the last school year of 2012/13? (see Study 3)	By school-related activities we mean the presence in class, doing homework, preparing classwork and tests, project work, preparing presentations and talks, written papers and assignments as well as the exchange of information about classes with classmates.
	Restricting definition	How many hours do you currently spend on a computer and on Internet usage in a week? (see Study 3)	By computer and Internet usage we refer to the usage due to school or vocational education-related purposes and private purposes such as creating and editing texts, tables and presentations, writing emails, searching for information, watching videos or movies, listening to music or downloading music, reading news or getting information about current events and shopping. Do not include the time exposure spent on communication with friends via social networks such as Facebook, Google Plus and Twitter.
Retrieval	Retrieval cues	Which sources did you use prior to your application to get information about studying at the Darmstadt University of Technology?	Besides offers of the Darmstadt University of Technology such as the central or specialized student counseling, the homepage or special information events of the Darmstadt University of Technology and conversations with students and professors, please consider also other sources of information such as student career counseling at school or at the labor agency, university rankings, study guides and the visit of an educational exhibition or of special homepages.
	Motivating statements	In which situations did you feel heavily burdened or even overchallenged during your choice of studies and the application process? (see Study 3)	Please try to remember each situation and the approximately duration of the respective situation.
Formatting	Formatting instructions (narrative)	On a personal level, which reasons were crucial to you, in applying for studies at a university instead of pursuing a vocational education or profession? (see study 3)	This question is important. Therefore, please report your answer in as much detail as possible.
	Formatting instructions (numeric)	When did you decide to apply for studies at the Darmstadt University of Technology? (see Study 3)	Please report the date in the format: mm.yyyy (e.g., 01.2013).

Table A4

*Wording of experimental questions and clarification features of Study 3 (English translations)*

Stage	Clarification feature type	Experimental question wording	Clarification feature wording
Comprehension	Extending definition	How many hours did you spent on school-related activities in a typical school week during the last school year of 2013/14? (see Study 2)	By school-related activities we mean your presence in class, doing homework, preparing classwork and tests, project work, preparing presentations and talks, written papers and assignments as well as the exchange of information about class with classmates.
	Restricting definition	How many hours do you currently spend on a computer and on Internet usage in a week? (see Study 2)	By computer and Internet usage we refer to the usage due to school or vocational education-related purposes and private purposes such as creating and editing texts, tables and presentations, writing emails, searching for information, watching videos or movies, listening to music or downloading music, reading news or getting information about current events and shopping. Do not include the time exposure spent on communication with friends via social networks such as Facebook, Google Plus and Twitter.
Retrieval	Retrieval cues	A new stage of life begins when you start your studies at a university. This new stage of life often bears new challenges. What are the main challenges for you in the very near future?	Besides the challenges of your studies at a university such as e.g., good grades, time management, and using expert knowledge, please consider also the challenges of your private life such as e.g., searching for an apartment, making and keeping new friends, doing your household chores and the compatibility of your studies with work or your studies with your free time.
	Motivating statements	In which situations did you feel heavily burdened or even overchallenged during your choice of studies and the application process? (see Study 2)	Please try to remember each situation and the approximate duration of each respective situation.
Formatting	Formatting instructions (narrative)	On a personal level, which reasons were crucial to you in applying for studies at a university instead of pursuing a vocational education or profession? (see Study 2)	This question is important. Therefore, please report your answer in as much detail as possible.
	Formatting instructions (numeric)	When did you decide to apply for studies at the Darmstadt University of Technology? (see Study 2)	Please report the date in the format: mm.yyyy (e.g., 01.2014).