Full length article

The impact of visual design and response formats on data quality in a web survey of MOOC students

CrossMark

Natalia Maloshonok [a, *], Evgeniy Terentev [b]

[a] The Institute of Education, National Research University Higher School of Economics (HSE), Moscow, Potapovskiy Alley, 16 Building 10, 101000, Russian Federation
[b] Centre for Institutional Research, National Research University Higher School of Economics (HSE), Moscow, Malaya Pionerskaya st. 12, 115054, Russian Federation

ARTICLE INFO

ABSTRACT

The objective of this paper is to gain more evidence regarding how the design of the rating scales and open-ended questions influence data quality in Web surveys of MOOC (Massive Open Online Course) students. We present the results of four full-factorial randomized experiments that investigate the impact of the following factors: 1) *order of response options*; 2) *user interface for rating questions* 3) *layout of question's options*; and 4) *size of answer boxes in open-ended questions*. We found that responses to scalar questions with ascending (from negative to positive) or descending (from positive to negative) order of response options do not differ substantially. The use of the radio button format allows a reduction in the percentage of respondents who choose the "Don't know" option and makes responding to questions less challenging in comparison with slider and text box interfaces. There are no significant differences in the answers of respondents who completed questionnaires with a vertical or horizontal orientation of the questions' options. In addition, respondents who answer the questions with larger answer boxes are more likely to write longer comments.

© 2016 Published by Elsevier Ltd.

1. Introduction

In recent decades, Web surveys have become a popular tool for collecting information regarding students. The survey form suggests easy access to students (for example, through e-mails) and allows the gathering of information regarding social and economic characteristics, attitudes, motivation, and satisfaction with different study aspects, which cannot be derived by using other instruments. However, Web surveys do not always provide valid data regarding students; this is caused by the lack of student motivation to participate in Web surveys. Chen argues that the data quality problem is particularly challenging for university students because of the rapid growth in demand for student survey data from college and university administration, governmental agencies, and marketing firms (Chen, 2011). The problem of data quality in student Web surveys can occur due to nonresponse and measurement error (Couper, 2000). The first reason concerns the research of methods

for improving the response rate by using incentives (see, for example, Zusman & Duby, 1987; Porter & Whitcomb, 2003; Fan & Yan, 2010; Veen, Goritz, & Sattler, 2015). The problem of irresponsible respondents can be partly solved by developing a highly motivated invitation to the Web survey and by improving the Web survey visual design features.

In certain research, a Web survey is the sole means to gather data regarding respondents. This is the case for surveying Massive Open Online Course students (MOOC students), where communication between the researcher and respondents is mediated by the Internet. In this paper, we consider how to improve the quality of MOOC participants' responses by manipulating certain attributes of the survey design elements. The objective of this study was to determine the features of visual design and response formats that positively affect data quality in the web surveys of MOOC students.

The following factors were examined in this research:

- Order of response options in scalar questions (from negative to positive/from positive to negative);
- User interface for rating question (radio button/slider/text box);
- Layout of question's options (horizontal/vertical);

* Corresponding author.
  E-mail addresses: nmaloshonok@hse.ru (N. Maloshonok), eterentev@hse.ru (E. Terentev).

- Size of answer boxes in open-ended questions (small answer box/big answer box).

The following criteria of data quality were employed: the share of non-substantive responses, numbers of words and characters in the open-ended questions. Moreover, we examined the effect of the different visual design and response formats in survey measurements.

2. Background

2.1. Order of response options in scalar questions

Scalar questions are likely the most widely used tool in Web surveys. With regard to MOOC students' surveys, these questions can be utilized to evaluate the participants' satisfaction with different aspects of courses. One important dilemma that arises in the process of designing such types of questions concerns the issue of option ordering (which option should be the first and which should be the last).

A number of studies on the cognitive process of responding to questions with ordinal scales in Web surveys suggested that this visual design feature has a significant influence on the behavior of respondents (Tourangeau, Couper, & Conrad, 2004, 2007, 2013). It was argued that respondents use several heuristics to interpret a question (Tourangeau et al., 2004). According to one heuristic, respondents view leftmost (in horizontally oriented list) and top (in vertically orientated list) options as first in a logical order. Other heuristics imply that respondents select leftmost (in horizontally oriented list) and top (in vertically orientated list) options as the most desirable. Consequently, the so-called primacy effect can be observed in which respondents are more prone to choose early-listed options.

The evidence of the primacy effect was shown in several empirical studies (Malhotra, 2008; Stern, Dillman, & Smyth, 2007; Toepoel, Das, & Soest 2006, 2009; Tourangeau et al., 2013). Toepoel et al. (2006, 2009) and Tourangeau et al. (2013) found that presenting response options in the negative to positive order significantly reduces the share of positive answers. Malhotra (2008) verified the findings of Toepoel, Das and Van Soest and observed the most significant primacy effects in the group of respondents who completed the questionnaire the fastest and among low-educated respondents. Stern et al. (2007) found that respondents are more likely to choose non-substantive options ("Don't know") when a scale begins with negative options. The researchers' experiment also showed that the use of a descending (from positive to negative) format of scale increases the number of respondents who choose neutral response categories.

However, certain research found no evidence of a primacy effect on data quality (Christian, Dillman, & Smyth, 2008; Christian, Parsons, & Dillman, 2009). At the same time, Christian et al. (2009) showed that respondents respond more quickly when positive options are presented first.

Consequently, there are contradictory inferences regarding the effect of options ordering in scalar questions that require clarification. Also in most experiments the effect of options ordering was examined for separately presented (one per screen) and vertically oriented radio button questions. Further research should be done to understand whether the order of options can affect the data quality for other types of rating questions' layout. In this research we examine the effect of options ordering for horizontally oriented separately presented radio button questions and matrix tables.

2.2. User interface for rating questions

One important advantage of Web surveys over paper-and-pencil surveys is an opportunity to visualize questions differently. Accordingly, slider scales are the new format that cannot be implemented in paper-based surveys but have become popular in Web surveys. Slider scales are considered less repetitive and more involved for respondents compared with traditional radio button user interfaces (Roster, Lucianetti, & Albaum, 2015). However, it is not clear whether the slider response format for questions may yield similar results as traditional question formats. Previous experiments with different user interfaces demonstrate inconsistent results regarding the effect of response formats on the quality of data derived from rating questions. Arnau, Thompson, and Cook (2001) investigated how slider and radio button user interfaces influence the latent structure of responses. The researchers found that both formats yield similar latent structures of responses. Similar results were observed in the experiment of Roster et al. (2015).

Bayer and Thomas (2004) explored the response rate and the average time that respondents spend answering questions with three user interfaces: sliders, radio buttons and text answer boxes for numbers. In contrast to the previous findings, the researchers showed that the slider format contributes to increasing the time-for-question and the dropout rate. The high dropout rate was caused by respondents using Internet browsers without Java, which thus could not load sliders. Cook, Heath, Thompson, and Thompson (2001) compared different types of sliders with the radio button format. Researchers found that both types do tend to yield reliable scores. However, a slider user interface is more time consuming for respondents.

Funke, Reips, and Thomas (2011) observed that problems with slider interfaces more often occur among respondents with less than average education. Moreover, Funke and Toepoel (2015) found that the slider user interface leads to increasing item nonresponse. Funke (2015) recommended avoiding slider scales, because they increase response time and negatively influence response rate and composition of sample. Waltson, Lissitz, & Rudner (2006) also found the negative effect of sliders on response rates.

The other traditional format of questions is the text answer box. This user interface was compared with the radio button format by Couper, Traugott, and Lamias (2001). The researchers concluded that respondents who were presented with the text answer box format more often skipped the question without answering or chose the "Don't know" option.

Therefore, previous studies on the effect of slider types of questions on the data quality compared with more traditional types have yielded mixed findings, and further research is required.

2.3. Layout of question's options

There are various means by which to organize a question's options: linear (vertical/one column, or horizontal/one row) and nonlinear (several rows and columns). Toepoel et al. (2009) found that the use of a linear versus nonlinear layout leads to statistically significant differences in responses. Researchers hypothesized that under nonlinear experimental conditions, respondents focus differently on the options depending on their place in a row and column (Toepoel et al., 2009). Similar results were observed in other studies (see, for example, Christian, 2003). However, a few studies were conducted to compare two types of linear layout: horizontal and vertical. Funke et al. (2011) showed that the layout of a question's options had no significant effect on respondent's answers under radio buttons and slider user interfaces conditions. In this article, we plan to replicate this experiment to ensure that

the employed horizontal or vertical linear layout of a question's options does not affect the data quality.

2.4. Size of answer box in open-ended narrative questions

Responses to open-ended narrative questions (for example, general comments regarding the online learning course) can be influenced by certain visual design features. Thus, previous studies showed that the size of the answer box in open-ended narrative questions has a significant impact on data quality in Web surveys (Emde & Fuchs, 2012; Zuell, Menold, Körber, & 2015). Emde and Fuchs (2012) compared the respondents' answers to questions with small, large and dynamically sized[1] answer boxes and found an increase in the length of responses for large boxes. At the same time, the use of large answer boxes did not increase item nonresponse. Such findings verified the results of experiments conducted on paper-based surveys (Christian, & Dillman 2004; Israel 2006, 2010). In contrast, Zuell et al. (2015) showed a significant decrease in the number of responses to open-ended questions with large answer boxes.

However, certain experiments did not show significant differences between responses to open-ended questions with different answer box sizes (Behr, Bandilla, Kaczmirek, & Braun, 2014; Smyth, Dillman, Christian, & McBride, 2009; Stern, Smyth, & Mendez 2012). Stern et al. (2007) showed a significant effect of answer size boxes solely for respondents with less than a college degree and for women. Smyth et al. (2009) found an effect of the answer box size solely for late respondents. Stern et al. (2012) reported an increase in the mean number of words that were written by respondents in large answer boxes; however, the differences were not statistically significant. However, there is no research on the effect of answer box size for questionnaires in different languages. In this paper, we examined the impact of answer boxes with different sizes on the number of substantive answers and the amount of words and characters provided in English and Russian languages.

3. Method and data

Four full-factorial randomized experiments were conducted to explore the impact of visual design features and the questions' formats on the data quality. Experiments were conducted in the post-course surveys for nine online-courses, which were launched by the National Research University Higher School of Economics

were taught in English, and six were presented in Russian. Accordingly, post-course Web surveys for the courses in English were in English and for the courses in Russian were in Russian. Surveys for all considered online courses consisted of nearly identical questions and had similar structure. According to Fisher (1925, 1935), the differences in student characteristics were controlled by implementing the procedure of randomization in each experiment.

The number of respondents varied from 128 to 635 for different MOOCs. The overall sample has 3088 students. The invitation to the survey and personalized links to the questionnaire were emailed to MOOC students by the Coursera.org mailing system. Two reminders were sent during the two weeks after the survey launched. Response rates ranged between 0.7% and 5.3% among students who signed in during an online course.[2] The Kinesis software platform was used to administer the surveys.

The detailed descriptions of each experiment are below.

Experiment 1. Order of response options (from negative to positive/ from positive to negative)

In this experiment, we tested the effect of option ordering for two types of layouts (vertical and horizontal) and for two question formats (matrix and radio buttons). It was hypothesized that respondents are more likely to choose positive options in scalar questions when these are presented on the left in a horizontally oriented list and on the top in a vertically orientated list.

In the first case, the responses to three questions regarding the overall estimation of the courses (*How likely would you be to revisit the course materials in the future?; How likely would you be to take this course again?; How likely would you be to recommend this course to a friend?*) were examined. Respondents were randomly assigned to one of four formats for each question: (1) vertically oriented radio buttons with options listed from negative to positive; (2) vertically oriented radio buttons with options listed from positive to negative; (3) horizontally oriented radio buttons with options listed from negative to positive; and (4) horizontally oriented radio buttons with options listed from positive to negative. Questions were measured on a four-pointed scale (very unlikely, somewhat unlikely, somewhat likely, very likely) with the "Don't know" and "I wasn't doing it" options in the rightmost area of a horizontally oriented format and on the bottom in a vertically oriented format (Fig. 1).

In the second case, five questions concerned with the evaluation of different types of activity in which respondents participated

Table 1
Description of the Coursera's online courses and the sample.

Title of the online course	Language	Number of responses	Response rate (RR2)
Core Concepts in Data Analysis	English	415	0.7%
Introduction to LaTeX	Russian	635	5.3%
Public Economics	English	249	1.0%
Understanding Russians: Contexts of Intercultural Communication	English	246	1.4%
Institutional Economics	Russian	393	4.6%
History of Economic Thought	Russian	128	0.9%
Fundamentals of Corporate Finance	Russian	529	2.2%
Microeconomics Principles	Russian	321	1.6%
Industrial Organization	Russian	172	2.0%

Note. Response rate (RR2) is accounted according to AAPOR standard definitions (AAPOR 2008).

(HSE) on Coursera.org in 2014–2015 (see Table 1). Three courses

---

[1] In this experiment, dynamically sized answer boxes changed size after 84 typed characters.

[2] Extremely low response rates can be explained by high attrition rates of MOOCs that usually range from 91% to 93%. (Xing, W., Chen, X., Stein, J., & Marcinkowski 2016).

How likely would you be to revisit the course materials in the future?

○ Very unlikely
○ Somewhat unlikely
○ Somewhat likely
○ Very likely
○ Don't know

How likely would you be to revisit the course materials in the future?

○ Very likely
○ Somewhat likely
○ Somewhat unlikely
○ Very unlikely
○ Don't know

How likely would you be to revisit the course materials in the future?

○ Very unlikely ○ Somewhat unlikely ○ Somewhat likely ○ Very likely ○ Don't know

How likely would you be to revisit the course materials in the future?

○ Very likely ○ Somewhat likely ○ Somewhat unlikely ○ Very unlikely ○ Don't know

Fig. 1. Screenshots of Experimental Manipulation 1 (1): Vertically Oriented Radio Buttons with Options Listed from Negative to Positive (top) versus Vertically Oriented Radio Buttons with Options Listed from Positive to Negative (2 from the top) versus Horizontally Oriented Radio Buttons with Options Listed from Negative to Positive (3 from the top) versus Horizontally Oriented Radio Buttons with Options Listed from Positive to Negative (bottom).

during the course (*watching video lectures, reading additional course materials, completing quizzes, reading postings/asking questions at the discussion forum, reading postings/communicating with other course participants at social networks*), were randomly presented in one of four formats: (1) matrix table with options listed from negative to positive; (2) matrix table with options listed from positive to negative; (3) horizontally oriented radio buttons (one item per screen) with options listed from negative to positive; and (4) horizontally oriented radio buttons (one item per screen) with options listed from positive to negative. Questions were measured on a four-pointed unipolar scale (absolutely useless, somewhat useless, somewhat useful, very useful) with the "Don't know" option on the far right (Fig. 2).

Experiment 2. User interface for rating questions (radio buttons vs slider and text box)

In the second experiment, the question "*How many hours per week did you spend on this course?*"[3] was randomly presented by using one of three user interfaces: (1) radio buttons, (2) slider and (3) text answer box for numbers (refer to Fig. 3). In previous experiments, we found no significant difference in the response patterns for two user interfaces: slider and text answer box for numbers (Maloshonok & Terentev, 2014). However, we have observed that both interfaces lead to a large share of respondents who choose the "Don't know" option. Accordingly, we generated a hypothesis that the slider and text box user interfaces are more challenging for respondents. Therefore, this experiment has the

goal to assess the impact of different user interfaces on the resulting data quality.

Experiment 3. Layout (horizontal vs vertical)

In the third experiment, we manipulated the spatial orientation of the questions' options on the screen. The randomly assigned half of respondents viewed seven questions in a horizontal layout, and the other half viewed them in a vertical orientation. This experiment was conducted for the following rating questions: "*Please rate the difficulty of the course*", "*Please rate the course workload*", "*Please rate the course pacing*", "*How would you rate your understanding of the subject matter after taking this course?*", "*How likely would you be to revisit the course materials in the future?*", "*How likely would you be to take this course again?*", "*How likely would you be to recommend this course to a friend?*". For both experimental conditions, each question was displayed on the separate screen (one question per screen). It was hypothesized that respondents are not affected by difference in the linear layout of a question's options.

Experiment 4. Size of answer box in open-ended narrative questions

In the fourth experiment, we tested the hypothesis that providing large answer boxes in narrative open-ended questions increases the length of responses and does not influence item nonresponse. Respondents were randomly assigned to one of two versions of the open-ended questions "*What did you like most about the course?*", "*What did you like least about the course?*" Both questions were displayed on one screen. The size of the large box was 6 rows with 80 characters each. The size of the small box was 2

---

[3] This question required answer about frequency (amount of hours per week).

How useful were the following types of activity for studying the course?

	Absolutely useless	Somewhat useless	Somewhat useful	Very useful	I wasn't doing it	Don't know
Watching video lectures	○	○	○	○	○	○
Reading additional course materials	○	○	○	○	○	○
Completing quizzes	○	○	○	○	○	○
Reading postings/asking questions at the discussion forum	○	○	○	○	○	○
Reading postings/communicating with other course participants at social networks	○	○	○	○	○	○

How useful were the following types of activity for studying the course?

	Very useful	Somewhat useful	Somewhat useless	Absolutely useless	I wasn't doing it	Don't know
Watching video lectures	○	○	○	○	○	○
Reading additional course materials	○	○	○	○	○	○
Completing quizzes	○	○	○	○	○	○
Reading postings/asking questions at the discussion forum	○	○	○	○	○	○
Reading postings/communicating with other course participants at social networks	○	○	○	○	○	○

Fig. 2. Screenshots of experimental manipulation 1 (2): Matrix table with options listed from negative to positive (top) versus matrix table with options listed from positive to negative (bottom).

rows with 80 characters each. Each version contained the "Don't know" option (Fig. 4).

Unfortunately, there is no full information about social and demographic characteristics of our sample of MOOC students. However, 55% of the sample respondents participated in the pre-course surveys for the same MOOCs. These surveys contained questions about gender, age, level of education and country of living. According to it, the audience of our study mainly consists of population with higher education degree (Males = 45.4%, Females = 54.6%, Mean Age = 33.8, SD (age) = 11.7, Respondents with higher education degree = 90.9%, Respondents without higher education degree = 9.1%, students from Russia = 56.4%; Students from foreign countries (including countries of Commonwealth of Independent States) = 43.6%).

4. Results

4.1. Order of response options in scalar questions

The results of the first experiment showed significant differences between the answers of respondents to scalar questions with an ascending or descending scale format only for one item for a horizontal layout (Table 2). Although a higher proportion of respondents who chose the most positive option in all three questions was observed among those who were presented with an ascendant scale format, the difference in the response structure in two out of three items is not significant at $p < 0.05$.

In the second case, we found significant differences in responses for one item in the matrix tables (*watching video lectures*) and three items in the radio buttons questions (*watching video lectures, reading additional course materials, reading postings/communicating with other course participants at social networks*) (refer to Table 3). Contrary to expectations, the findings for questions presented on separate screens indicated that the most positive option is less often chosen when it is presented on the left. Although the differences were statistically significant for three items, the two remaining questions also demonstrate the same trend. For matrix tables, such trend wasn't observed.

4.2. User interface for rating questions

According to our results, students significantly more often responded that they spent less than 1 h per week on online course activities and 1–2 h per week when the question was displayed in the radio button format (Table 4). Moreover, 3.3% of such students chose the "Don't know" option, whereas, among students who were presented slider and text answer box user interfaces, the proportions were 11.5%, and 14.3%, respectively.

These results allow us to support the hypothesis that the radio button user interface is easier for respondents than sliders and text answer boxes for rating questions.

4.3. Layout of question's options

The effect of an option's layout was tested by using two types of questions: for assessing satisfaction with different aspects of an online course (difficulty, workload, and pacing) and for indirect measurement of general satisfaction with a course. For all considered questions, the effect of layout was not significant (refer to Tables 5 and 6). The difference in the response structure is not significant at $p < 0.05$.

In addition, there is no significant difference in the time that respondents spend on answering questions that have a horizontal or vertical format. Consequently, we can conclude that the layout of the questions' options does not affect responses.

4.4. Size of answer box in open-ended narrative questions

Results of the fourth experiment show the significant influence of the answer box size on the length of responses to open-ended questions, with those receiving the small answer box providing shorter answers than those receiving the large answer box (Table 7). Although large answer boxes yielded 81 (*What did you like most about the course?*) and 133 (*What did you like least about this course?*) characters on average, small answer boxes yielded 67 and 103 characters on average for the same questions. The differences in the mean number of words was approximately 2 for the first question (*What did you like most about the course?*) and nearly 5 for the second question (*What did you like most about the course?*). However, significant differences were detected solely for questionnaires in Russian. While similar trend was observed for English questionnaires too, it wasn't statistically significant (Table 8).

As was hypothesized, the size of the answer box did not significantly influence the item nonresponse ("Don't know"). For

How many hours per week did you spend on this course?

- ○ Less than 1 hour
- ○ 1-2 hours
- ○ 3-4 hours
- ○ 5-6 hours
- ○ 7-8 hours
- ○ 9-10 hours
- ○ 11-12 hours
- ○ 13-14 hours
- ○ 15-16 hours
- ○ 17-18 hours
- ○ 19-20 hours
- ○ More than 20 hours
- ○ Don't know

How many hours per week did you spend on this course?

0 hours ▭━━━━━━━━━━▭ more than 20 hours	
Don't know	☐

How many hours per week did you spend on this course?

▭ hours	
Don't know	☐

Fig. 3. Screenshots of Experimental Manipulation 2: The question with the radio button user interface (top), the question with the slider user interface (middle), and the question with the text answer box user interface (bottom).

both small and large answer boxes, approximately 70% of respondents provided a substantive answer to each of the two open-ended questions. Significant differences were detected solely for questionnaires in English. Respondents who answered questions with small boxes on English questionnaires less often chose the "Don't know" option than those who answered questions with large boxes. The percentages for the second question (*What did you like least about this course?*) differed on a marginally significant level ($p < 0.10$), whereas differences for the first question (*What did you like most about the course?*) were significant and achieved 6% ($p < 0.05$). The possible explanation of this finding originates from the fact that English is not the native language for a large share of the respondents. Therefore, the respondents can be more influenced by a box size. Large answer boxes can discourage them from providing responses to open-ended questions.

5. Discussion

This paper's objective was to explore how different visual design features and response formats influence data quality in Web surveys of MOOC students. We presented the results of four full-factorial randomized experiments testing the following factors: 1) order of response options; 2) user interface for rating questions; 3) layout of question options; and 4) size of answer boxes in open-ended questions.

The first experiment showed that responses to scalar questions do not differ substantially when response options are presented in either ascending or descending order. Our findings verify the results of Christian et al. (2008) and Christian et al. (2009), while contradicting those of Stern et al. (2007), Toepoel et al. (2006, 2009), Malhotra (2008) and Tourangeau et al. (2013) for separately presented (one per screen) vertically oriented radio button questions. One possible explanation for such an inconsistency

What did you like most about the course?

What did you like most about the course?

Fig. 4. Screenshots of experimental manipulation 4: Large answer box (top) versus small answer box (bottom).

Table 2
Effect of response options order.

Please rate the difficulty of the course (%)							
Format		Too easy	Somewhat easy	Just right	Somewhat difficult	Too difficult	Don't know
Vertical orientation[a]	Ascending (N = 792)	1.3	16.4	49.4	25.9	2.9	4.1
	Descending (N = 737)	1.4	15.5	52.6	24.3	2.4	3.8
Horizontal orientation[b]	Ascending (N = 780)	0.9[*]	13.8[*]	48.2[*]	29.5[*]	4.1[*]	3.5[*]
	Descending (N = 771)	1.9[*]	13.6[*]	54.2[*]	25.7[*]	1.9[*]	2.7[*]
Please rate the course workload (%)							
Format		Too light	Somewhat light	Just right	Somewhat heavy	Too heavy	Don't know
Vertical orientation[c]	Ascending (N = 770)	1.0	16.6	54.7	19.6	2.7	5.4
	Descending (N = 722)	1.1	14.8	55.1	21.2	2.1	5.7
Horizontal orientation[d]	Ascending (N = 760)	0.8	12.6	59.7	20.5	2.4	3.9
	Descending (N = 760)	1.4	14.2	58.6	19.3	1.8	4.6
Please rate the course pacing (%)							
Format		Too slow	Somewhat slow	Just right	Somewhat fast	Too fast	Don't know
Vertical orientation[e]	Ascending (N = 764)	2.4	11.6	65.6	14.1	1.8	4.5
	Descending (N = 713)	1.3	10.7	64.8	16.7	2.4	4.1
Horizontal orientation[f]	Ascending (N = 748)	0.9	12.3	66.6	15.0	2.0	3.2
	Descending (N = 748)	2.3	11.1	65.4	17.5	1.3	2.4

Note. The response structures are presented. [*]p < 0.05; [**]p < 0.01; [***]p < 0.001; Significance tests are Chi-square tests.

[a] $\chi^2 = 1.865$, df = 5.
[b] $\chi^2 = 14.705$, df = 5.
[c] $\chi^2 = 1.993$, df = 5.
[d] $\chi^2 = 3.419$, df = 5.
[e] $\chi^2 = 4.922$, df = 5.
[f] $\chi^2 = 8.054$, df = 5.

derives from the fact that the experiments were conducted on different respondent groups. The respondents in the studies by Christian et al. (2008) and Christian et al. (2009) were university students, whereas the surveys of Stern et al. (2007), Toepoel et al. (2006, 2009), Malhotra (2008) and Tourangeau et al. (2013) were conducted using samples taken from the general population. It is likely that the university students and graduates comprising the main body of respondents in MOOC participants' surveys will better be able to cope with the task of completing surveys online. The process of online surveying is easier for them due to their greater skill in working with computers and the Internet than older population groups and those with lower levels of education.

In addition, it was found over the course of this research that the order of options has no impact on the data quality for matrix table questions. However, some evidence of the effect of options ordering on the data quality was observed for separately presented (one per screen) horizontally oriented radio button questions. Those who responded to questions with descending scales (with the most

positive option placed furthest to the left) chose the most positive options less often. This finding can be explained by the character-istics of the experimental design. The non-substantive ("Don't know") option was positioned furthest to the right in both exper-imental conditions. As such, the most positive option was extreme solely in the descending scale arrangement. Accordingly, the reduced share of most positive answers for questions with a descending scale can be explained by the tendency of respondents to avoid extreme answers, something well-known in self-administrated surveys (Kiesler & Sproull, 1986).

In the second experiment, we found a significant effect of questions' user interface on data quality. The use of the radio button format for questions regarding the average number of hours per week spent by respondents on the course led to a reduction in the percentage of respondents choosing the non-substantive ("Don't know") option and is likely to have eliminated overestimation bias. Such results support the findings of Couper et al. (2001), Bayer and Thomas (2004), and Funke et al. (2011); that slider and text box

Table 3
Effect of response options order.

Format		Absolutely useless	Somewhat useless	Somewhat useful	Very useful	Don't know
Watching video lectures (%)						
Matrix tables[a]	Ascending (N = 451)	1.8**	3.6**	18.8**	73.6**	2.2**
	Descending (N = 488)	1.0**	1.6**	22.5**	74.6**	0.2**
Radio Buttons[b]	Ascending (N = 486)	1.0**	1.9**	19.5**	76.5**	1.0**
	Descending (N = 482)	1.9**	3.7**	26.8**	66.2**	1.5**
Reading additional course materials (%)						
Matrix tables[c]	Ascending (N = 331)	2.4	7.6	41.7	42.0	6.3
	Descending (N = 361)	1.4	7.8	50.7	34.1	6.1
Radio buttons[d]	Ascending (N = 363)	1.4***	3.6***	48.2***	44.9***	1.9***
	Descending (N = 373)	2.1***	7.5***	56.0***	29.0***	5.4***
Completing quizzes (%)						
Matrix tables[e]	Ascending (N = 426)	1.2	6.1	30.0	58.9	3.8
	Descending (N = 453)	2.0	7.1	36.9	52.1	2.0
Radio buttons[f]	Ascending (N = 451)	1.3	4.4	29.9	62.1	2.2
	Descending (N = 456)	2.0	3.7	36.4	54.2	3.7
Reading postings/asking questions in the discussion forum (%)						
Matrix tables[g]	Ascending (N = 273)	4.0	16.1	44.0	23.1	12.8
	Descending (N = 281)	6.0	22.1	38.4	24.2	9.3
Radio buttons[h]	Ascending (N = 297)	2.7	18.5	48.5	25.6	4.7
	Descending (N = 290)	4.5	18.6	51.4	20.0	5.5
Reading postings/communicating with other course participants on social networks (%)						
Matrix tables[i]	Ascending (N = 223)	12.1	21.5	30.5	15.2	20.6
	Descending (N = 221)	11.8	24.4	29.4	14.0	20.4
Radio buttons[j]	Ascending (N = 190)	11.1*	28.9*	34.2*	18.9*	6.8*
	Descending (N = 172)	11.0*	29.1*	36.6*	8.7*	14.5*

Note. The response structures are presented. $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; Significance tests are Chi-square tests.
[a] $\chi^2 = 14.011$, df = 4.
[b] $\chi^2 = 13.686$, df = 4.
[c] $\chi^2 = 6.883$, df = 4.
[d] $\chi^2 = 26.481$, df = 4.
[e] $\chi^2 = 8.520$, df = 4.
[f] $\chi^2 = 7.890$, df = 4.
[g] $\chi^2 = 6.378$, df = 4.
[h] $\chi^2 = 3.753$, df = 4.
[i] $\chi^2 = 0.580$, df = 4.
[j] $\chi^2 = 11.940$.

Table 4
Effect of three different formats of questions on data quality.

	Radio buttons (N = 1002, %)	Slider (N = 964, %)	Text answer box (N = 969, %)
Less than 1 h	4.5	3.3	0.6
1–2 h	20.2	16.4	17.9
3–4 h	35.7	30.9	35.9
5–6 h	20.9	18.2	18.6
7–8 h	8.5	9.5	6.8
9–10 h	3.8	5.8	3.3
11–12 h	2.0	1.6	0.6
13–14 h	0.3	0.4	0.0
15–16 h	0.5	1.5	0.8
17–18 h	0.0	0.3	0.1
19–20 h	0.1	0.4	0.4
More than 20 h	0.3	0.2	0.6
Don't know	3.3	11.5	14.3

Note. The response structures are presented. Question: How many hours per week did you spend on this course? $\chi^2 = 141.819$, df = 24, $p < 0.001$.

interfaces are more challenging for respondents. At the same time, further research should be conducted to estimate the effect of questions' user interface on responses.

The main goal of the third experiment was to ascertain whether the layout of the question options has an impact on data quality. The results of the experiment showed no significant differences between the answers of respondents who completed questionnaires with vertical or horizontal orientations of question options. Nor did the layout of questions affect the average response time. This finding supports our hypothesis and verifies inferences made in previous studies.

Finally, we tested the effect of text answer box size on responses

to narrative open-ended questions. The results of this experiment support the findings of several previous investigations (Christian & Dillman 2004; Emde & Fuchs, 2012; Israel 2006, 2010). It was observed that respondents who answer questions provided with larger answer boxes are more likely to write longer comments. At the same time, the use of larger text boxes resulted in an increase in the share of "Don't know" answers in surveys conducted in English. For surveys in Russian, this effect was not observed.

The main limitation of this research is the lack of generalization and validity of the results. The first limitation is related to the low response rate, ranging between 0.7% and 5.3%. The problem of validity concerns the lack of information on differences between

Table 5
Effect of layout in rating course aspects.

Format/Question	Question's option					
Please rate the difficulty of the course (%)[a]	Too easy	Somewhat easy	Just right	Somewhat difficult	Too difficult	Don't know
Horizontal	1.4	13.7	51.2	27.6	3.0	3.1
Vertical	1.3	16.0	50.9	25.1	2.7	4.0
Please rate the course workload (%)[b]	Too light	Somewhat light	Just right	Somewhat heavy	Too heavy	Don't know
Horizontal	1.1	13.4	59.1	19.9	2.1	4.4
Vertical	1.1	15.8	54.9	20.4	2.4	5.4
Please rate the course pacing (%)[c]	Too slow	Somewhat slow	Just right	Somewhat fast	Too fast	Don't know
Horizontal	1.6	11.7	66.0	16.2	1.7	2.8
Vertical	1.8	11.2	65.2	15.4	2.1	4.3
How would you rate your understanding of the subject matter after taking this course? (%)[d]	Poor	Fair	Good	Very good	Excellent	Don't know
Horizontal	4.1	31.2	41.2	17.2	3.1	3.2
Vertical	3.7	32.8	41.8	14.2	3.0	4.5

Note. The response structures are presented. Significance tests are Chi-square tests. There are no significant differences.
[a] $\chi^2 = 6.792$, df = 5.
[b] $\chi^2 = 7.888$, df = 5.
[c] $\chi^2 = 6.398$, df = 5.
[d] $\chi^2 = 8.260$, df = 5.

Table 6
Effect of Layout in Assessing Probability to Revisit Course Materials or Recommend it to Others.

Format/Question	Question's option				
	Very unlikely	Somewhat unlikely	Somewhat likely	Very likely	Don't know
How likely would you be to revisit the course materials in the future? (%)[a]					
Horizontal	2.5	8.1	40.4	44.0	5.0
Vertical	3.0	8.8	40.8	42.3	5.1
How likely would you be to take this course again? (%)[b]					
Horizontal	16.9	26.4	29.0	16.9	10.8
Vertical	15.5	29.4	25.9	17.9	11.3
How likely would you be to recommend this course to a friend? (%)[c]					
Horizontal	3.0	5.5	36.6	50.7	4.2
Vertical	3.6	4.3	38.6	48.5	5.0

Note. The response structures are presented. Significance tests are Chi-square tests. There are no significant differences.
[a] $\chi^2 = 1.422$, df = 4.
[b] $\chi^2 = 6.007$, df = 4.
[c] $\chi^2 = 5.334$, df = 4.

Table 7
Effect of answer box size.

	Mean number of words	Mean number of characters	Don't know (%)
What did you like most about the course?			
Small answer box (N = 858)	9.4*	67.2*	28.7
Big answer box (N = 818)	11.6*	81.3*	31.0
Sig. tests	−2.983	−2.832	1.740 (df = 1)
What did you like least about this course?			
Small answer box (N = 690)	16.0*	103.7*	29.9
Big answer box (N = 674)	20.6*	133.4*	29.7
Sig. tests	−3.714	−3.967	0.008 (df = 1)

Note. *$p < 0.05$; Significance tests are one-sided t-tests for the number of words and signs and chi-squared tests for percent elaborated. Outliers, who typed more than 1000 characters, were excluded from the analysis.

MOOC students who participated in the post-surveys, thus limiting the control for external factors and covariates (gender, education, age, etc.). Consequently, it is desirable that scholars develop these research conclusions accordingly.

6. Conclusion

According to the results of the experiments, certain general conclusions and recommendations can be made regarding MOOC Web survey design features. Firstly, it is better to use traditional radio button interfaces for rating questions regarding, for example, the number of hours spent on courses. Secondly, to obtain detailed answers in open-ended narrative questions (for example, regarding general comments on the course), it is better to use relatively large answer boxes if the survey language is native to most respondents, and relatively small answer boxes in other situations. These conclusions should be verified in future research. In addition, more detailed investigation on the effect of visual design and response formats should be carried out in groups of respondents with different characteristics (education, psychological features, social

Table 8
Effect of answer box size.

	Mean number of words		Mean number of signs		Don't know (%)	
	Russian	English	Russian	English	Russian	English
What did you like most about the course?						
Small answer box	8.2[*]	11.3	60.5[*]	76.9	31.7	23.8[*]
Big answer box	10.5[*]	13.6	75.6[*]	90.7	31.6	29.9[*]
Sig. tests	−3.271	−1.843	−3.271	−1.849	0.001 (df = 1)	4.690 (df = 1)
What did you like least about this course?						
Small answer box	13.4[*]	19.4	91.2[*]	119.9	44.1	33.1
Big answer box	19.3[*]	22.7	128.3[*]	141.8	40.4	38.5
Sig. tests	−4.115	−1.502	−4.163	−1.685	2.353 (df = 1)	3.146 (df = 1)

NOTES - [*]$p < 0.05$. Significance tests are one-sided t-tests for the number of words and signs and chi-squared tests for percent elaborated. Outliers, who typed more than 1000 characters, were excluded from the analysis.

background, etc.).

References

AAPOR (American Association for Public Opinion Research) Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 2008. URL: http://www.aapor.org/AM/Template.cfm?Section=Standard_Definitions2&Template=/CM/ContentDisplay.cfm&ContentID=3156.

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

Bayer, L., & Thomas, R. (2004). A comparison of sliding scales with other scale types in online survey. In *Paper presented at the RC33 6th International Conference on Social Science Methodology, Amsterdam*.

Behr, D., Bandilla, W., Kaczmirek, L., & Braun, M. (2014). Cognitive probes in Web surveys on the effect of different text box size and probing exposure on response quality. *Social Science Computer Review, 32*, 524–533.

Chen, P. S. D. (2011). Finding quality responses: the problem of low-quality survey responses and its impact on accountability measures. *Research in Higher Education, 52*, 659–674.

Christian, L. M. (2003). *The influence of visual layout on scalar questions in web surveys* (Master's thesis). Washington State University.

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. (2008). The effects of mode and format on answers to scalar questions in telephone and web surveys. *Advances in Telephone Survey Methodology, 12*, 250–275.

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

Cook, C., Heath, F., Thompson, R., & Thompson, B. (2001). Score reliability in Web- or Internet-based surveys: unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement, 61*(4), 697–706.

Couper, M. (2000). Web surveys – a review of issues and approaches. *Public Opinion Quarterly, 64*(4), 464–494.

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

Emde, M., & Fuchs, M. (2012). Using adaptive questionnaire design in open-ended questions: a field experiment. In *Paper presented at AAPOR Annual Conference. May 17–20, 2012, Orlando, USA*.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: a systematic review. *Computers in Human Behavior, 26*(2), 132–139.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Fisher, R. A. (1935). *The design of experiments*. Edinburgh and London: Oliver and Boyd.

Funke, F. (2015). A Web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales. *Social Science Computer Review*. http://dx.doi.org/10.1177/0894439315575477.

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

Funke, F., & Toepoel, V. (2015). Click, touch, slide: impact of the implementation of graphical rating scales on data quality in mobile and desktop settings. In *Paper presented on the General Online Research Conference, 18–20 March 2015*.

Israel, G. D. (2010). Effects of answer space size on responses to open-ended questions in mail surveys. *Journal of Official Statistics, 26*, 271–285.

Israel, G. (2006). Visual cues and response format effects in mail surveys. In *Paper presented at the Annual Meeting of the Southern Rural Sociological Association. February 7, 2006, Orlando, USA*.

Kiesler, S., & Sproull, L. S. (1986). Response effects in the electronic survey. *Public Opinion Quarterly, 50*(3), 402–413.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly, 72*, 914–934.

Maloshonok, N., & Terentev, E. (2014). Vliyanie dizayna ankety na kachestvo dannyh v onlain-oprosah studentov (in Russian) (Effects of questionnaire design on data quality in student online surveys). *The Monitoring of Public Opinion: Economic and Social Changes Journal, 124*, 15–27.

Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education, 44*, 389–407.

Roster, C. A., Lucianetti, L., & Albaum, G. (2015). Exploring slider vs. categorical response formats in Web-based surveys. *Journal of Research Practice, 11*. Retrieved from http://jrp.icaap.org/index.php/jrp/article/view/509/413.

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*, 325–337.

Stern, M. J., Dillman, D. A., & Smyth, J. D. (2007). Visual design, order effects, and respondent characteristics in a self-administered survey. *Survey Research Methods, 1*, 121–138.

Stern, M. J., Smyth, J. D., & Mendez, J. (2012). The effects of item saliency and question design on measurement error in a self-administered survey. *Field Methods, 24*, 3–27.

Toepoel, V., Das, M., & van Soest, A. (2006). *Design of Web questionnaires: The effect of layout in rating scales*. CentER Discussion Paper 2006-30. Tilburg: CentER.

Toepoel, V., Das, M., & van Soest, A. (2009). Design of Web questionnaires: the effect of layout in rating scales. *Journal of Official Statistics, 25*, 509–528.

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

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

Tourangeau, R., Couper, M. P., & Conrad, F. G. (2013). "Up means good" the effect of screen position on evaluative ratings in Web surveys. *Public Opinion Quarterly, 77*, 69–88.

Veen, F., Goritz, A. S., & Sattler, S. (2015). Response effects of prenotification, prepaid cash, prepaid vouchers, and postpaid vouchers: an experimental comparison. *Social Science Computer Review*. Retrieved from http://doi.org/10.1177/0894439315585074.

Walston, J. T., Lissitz, R. W., & Rudner, L. M. (2006). The influence of Web-based questionnaire presentation variations on survey cooperation and perceptions of survey quality. *Journal of Official Statistics, 22*(2), 271–291.

Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior, 58*, 119–129.

Zuell, C., Menold, N., & Körber, S. (2015). The influence of the answer box size on item nonresponse to open-ended questions in a web survey. *Social Science Computer Review, 33*, 115–122.

Zusman, B. J., & Duby, P. (1987). An evaluation of the use of monetary incentives in postsecondary survey research. *Journal of Research and Development in Education, 20*, 73–78.