RECSM Working Paper Number 37

2014

Universitat Pompeu Fabra - Research and Expertise Centre for Survey Methodology
Edifici ESCI-Born - Despatxos 19.517-19.527
Passeig Pujades, 1 - 08003 Barcelona
Tel.: +34 93 542 11 63 web: http://www.upf.edu/survey/ e-mail: recsm@upf.edu

Forcing respondents of online surveys to slow down: An effective strategy?

Melanie Revilla

RECSM, Universitat Pompeu Fabra

Carlos Ochoa,

Albert Turbina,

Netquest

Willem Saris,

RECSM, Universitat Pompeu Fabra

Abstract:

The goal of this paper is to investigate the effects of forcing the respondents to slow down while answering online surveys. This is done by setting up an experiment with respondents from the Netquest panel in Mexico in which respondents are randomly assigned to a control group (free to go as fast as their Internet connection allows them) or to different treatment groups. The treatment groups are forced to slow down by using a timing control (minimum time per page necessary before they can go on), by showing the different blocks on the page progressively (first the question, later the scale, later the "next" button), or by fading-in the text of the question character by character. The minimum time per page is fixed by us in half of the groups and determined by tuning in the other half. The results show that there are some positive

effects on the quality, mainly from the treatment which shows progressively the different blocks (question-scale). There are also some negative effects in all treatments (more complaints) but the abandon rates are higher only in the timing control condition. Overall, forcing respondents to slow down seems to be a strategy with more positive than negative effects if it is done by presenting progressively the different elements. We believe it should be further considered for future research.

Keywords: Web surveys, response times, speeding, quality, timing control, Netquest

Acknowledgements: We would like to thank Germán Loewe for his support and his helpful comments on a previous draft of this paper.

Number of words excluding tables, etc: 6497

Forcing respondents of online surveys to slow down:

An effective strategy?

Introduction

Since the middle of the 19th century, psychologists have been collecting and studying response times (starting with Donders, 1868). But with the development of computer-assisted modes of data-collection, which makes it much easier to capture paradata (i.e. data generated directly by the survey data collection process that can be used to describe and evaluate that process, Couper, 2000), it became much more common, not only in psychology but also in social sciences or marketing.

Response times have been used as a measure of the "amount of information processing necessary to answer a question" (Bassili and Scott, 1996, pp391). In that way, long response times have been seen as potential indicators of problems in the questions. They were also related to the complexity of the question (Yan and Tourangeau, 2008). Moreover, they have been used to assess the strength of attitudes or intentions like voting (Bassili, 1993). If respondents are certain of their position, they can answer quickly. On the contrary, in the case of uncertainty, they may need more time to decide on their answer.

However, interpreting response times is difficult because human beings tend to satisfice, i.e. to seek solutions that are simply satisfactory instead of optimal in order to minimize the psychological costs (Simon, 1957). In the frame of surveys, it means respondents select the first acceptable answer, in order to reduce their efforts, instead of looking for the best answer (Krosnick and Alwin 1987; Krosnick, 1991). As a

consequence, short response times could also be the result of using shortcuts during the response process instead of following all steps of the cognitive process (described for instance by Strack and Martin, 1987; Tourangeau and Rasinski, 1988; Tourangeau, Rips and Rasinski, 2000). If this is so, short response times could indicate low quality answers. In web surveys, this may be even truer, since the absence of interviewer gives complete freedom to the respondents to choose the pace for answering the questions.

A study of Malhotra (2008) looks at this possibility. He finds that shorter response times are associated with higher satisficing, measured by primacy effects (tendency to select more the first answer category, cf. Schuman and Presser, 1981), mainly for low educated respondents. Zhang (2013) considers a different measure for satisficing: the level of straight-lining (systematic selection of the same answer category for all items in a battery) and finds again an association with shorter response times. Revilla and Ochoa (2013) use a Structural Equation Modeling approach to estimate the relationship between response times and quality (each considered as latent concepts measured by several indicators) while controlling for potential spurious effects and find once more a significant association between response times and quality: the quality is lower when the speed is higher. Our own experience with web surveys is that response times tend to be lower compared to other modes. In some cases, they are even so low that it seems impossible that the respondents answer properly.

If response times and quality are associated, maybe by acting on the response times it is possible to improve the quality. Researchers cannot force people to think. But in web surveys, they can force them to spend a minimum time on each question. There are two potential limits in this approach. First, previous research only demonstrated an association. If the relationship is not causal, then acting on the response times will not

produce any effect on quality. Second, even if the respondents have to spend a minimum time on each question, this does not guarantee they will think more. Nevertheless, the incentives they have to satisfice are reduced. Even if they select quickly an answer, they will still have to wait before going on, such that by satisficing, they can reduce their cognitive efforts but they cannot gain time. Therefore, we can argue that acting on the response times reduce the added value of satisficing and by consequence, may improve the quality of the answers.

Kapelner and Chandler (2010) did experiments in this direction. They force respondents to slow down while answering the survey using two different treatments (pp3): a "timing control" and a "Kapcha". As they explain, the "*Timing control* treatment is identical to the *Control* treatment except that the continue button is disabled and has a spinning graphic during a waiting period after which the continue button is enabled. The *Kapcha* treatment goes one step further and, in addition to slowing down the respondent for a time equal to the *Timing control* treatment, also draws additional attention to the instructions and answer choices by "fading-in" the survey's words at 250 words per minute". They find some positive impact of the treatments on the quality, especially for the Kapcha. However, they also find more attrition, again especially for the Kapcha.

The goal of this paper is to further investigate the effects of forcing respondents of online surveys to slow down. The experiment by Kapelner and Chandler (2010) used Amazon's Mechanical Turk (MTurk) workers to answer the survey. This is a very specific kind of online surveys. Besides, they consider a limited number of treatments and assess their effects looking only at a few indicators of quality. We propose an experimental design with more treatments, tested using data from on online panel

oriented toward marketing and by considering more indicators of quality as well as more potential negative effects.

In section 2, the design of the experiment is explained. In section 3, hypotheses about the expected positive and negative effects are developed. Section 4 gives some details about the data and method used to test these hypotheses. In section 5, the main results are presented. Finally, section 6 concludes.

I) Design of the experiment

The experiment wants to test the effect of forcing the respondents to slow down while answering online questionnaires. To do that, different treatments groups are compared with a control group, which gets the survey in the usual way.

Several options are tested for the treatments. “Treatment 1” does not let the respondents go to the next page before a certain time has been spent on a question. This is similar to the “timing control” group in Kapelner and Chandler (2010). In that case, an automatic control is programmed such that the computer allows going on only once some minimum time is over. If respondents spend more than this minimum time before clicking the “Next” button, then nothing differs for them compared to the control group. But if respondents click on the “Next” button before, they get a pop-up window with a warning message: “we consider that you did not spend enough time reading this question. Please, read it carefully before going on”. When they close the pop-up window to come back to the questionnaire, the “Next” button appears as a clock until the minimum time is finished. At that point, the button takes its original appearance again and respondents are allowed to go on.

In line with Kapelner and Chandler (2010), we also implement the “Kapcha” treatment. This will be our treatment 3. In that case, the text enters on the page character by character. In that way the respondents cannot go directly to the answers but have to follow little by little what is coming in the page. This could incite respondents to read more linearly the information in the page instead of reading in diagonal or not reading at all.

In addition, we propose one new condition, intermediate between treatments 1 and 3. We call it treatment 2. Now, the information appears per blocks: first, the introduction (if there is one), then, the request for an answer, later, the answer categories and finally, the “Next” button.

The minimum time respondents have to spend on one page is the same in the case of treatment 1 (Timing Control), treatment 2 (Question-scale) and treatment 3 (Kapcha). Treatment 3 is used in order to determine the minimum time for each page. It depends on the velocity with which the characters enter on the page. Then, we apply the same time to treatments 1 and 2 (with a time for each block).

Another novelty of this study stands in the way of determining this minimum time. Kapelner and Chandler (2010) used a fixed speed of 250 words per minute. We replicate this condition. Besides the time of entering of the words, we add one second for the time of answering each item and another second for the time of clicking the “Next” button.

However, there are alternatives ways of doing it. For instance, the data could be collected first for a small control group. The response times per page in this sample could be used to determine reasonable minimum times for each page. Another option

would be to determine the minimum times based on the first questions of the survey. A few questions could be added at the beginning of the survey that would not be used for substantive purposes but to assess what should be the minimum response time. Still another option is to do a "test" of how fast the respondents read before starting the survey. Then we can adapt the minimum times per page for each respondent such that these times fit with his/her reading skills.

In this study, we implement the last option, that we call "tuning option". The initial test of speed of reading allows respondents to adjust the speed between +-ΔT% in relation to the fixed time of 250 words per minute with which we start. Just after the introduction page, half of the respondents of the control group and of each of the treatments groups get the following instruction:

"In this page you will see a text which will get in progressively as soon as you click the button 'start reading'. When it is finished, tell us if the speed in which it appeared was good. Pay a lot of attention since we will ask you some questions about the text at the end".

If respondents report that the speed was good, then they directly get to the next step. If they answer that it was too slow, they are proposed a new text which fade-in at 200 words per minute. If they say it was too quick, they get a new text fading-in at 300 words per minute. If in the second attempt they still think the speed is too quick or too slow, they get a third and last chance to adapt the velocity.

Once they said the speed is adequate or once they have finished the third attempt of adaptation, all respondents get a test question. The goal of this question is to assess if the respondents did adapt well the speed to their own reading skills by asking them a

question about a short text that they got at the chosen speed. This question included an explicit “I could not read well the text” option. If respondents selected it or failed to give the correct answer, then we consider that they did not adapt well the speed to their reading skills. Therefore, in case they were in the treatment groups, we reduced the speed for them (50 less) for the rest of the survey.

Table 1 summarizes the groups for this experiment. Respondents were randomly assigned to one of them. From now on, we will use “with test” to refer to the groups that got the test of speed of reading at the beginning of the survey (tuning condition) and “without test” for the others.

Table 1: the experimental design

	With test	**Without test**
Control	Control-T	Control
Treatment 1: minimum time	TimingControl-T	TimingControl
Treatment 2: progressive blocks	QuestionScale-T	QuestionScale
Treatment 3: progressive characters	Kapcha-T	Kapcha

Note: names for 1 and 3 from Kapelner and Chandler (2010)

Finally, we should notice that in treatments 2 and 3 some more exact specifications are required depending on the kind of questions. For example, in the case of grids of questions, we decided that each new row enters only once the previous row has been answered. All the questions are available using this link:

http://test.nicequest.com/surveys/esnpa/force

Our goal is to study the effects of these different treatments. The effects can be at different levels and be negative or positive. In the next section, we propose a set of hypotheses about the effects expected.

II) Hypotheses

1) Some positive effects are expected

Since we are fixing minimum times per page, we obviously expect a positive impact on the overall speed of answer. If respondents do not slow down, then, the treatments did not do their work.

a) Reduced speed

Therefore, our first hypothesis is the following:

H1: speeding will be reduced in the treatment groups.

We need this hypothesis to hold before going on.

b) Better quality

Nevertheless, the speed in itself is not a very interesting indicator, first because we put constraints on it, and second because if we are worried about speeding, it is not *per se* but as a potential indicator or cause of low quality answers. Therefore, the main positive effect expected when the treatments are used is a higher quality of the answers.

H2: the quality of answer will be higher in the treatment groups.

In order to assess the quality we will consider different indicators as straight-lining, incoherence and precision of answers, etc. We will explain them later.

2) Some negative effects are expected

However, by limiting the freedom of the respondents in choosing the pace of answering the interview, the treatments may also have some negative effects.

Comparing the different treatments, we expect the negative effects to be higher when the minimum times per page are fixed arbitrarily at the same value for all respondents compared to the conditions where there are adapted by tuning with the test of speed of reading at the beginning of the survey. Indeed, the test should allow the adaptation of the minimum times to the respondents reading skills such that the risks of frustration should be lower. If a respondent reads quickly, this should appear at the moment of the test, allowing him/her to do all the rest of the survey quicker. So our third hypothesis is:

H3: The negative effects are smaller when the test of reading is undertaken.

This said, what are the negative effects that we expect?

a) Higher abandon rate

If respondents wants to go quick but they cannot, they may get frustrated and decide to stop answering the survey. Therefore, we expect more abandons when the treatments are used.

H4: the abandon rate will be higher in the treatment groups.

This is what Kapelner and Chandler (2010) found in their experiment. Besides, they found that it was even higher in the Kapcha condition (our treatment 3) than in the timing control condition (our treatment 1).

We believe that our intermediate condition where the information appears on the page per blocks (treatment 2) should be more comfortable for respondents than both the

timing and the Kapcha ones. Therefore, we expect the abandon rates to be the lowest in the control groups, then, in treatment 2, then in 1 and finally in 3.

b) More negative evaluation

Because of the novelty of the method (which can be disturbing) and the limits it puts on respondents freedom, we propose hypotheses 5 and 6.

H5: the respondents will evaluate the questions as being more difficult in the treatment groups.

H6: there will be more complaints in the treatment groups.

Again, we expect that they will evaluate it even more difficult and they will complain even more for treatment 1 (because it may be more difficult for respondents when everything appears at once to really get an idea of what is done) and for treatment 3 (because the fact that the text enters character by character is something unusual that some respondents may not like) than for treatment 2.

3) More positive effects than negative effects

Finally, the general point that we want to test is if it is worth it to use such a design with constraints on the minimum time per page. It will be worth it if Hypothesis 7 holds.

H7: overall, there are more positive effects than negative effects.

III) Data and Method

In order to test all these hypotheses, we will compare different indicators for the two control and six treatment groups. Since most of the indicators compared are classic ones they will be defined or explained directly in the results' section. But we should mention that when nothing is specified, it means that the tests of significance of differences are done between the control group with test (respectively without test) and the treatment groups with test (respectively without test) at the 5% level. In the tables, the stars are next to the values for the group that differ from the control group.

The data comes from an experiment realized with the online fieldwork company Netquest[1] which is accredited with the ISO 26 362 quality standard. The company is present in Spain, Portugal, central and Latin America. Netquest recruits its panellists "by invitation only" using a frame of email addresses of people that agreed at the end of a short satisfaction survey proposed by different websites to give their contact and participate in further surveys.

The data was collected in Mexico between the 5th and the 19th of December 2013. The numbers of observations in each group (control and treatments) are between 200 and 250. In order to get a sample representative of the panel distribution, quotas were used for age and gender (crossed-quotas).

IV) Results

We start by a short overview of results related to the test of speed of reading. Then, we will look at the expected positive and the expected negative effects.

[1] For more information: www.netquest.com

1) The test of speed of reading

In this section, we focus on the respondents that did the test and finished the survey, i.e. 965 respondents. The treatments started to be applied only after the test was achieved. So there is no need to separate control and treatment groups here.

a) Process of choosing the final speed

First, we look at the different patterns of responses for the set of three initial questions proposed to adapt the speed of the survey: for instance, saying that the text was too slow at the first attempt and then saying that it was ok at the second attempt is a pattern. Table 2 gives the percentages of respondents for different patterns.

Table 2: Percentages of individuals in the main patterns

1st question	2nd question	3rd question	% pattern
Ok			**41.87**
Too slow	ok		0.41
Too quick	ok		**40.93**
Too quick	Too quick	ok	**12.64**
Too quick *or* Too slow	Too slow Too quick	ok ok	0.62
Slow	Too slow	ok	0.21
Too slow	Too slow	Too slow	1.04
Too quick	Too quick	Too quick	1.66
Other			0.62
Total			100

We can see that 41.87% of the respondents answered already in the first attempt that the speed was ok. Another 40.93% said that it was too quick but at the second attempt they said it was ok. Aditionally, 12.64% said twice too quick and then ok. These are clearly the three main patterns of responses. Therefore, it seems that the value we

selected (following Kapelner and Chandler, 2010) as basis when fixing the minimum time is too quick for most of the respondents.

We should also notice that with three steps there are only 3.32% of the respondents that did not adapt well the speed yet (i.e. respondents that did not say ok).

b) Results to the "test" question

The "test" question came just after. The objective was to get an indication of how well respondents adapted really the speed to their own reading skills. The text shown mentioned there were three species of birds. The "test" question asked: "how many species of birds are there ?" The option "I could not read well the text was explicitely offered". Table 3 gives the cross distribution of the speed chosen by the respondents and their answer to this test question: "correct" (if they say "3"), "wrong" (if they say another number) or "could not read".

Table 3: Cross distribution speed choosen & answer to the test question: percentages (and number of observations)

Speed / answer test	Wrong	Correct	Could not read	Total
100 or 150	16.67% (23)	69.57% (96)	13.77% (19)	100% (138)
200	15.83% (63)	64.82% (258)	19.35% (77)	100% (398)
250	21.65% (89)	59.37% (244)	18.98% (78)	100% (411)
300 or 350 or 400	5.56% (1)	72.22% (13)	22.22% (4)	100% (18)
Total % (no. obs)	18.24% (176)	63.32% (611)	18.45% (178)	100% (965)

Overall, 63.32% of the respondents passed the test. However, there are still 18.45% that said they could not read well and 18.24% that failed. If we look for different speeds, the highest percentage of correct answers is found for the group with highest speed (300 to 400). This suggests that there are some respondents highly skilled that are

able to go quick and understand what they read. But this is a reduced group. For the rest, when the speed is lower, the percentage of correct answers is higher.

c) Final speed

For all respondents we start with a speed of 250 words per minute. Then, for the groups with test, the final speed depends on the pattern of answers at the three questions of adaptation of the speed (presented in subsection a) and the answer to the "test" question (presented in subsection b). If respondents failed this question or said they could not read well the text, then we subtract 50 from the final speed they choose to fix their final speed. Doing so, we get the distribution of final speeds presented in Table 4.

Table 4: Final speed for the respondents with tuning

Speed	%	(cumul%)
50	.31	(.31)
100	5.39	(5.70)
150	23.11	(28.81)
200	44.04	(72.85)
250	25.70	(98.55)
300	.21	(98.76)
350	.31	(99.07)
400	.93	(100.0)

Table 4 shows that 72.85% of the respondents get a final speed that is lower than in the groups without tuning. Another 25.70% get 250 words per minute. Only very few get a higher speed. Therefore, our third hypothesis may not be verified. We expected the negative effects to be lower when the respondents got the tuning condition because we thought that if they read quickly they will choose higher speed. However, most of the respondents in the tuning condition ended up with a lower speed than in the fixed condition. By consequence, the constraints are stricter for the groups with tuning.

The fact that so many respondents ended up with a speed lower than 250 words per minute may be related with the country studied. It might be that in Mexico the reading skills are not as high as in the USA, which is the country used by Kapelner and Chandler (2010) for their experiment. It may also be that respondents got surprised by the letters fading-in progressively and because of that, got the impression it was too quick when in fact they were just not used to it yet.

To summarize, the adaptation of the speed did not work as expected. We though that the tuning will allow some respondents to choose a quicker speed than in the conditions without test, such that the treatment with tuning will be less strict. However, the large majority of respondents ended up with a speed lower than 250 words per minute: part of them because they chose so, part of them because they fail at the test question and so we reduce their final speed compared to the one they chose.

2) Expected positive effects

a) Reduced speed

We first check whether the constraints on the response times really reduced the overall speed. Table 5 gives the distribution of completion times[2] for the different groups. We should notice that with a speed of 250 words per minutes, the minimum time to fill in the questionnaire was around 14 minutes.

[2] The server-side times are used here, meaning that the downloading times are included. In order to make the comparisons across groups, the total time of completions are computed starting with the first common question to all groups (i.e. the test of speed of reading is excluded). Also, since respondents sometimes keep the survey open but go away for a while, we had to get rid of the abnormally long response times. To do so, we replaced the highest 1% times for each page by the mean of the other 99%.

Table 5: Total time of completion

	ALL	Control		Treatment 1 Timing control		Treatment 2 QuestionScale		Treatment 3 Kapcha	
		with	without	with	without	with	without	with	without
%	10 min or less		0.87	0	0	0	0	0.80	0
	10.01 min to 15 min	11.24	12.12	0	0	0	0	0	0
	15.01 min to 20 min	27.31	24.68	0	0.89	.40	0.45	0	0
	20.01 min to 25 min	22.89	21.21	6.57	15.11	3.16	11.71	8.00	11.20
	25.01 min to 30 min	14.86	17.32	14.08	22.22	13.04	23.87	14.40	21.99
	30.01 min to 45 min	16.87	16.88	43.19	40.44	48.62	41.44	50.40	42.32
	45.01 min to 60 min	3.61	3.46	15.96	13.78	18.58	14.86	16.00	17.43
	More than 60	3.21	3.46	20.19	7.56	16.21	7.66	10.40	7.05
Cumul %	10 min or less	0	.87	0	0	0	0	0.80	0
	10.01 min to 15 min	11.24	12.99	0	0	0	0	0.80	0
	15.01 min to 20 min	38.55	37.66	0	0.89	0.40	0.45	0.80	0
	20.01 min to 25 min	61.45	58.87	6.57	16.00	3.56	12.16	8.80	11.20
	25.01 min to 30 min	76.31	76.19	20.66	38.22	16.60	36.04	23.20	33.20
	30.01 min to 45 min	93.17	93.07	63.85	78.67	65.22	77.48	73.60	75.52
	45.01 min to 60 min	96.79	96.54	79.81	92.44	83.79	92.34	89.60	92.95
	More than 60	100	100	100	100	100	100	100	100
Median in minutes		21.98	22.73	38.72	32.58	39.22	33.30	36.28	34.25

Around 11%- 13% of the respondents in the control groups spent less than 15 minutes filling in the questionnaire. In the treatment groups an equivalent percentage is therefore forced to slow down. But Table 5 shows that there are even less than 1% of respondents in the control groups answering in less than 20 minutes, even in the groups without test where respondents could therefore have answered in around 14 minutes. The treatments make respondents slow down really quite a lot and more than just the minimum authorized.

We also consider the number of times the respondents clicked before the minimum time in the Timing control condition (treatment 1) as an indicator of how often they would like to go quicker. Table 6 gives the results. The variable “total” counts the total number of clicks before time each respondent did through the complete survey. The variable “average” divides this total by the number of pages in the survey (87) to get the number of clicks before time per page.

Table 6: Number of clicks before time, in total and per page

Treatment 1	Total				Average			
Timing Control	mean	sd	min	max	mean	sd	min	max
With test	24.02	23.61	3	156	.2761	.2714	.0345	1.7931
Without	16.63	11.82	2	106	.1911	.1359	.0230	1.2184

In average the respondents clicked 24 times before the minimum time in the group with test and 17 times in the group without test, which corresponds to a mean of .27 clicks per page in the group with test and .19 in the group without test. The differences in means between the group with and without test are statistically significant. This is the consequence of higher minimum times imposed in the group with test (as seen in Table 4).

A few respondents are clicking up to 1.8 times per page, whereas the majority is clicking less than once every five pages. Still, almost all the respondents clicked before the time at least a few times in the survey. So the control on time affected almost all the respondents.

b) Higher quality

Instructional Manipulation Check (IMC)

We consider several indicators of quality. The first one is the fact that respondents pass an Instructional Manipulation Check (IMC). As defined by Oppenheimer, Meyvis and Davidenko (2009, pp.867), an IMC "consists of a question embedded within the experimental materials that asks participants (…) to provide a confirmation that they have read the instruction." If people do well the manipulation asked, it suggests that they really read the instruction. In this survey, the IMC asked respondents to select all buttons in the corresponding row besides the five preferred options they were asked to

select. Table 7 indicates the percentage of respondents that correctly followed this instruction.

Table 7: Percentages of respondents that passed the IMC

In %	With test	Without test
Control	29.72	24.24
Treatment 1 - Timing Control	35.68	33.78*
Treatment 2 - QuestionScale	4.35*	3.15*
Treatment 3 - Kapcha	4.00*	7.88*

Note: * indicates a significant difference at the 10% level. This is the same in all tables.

In the control group a large majority of respondents fail the IMC. When a minimum time per page is imposed, there is still a majority of respondents failing the IMC. But a few more answer it correctly (significant difference for the group without test). Therefore, the timing control seems to have some positive effect on the attention respondents pay to reading, even if it may not be very large.

For treatments 2 and 3, the results are completely opposite of what we expected: almost all the respondents are failing the IMC. We were first surprised by these results, but coming back to the exact way the IMC was presented we found the reason. The instruction was in the 9^{th} row of a grid where people had to select five options in each of three different columns. But in treatments 2 and 3, as soon as the respondents selected one option in the first column, the second column appeared (no check was implemented on the fact that they selected indeed five options). In this already really complex grid, the fact that the next column appeared suggested to the respondents that it was enough to select only one option. Instead of pushing them to pay more attention to the options, on the contrary, it pushed them to go on.

Therefore, we believe that these unexpected results for the progressive treatments are more due to the way the IMC were designed than to the treatment itself. A better

design including a check on the number of items and a second column appearing only once the five items were selected would most probably lead to an improvement. A more classical IMC (for instance, in a traditional battery or alone on one page) would also probably perform better, as it did in Kapelner and Chandler (2010).

Straight-lining

Straight-lining consists in selecting the same answer for all the items in a battery without differentiating between them. This is considered as one of the main indicators of satisficing.

In our survey, a battery of 16 items about drink consumptions with seven answer categories came first. The same battery was repeated with a reversed scale at the end of the survey. Finally, we had a battery of 12 items about several brands of cars using a 5-point agree-disagree scale.

Table 8 gives the percentages of pure straight-liners on 0 to 3 of these batteries. Since respondents can also straight-line not on all items but on almost all items, we consider the percentages of respondents that give the same answer on at least N-2 items in the battery.

Table 8: Pure straight-liners and straight-liners on at least N-2 items in 0 to 3 batteries

	In%	Control		Treatment1		Treatment2		Treatment3	
				Timing control		QuestionScale		Kapcha	
		with	without	with	without	with	without	with	without
Pure	**0**	88.35	89.61	**93.43***	**92.44**	**96.05***	**93.69**	**95.60***	**97.10***
	1	10.84	8.66	6.10*	7.56	3.95*	6.31	4.40*	2.49*
	2	.80	1.73	.47	0*	0	0*	0	.41
	3	0	0	0	0	0	0	0	0
At least N-2	**0**	83.94	86.58	**88.26**	**88.89**	**91.70***	**91.89***	**92.40***	**91.70***
	1	12.45	9.52	8.45	9.78	7.51*	5.86	6.40*	6.64
	2	1.20	3.46	1.41	.89*	0*	.90*	1.20	.41*
	3	2.41	.43	1.88	.44	.79	1.35	0*	1.24

The percentages of respondents that never straight-line are higher for the treatment groups. Considering also the straight-lining on N-2 items, the best results are obtained for treatments 2 and 3. The fact that the rows enter progressively helps reducing the non-differentiation across items in batteries of questions.

Incoherence response

Next, we look at the incoherence across responses for the same respondents. We use the battery of questions about drink consumption since it is repeated with a reversed scale. If respondents give twice the same answer (in terms of labels), then they get a score of 0 for the incoherence variable of the corresponding item. If they answer the second time one category (respectively two,…, *n*) next to the one they chose at first, they will get a score of 0.01 (respectively 0.02,…, 0.01**n*) for the incoherence variable. The scores for the 16 items of the battery are added. Besides, in the battery about brand of cars, there are three reversed items that are meant to be opposite of three others. We use the same system for these items. Table 9 gives the mean final score of incoherence (adding drinks and cars).

Table 9: Incoherence of answers: mean (median) and minimum – maximum

Incoherence	**With test**		**Without test**	
	Mean (median)	Min-max	Mean (median)	Min-max
Control	.1476 (.10)	0-.90	.1521 (.10)	0-.90
Treatment 1-TimingControl	.1261 (.09)	0-.82	.1233* (.09)	0-.79
Treatment 2-QuestionScale	.1150* (.09)	0-.67	.1171* (.10)	0-.59
Treatment 3-Kapcha	.1320 (.10)	0-.83	.1351 (.11)	0-.78

The average level of incoherence is lower in the treatment groups. In particular, treatment 2 is leading to significantly less incoherence in average. However, the median is similar over groups. This suggests that the improvement in the treatment groups consists mainly in a reduction of the very incoherent responses (cf. the maximum values). Nevertheless, lower levels of incoherence are still present.

For open questions, it is also possible to look at the incoherence of the answers to single questions. The questionnaire contains two narrative open questions for which the percentages of respondents answering a non-sense (just a letter or a question mark or a number…) are reported in Table 10. The first question is about suggesting improvements for the survey. The second one asks a list of all the topics they remember from the questionnaire.

Table 10: Percentages of respondents answering a non-sense in open narrative questions

	1st question		**2nd question**	
	With test	**Without test**	**With test**	**Without test**
Control	.80	.85	.80	1.71
Treatment 1-TimingControl	.46	.88	3.24*	.88
Treatment 2-QuestionScale	.39	.90	.78	1.79
Treatment 3-Kapcha	1.19	0	2.76*	.83

Overall, the percentages of respondents answering a non-sense in the two narrative open questions are quite low. There are no significant differences between treatment and control groups for the first question. For the second one, treatments 1 and 3 with test lead to a significantly higher number of non-senses. This is different from what we expected.

Precision of answers in open questions

Finally, we consider the precision of the answers to the same two open questions. We consider more written characters as an indication of higher quality. Table 11 gives the average number of characters for the two questions.

Table 11: Number of characters in open questions: mean (min- max)

	1st question		2nd question	
	With test	**Without test**	**With test**	**Without test**
Control	58.5 (1-621)	50.5 (4-432)	67.0 (1-430)	66.5 (3-281)
Treatment 1-TimingControl	84.9* (4-936)	71.9* (4-321)	81.5* (5-613)	73.1 (5-715)
Treatment 2-QuestionScale	65.9 (2-967)	56.8 (2-611)	68.6 (5-587)	65.0 (4-558)
Treatment 3-Kapcha	76.3* (1-559)	73.7* (4-396)	62.5 (2-296)	72.1 (2-556)

The numbers of written characters are generally higher in the treatment groups, even if the differences with the control groups are not always significant. For the first question, the highest numbers are found for treatments 1 and 3 with test. However, since this question is about improvements of the survey, we should be careful. The average number of characters may be higher in treatments 1 and 3 (especially with test) because the respondents were really not satisfied with the experience and took this question as an opportunity to complain about the survey. We will see in the next section that these groups are indeed the ones with lower percentages of satisfied respondents.

Nevertheless, for the second question there is no reason to believe that the length of the answers relates with the satisfaction with the survey experience. And again, there is a higher number of written characters in average for treatment 1, even if the difference is significant only in the group with test. So it seems that this treatment leads to more precise answers to open questions.

To summarize the results about quality, we observe some positive impact of the treatments on straight-lining (treatments 2 and 3 mainly), incoherence of responses

(treatment 2 mainly) and precision of answers (treatment 1) but not for the IMC (even if this is probably due to a bad design of the question) or for the non-sense in open questions (in most cases, it stays stable). We conclude that overall there are some improvements due to the treatments. However, they may not be as large and systematic as expected. Now, what about the negative effects?

3) Expected negative effects

a) Abandon rate

The first expected negative effect is an increase of the abandon rates. We call abandon rate the percentage of people that start the survey but do not finish it. Table 12 gives these abandon rates.

Table 12: abandon rate

In %	With test	Without test
Control	30.06	34.56
Treatment 1-TimingControl	46.75*	39.84
Treatment 2-QuestionScale	29.53	30.41
Treatment 3-Kapcha	32.98	31.34

The abandon rates are 30-35% in the control groups and very similar in treatments 2 and 3. In treatment 1, they are higher and the difference is statistically significant in the group with test. This is against our hypothesis 3 but in line with what we saw analyzing the results of the test of speed of reading. Since most respondents ended up with a speed slower than 250 words per minute, the group with test is facing stricter conditions while responding the survey which leads to more abandons.

b) More negative evaluation

We included a few questions to evaluate the experience of the respondents in completing the survey. In particular, respondents were asked at the end to auto-evaluate how difficult the questions were.

We expected respondents to say the questions were more difficult when the treatments were used, because they were not used to this way of presenting questions and because we thought they will evaluate as part of the difficulty of the question the design used to ask it. Table 13 reports the auto-evaluation of the respondents.

Table 13: Percentages of respondents evaluating the questions as difficult, middle or easy

In%	**Control**		**Treatment 1** Timing control		**Treatment 2** QuestionScale		**Treatment 3** Kapcha	
	with	without	with	without	with	without	with	without
Difficult (0-4)	3.61	4.33	**5.63**	**9.33***	**5.93**	4.95	**5.20**	**5.81**
Middle (5)	8.03	6.06	6.57	10.22	7.91	7.66	10.00	7.47
Easy (6-10)	88.35	89.61	87.79	80.44*	86.17	87.39	84.80	86.72

Higher percentages of respondents say the questions are difficult in the treatment groups. But the difference is significant only for treatment 1 without test.

Moreover, respondents got an opportunity to complain about the design in the open-question about improvements of the survey. Because they were restricted in their freedom of completing the survey the way they want, we expected fewer respondents to say that everything was ok or that they were completely satisfied with the way the survey was done in the treatment groups. Also, we expected a high number of complaints focusing on aspects related to the treatments and less complaints about other aspects. Table 14 presents the corresponding percentages.

Table 14: Percentages of respondents saying everything ok / complaining about the speed

In%	**Control**		**Treatment 1** Timing control		**Treatment 2** QuestionScale		**Treatment 3** Kapcha	
	with	without	with	without	with	without	with	without
Everything ok	32.00	34.19	19.91*	19.03*	27.91	26.46*	20.55*	25.21*
Complaints speed	4.80	8.12	41.20	28.76	29.46	32.29	36.36	40.50

There are fewer respondents completely satisfied with the survey in the treatment groups. The lowest percentage is in treatment 1. Looking at the complaints about the speed of the survey, even in the control groups we can find some, the percentages of such complaints are clearly much higher in the treatment groups. Treatments 1 with test and 3 without test are the ones leading to more complaints. Treatment 1 without test generated much less complaints. Also, there were fewer complaints in treatment 2, which is in line with our expectations.

To summarize the results about the negative effects, we can say that respondents complain more in the groups with treatments, but they did not evaluate the questions as more difficult (except in treatment 1 without test) and the abandon rate is not higher (except in treatment 1). So mainly for treatment 2 but also for treatment 3, the negative effects are quite limited.

V) Conclusion

In this paper, we wanted to test the effects of forcing respondents to slow down while completing online surveys. We differentiated several ways of doing that: using a minimum time per page and presenting the text of the questions progressively (by blocks or by characters), either when the speed was fixed arbitrarily for all respondents or when it was adjusted by introducing a test of speed of reading at the beginning of the

survey. We developed a set of hypotheses about the expected effects, that we tested using data from Netquest in Mexico.

The results for the six first hypotheses have been presented already so we will just summarize them:

- **H1**: “speeding will be reduced in the treatment groups” is supported by the data
- **H2**: “the quality of answer will be higher in the treatment groups” gets some support but it depends on the indicators considered
- **H3**: “The negative effects are smaller when the test of reading is undertaken” is not supported
- **H4**: “the abandon rate will be higher in the treatment groups” is only supported for treatment 1 with test
- **H5**: “the respondents will evaluate the questions as being more difficult in the treatment groups” is only supported for treatment 1 without test
- **H6**: “there will be more complaints in the treatment groups” is supported

Now what about the last hypothesis **H7**: “overall, there are more positive effects than negative effects”?

We do not have a scientific test to accept or reject this hypothesis. The decision has to be based on the overall judgment of the results presented. In our opinion, there are indeed more positive effects than negative effects observed for this survey. However, all treatments are not equal: treatment 1 is the one for which more negative effects are observed. In this case, it is not so clear that it is worth it using such a treatment. So it is probably better for future research to focus on the two other treatments. These two treatments not only force respondents to slow down but also push them to proceed the

information stepwise. This can help separating the different cognitive steps of reading, understanding, and answering of the question.

The difference with and without test did not lead to the expected results because most respondents by adapting the speed of reading get to a slower overall speed. But this can be due also to the country studied.

We should also notice that if using progressive designs as treatments 2 and 3, we should pay a lot of attention in designing specific kinds of question. The complex grid in which our IMC was included clearly did not work.

Moreover, we should not conclude too quickly about the abandon rates: since the experiment was part of a panel, people may have decided to go on until the end because it was a one-time experiment. But if this would be systematic, maybe they would not make the effort anymore.

More research is needed in order to oversee the limits of this study and to look at the robustness of the results across different surveys and contexts. However, we believe that this experiment was very interesting because it shows that innovative ways of presenting the surveys are feasible in practice: at least punctually, respondents went through the complete survey with limited abandon rates and limited amount of complaints, while the indicators of quality (except for the IMC but this is probably due to its bad design) were either similar or better. Online surveys present huge opportunities to develop interactive designs and to adapt the survey conditions to each respondent as function of their preferences or skills, but more experiments need to be done to test the effects of innovative ways of doing online surveys.

References

Bassili, J. N. (1993). "Response latency versus certainty as indexes of the strength of voting intentions in a CATI survey". *Public Opinion Quarterly*, 57, 54–61.

Bassili, J. N., and B.S. Scott (1996). "Response latency as a signal to question problems in survey research". *Public Opinion Quarterly*, 60, 390–399.

Couper, M.P. (2000). "Web surveys: A review of issues and approaches". *Public Opinion Quarterly*, 64, 464–494.

Donders, F.C. (1868). "Over de snelheid van psychische processen". Onderzoekingen gedaan in het *Physiologisch Laboratorium der Utrechtsche Hoogeschool*, 1868–1869, Tweede reeks, II, 92–120.

Kapelner, A. and D. Chandler (2010). "Preventing Satisficing in Online Surveys". In *Proceedings of 2010 CrowdConf*, 2010.

Krosnick, J.A. (1991). "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys". *Applied Cognitive Psychology* 5:213-36. doi: 10.1002/acp.2350050305

Krosnick, J. A., and D.F. Alwin (1987). "An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement". *Public Opinion Quarterly*, 51(2):201-219. Stable URL: http://www.jstor.org/stable/2748993

Malhotra, N. (2008). "Completion time and response order effects in web surveys". *Public Opinion Quarterly*, 72(5), 914-934.

Oppenheimer, D.M., Meyvis, T. and N. Davidenko (2009). "Instructional manipulation checks: Detecting satisficing to increase statistical power." *Journal of Experimental Social Psychology* (2009) 45:867–872. doi:10.1016/j.jesp.2009.03.009

Revilla, M. and C. Ochoa (2013). "What are the links in a web survey between response time, quality and auto-evaluation of the efforts done?" RECSM Working Paper 36.

Schuman, H., and S. Presser (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context.* New York: Academic Press.

Simon, H.A. (1957). *Models of Man*. New York: Wiley

Strack, F., and L.L. Martin (1987). "Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitude Surveys". In *Social Information Processing and Survey Methodology*, ed. H.J. Hippler, N. Schwarz, and S. Sudman, pp. 123-48. New York: Springer.

Tourangeau, R., and K. Rasinski (1988). "Cognitive Processes Underlying Context Effects in Attitude Measurement." *Psychological Bulletin* 103:299-314.

Tourangeau, R., Rips, L.J., and K. Rasinski (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Yan, T., and R. Tourangeau (2008). "Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times". *Applied Cognitive Psychology*, 22: 51–68, DOI: 10.1002/acp.1331

Zhang, C. (2013). "Satisficing in web surveys: implications for data quality and strategies for reduction". Doctoral dissertation University of Michigan 2013. Available at: http://deepblue.lib.umich.edu/handle/2027.42/97990