Survey Sampling International
info@surveysampling.com • surveysampling.com

Questionnaire Length, Fatigue Effects and Response Quality Revisited

By Pete Cape, Global Knowledge Director

Introduction

At the first ESOMAR Panel Conference in 2005, the award for best paper went to Sandra Rathod and Andrea la Bruna for their paper "Questionnaire Length and Fatigue Effects." The conclusions they drew are as relevant today, if not more relevant, as they were then.

Their main conclusions were:

- Response rate does not depend on interview length.
- Longer surveys do not necessarily mean increased drop-out; most drop-out has occurred by the halfway stage, irrespective of interview length.
- Respondents get fatigued, pay less attention to the task in hand and increase their speed of response as the interview progresses.
- Data quality suffers as interview length increases.

For many, 2004 (the year the research was conducted) seems light years away. The world, especially the online world, has moved on apace. Tim O'Reilly had already coined the term Web 2.0 in 2004 but the impact in terms of how users viewed and used the web had yet to emerge; no YouTube, no Facebook, no Twitter, not even Lego mindstorms.com.

Most would agree that people now have many more things to do online than they did in 2004, things that are arguably more interesting than market research surveys. The amount of leisure time available has not increased so attention spans must have declined. By "attention spans" I mean the amount of time given to a particular task, not the individual's ability to concentrate on a given task.

Research Design

In order to more fully understand the possible effects of survey length, fatigue and subsequent response quality, two surveys were designed: a long version and a short version. The long survey was designed first and the shorter version was simply an extract of it. Each of the surveys utilized a block design and was divided into four blocks of questions, each representing a different subject matter. The blocks were randomized for each respondent so that the effect of survey length on response quality could be investigated by comparing whether the different order of the blocks led to different response patterns as the block position varied in the survey.

In August 2009 we re-fielded the same survey using the same sample specifications and, as best we could, the same practices as were common in 2004. Some minor changes were inevitable;

we had to change the charity, for example. For the most part, the survey was conducted, warts and all, exactly as it had been in 2004. We were unable to take advantage of innovations like SSI Verify to weed out potentially undesirable, fraudulent respondents and did nothing in data analysis to exclude inattentive participants.

Response Rates

It will come as little surprise to most to see how response rates have collapsed over the past 5 years. The best country response rate in 2004 was the Netherlands at 59% response for the short version of the survey, and a commendable 49% for the longer survey. By 2009, these dropped to 13% for both surveys. While the Netherlands still boasts the best response rates of all three countries under test, the marked inter-country differences seen in 2004 are no longer visible.

***Table 1**–Response Rates*

	France		UK		Netherlands		Average	
	Long	Short	Long	Short	Long	Short	Long	Short
Response Rate 2004	24%	24%	18%	10%	49%	59%	30%	31%
Response Rate 2009	11%	11%	5%	7%	13%	13%	10%	10%

In both 2004 and 2009, Response Rates do not depend on the length of the interview, at least not at these interview lengths.

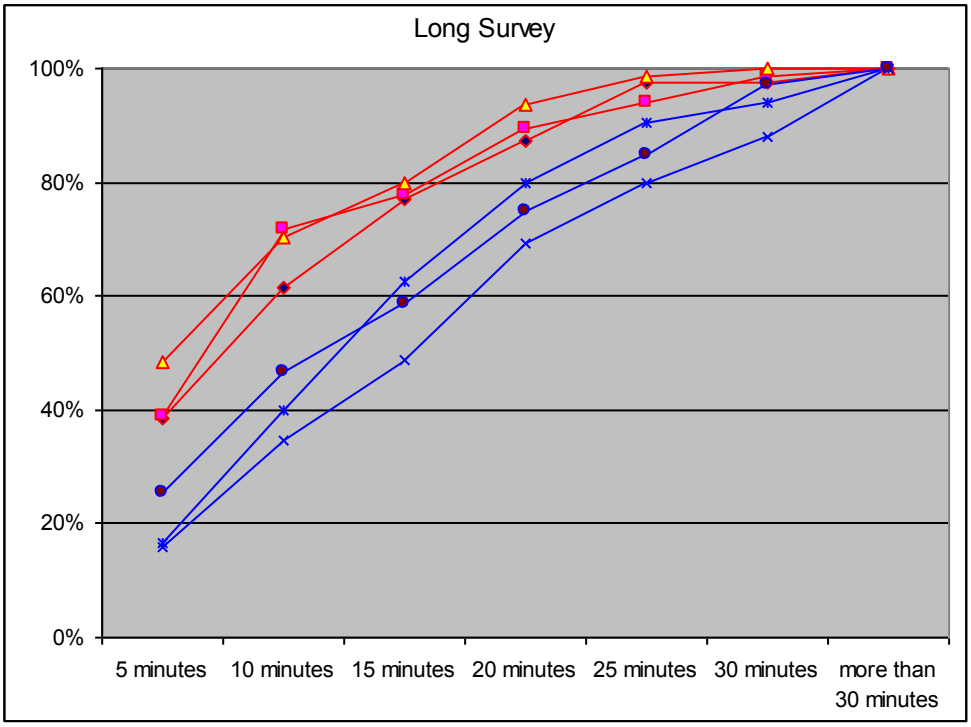
Drop-out Rates

Although response rates have plummeted over the period, drop-out rates have remained remarkably constant. This attests, perhaps, to people having the same level of attention span, or at least the same level of commitment to complete a task once started.

***Table 2**–Drop-out Rates*

	France		UK		Netherlands		Average	
	Long	Short	Long	Short	Long	Short	Long	Short
Drop-out Rate 2004	27%	25%	26%	19%	24%	30%	26%	25%
Drop-out Rate 2009	30%	23%	26%	37%	17%	17%	24%	26%

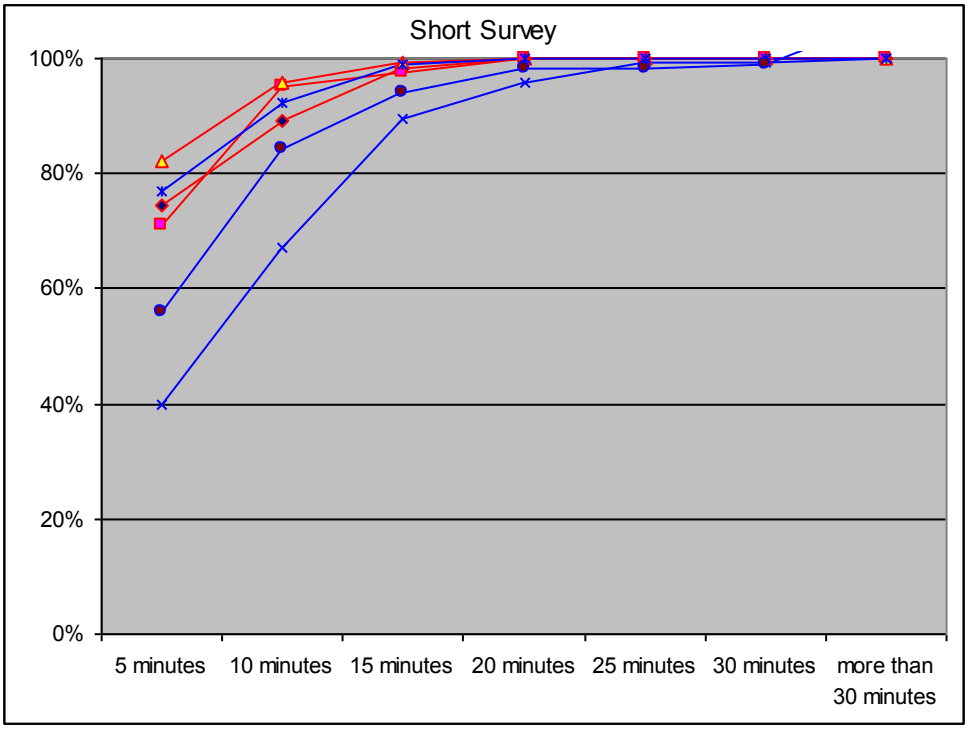
***Chart 1**–Cumulative Drop-Out by Country and Year (Blue 2004, Red 2009)*



In 2004, most drop-out had occurred by the halfway point of the survey. This remained true in 2009. Chart 1, for the long survey, appears to show a different cumulative distribution of drop-outs but in fact this is a function of the interview lengths having decreased overall and quite dramatically in some cases. (See next section.)

Even on the short survey, panelists shaved off between 33% (France) and 14% (the Netherlands) of the time they took to complete the survey. In the short survey, we still see the majority of drop-out occurring by the halfway point.

***Chart 2**–Cumulative Drop-Out by Country and Year (Blue 2004, Red 2009)*



Interview and Block Length

As mentioned previously, overall interview lengths were substantially reduced in both the long and short versions of the survey. It might be argued that panelists now are more experienced in survey taking than their predecessors, since they are asked to do more surveys during their tenure on the panel. But panel tenures are much shorter now than they were in 2004. Panelists completing surveys in 2009 were no more experienced (at least on OpinionWorld) than their predecessors in 2004.

***Table 3**–Average Interview Lengths (minutes)*

	France		UK		Netherlands	
	Long	Short	Long	Short	Long	Short
2004	37	18	33	15	30	14
2009	25	12	23	11	24	12

One of the hypotheses of the 2004 study was that respondents would take less time and exert less effort on the later parts of the questionnaire than they did on the earlier parts, due to

fatigue. This hypothesis was proven in 2004. By employing a block-rotation design, we see that as the same block of questions are moved further back in the study, the time taken to complete them gradually reduces.

One might argue that the decrease in block completion time could be due to increased familiarity with the question set. It is true that the question blocks were similar in their construction and contained somewhat similar questions. The evidence presented in the next section (Panelist Fatigue) shows that at least some of the increased speed was due to fatigue.

In 2009, the picture is somewhat different. The block completion times per position are, in 3 out of the 4 sections of the long survey and for all of the short survey, statistically significantly different. However they do not follow the expected pattern, as can be seen in tables 4 and 5.

***Table 4**–Block Mean Completion Times (minutes): Long Survey 2009*

	Position 1	Position 2	Position 3	Position 4
Section A	5.78	6.94	6.67	5.82
Section B	7.07	6.71	7.36	6.91
Section C	6.1	5.26	5.09	5.31
Section D	5.07	4.44	4.08	4.02

***Table 5**–Block Mean Completion Times (minutes): Short Survey 2009*

	Position 1	Position 2	Position 3	Position 4
Section A	2.76	3.74	3.34	3.52
Section B	1.83	1.69	2.06	1.89
Section C	2.06	1.72	1.5	1.48
Section D	2.12	1.79	1.68	1.58

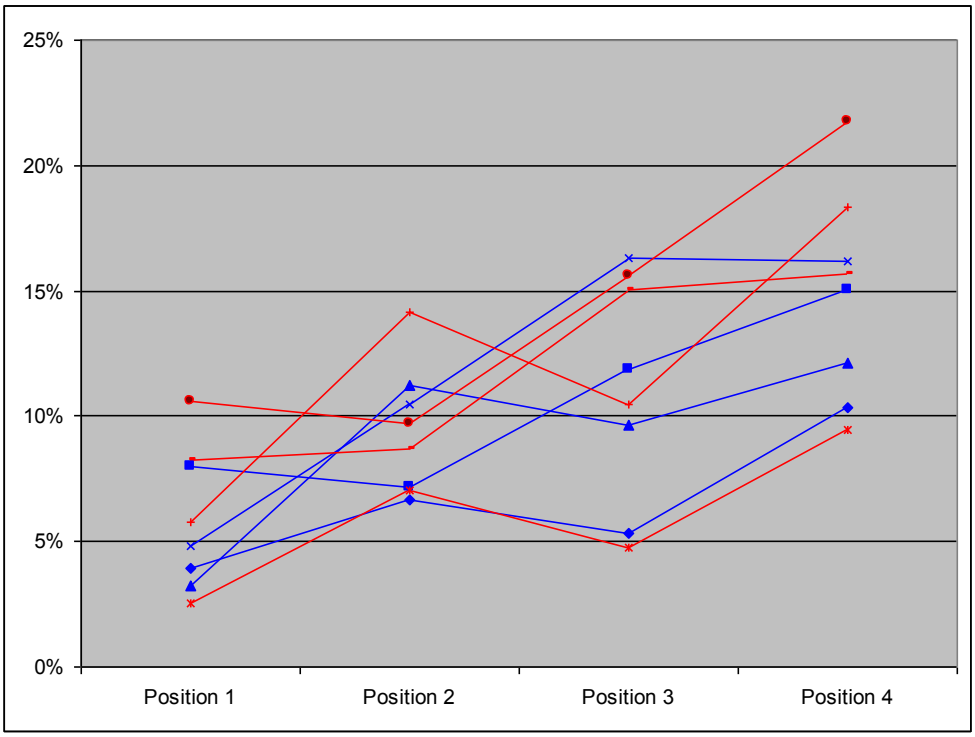
As we noted earlier, average overall interview lengths were much shorter in 2009 than 5 years ago. This implies perhaps that the panelist, in contrast to his 2004 counterpart, starts fast and stays fast throughout the survey.

Panelist Fatigue and Satisficing

One of the behavior outcomes of cognitive fatigue is satisficing, which is defined as doing just enough work to satisfy the task. In 2004, the authors looked to a question in each block that it was possible to skip to see if this behavior was present. This question was a set of scales, presented in the form of sliders, which measured involvement with the subject matter of the block. The slider bar was positioned at the mid-point so it was possible to click on "next" without moving the slider and still leave some data behind. It was demonstrated that the likelihood of skipping the question rose as the skippable question was encountered further and further into the questionnaire. In 2009, we found precisely the same pattern. The first time the skippable question was encountered, it was more likely to be completed than on subsequent occasions when it was seen.

This was particularly true for the long survey as can be seen in Chart 3. The individual lines represent the 4 different sections and the colors represent the years.

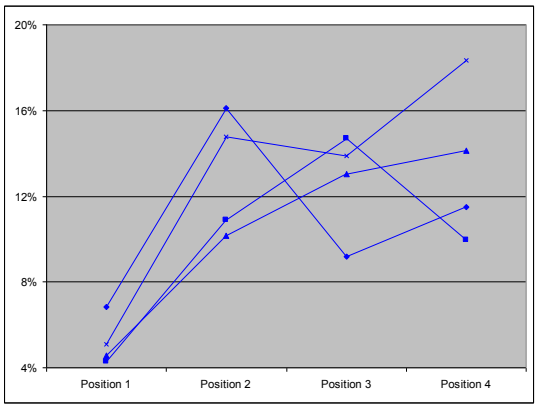
***Chart 3**–Instrument Non-Response by Position: Long Survey (Blue=2004, Red= 2009)*

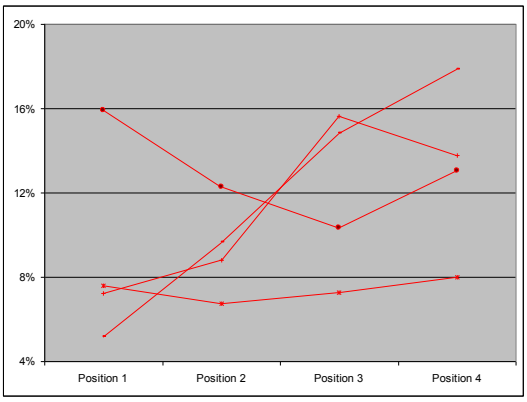


Generally speaking, on the long survey non-use of the slider increases as the survey progresses. There are no discernable differences between 2004 and 2009.

On the short survey, particularly in 2009, the effect is less dramatic and less clear. In only two of the section types do we see the pattern.

***Chart 4**–Instrument Non-Response by Position: Short Survey (Blue=2004, Red=2009)*





What is clear is that there is satisficing behavior in evidence, particularly on the long survey. The reduction in elapsed survey times does not mitigate the effect. The long survey, at nearly 25 minutes, is still too long.

Rathod and la Bruna, working perhaps in more innocent times, did not actively look for any fraudulent behavior. One of the sections gives an ideal opportunity to look for deliberate misrepresentation. The entire short break holiday section was skipped if the respondent claimed not to have been on a short break holiday. Since respondents were randomly assigned to see this block either 1st, 2nd, 3rd or 4th we should expect to see the incidence of qualification to be the same irrespective of where it was positioned, given a certain latitude for sampling error. Furthermore, we should expect to see the same incidence in both the long and short surveys for the same reason.

Examination of the data shows that this is not the case now and was not the case in 2004. The highest incidence of qualification for the short break's block of questions was 68%. This occurred on the short survey when short break holiday was positioned first and, of course, the respondent was at his or her freshest. The level of qualification declined at each position further into the questionnaire to finally reach 50% when the block was positioned 4th. The same phenomenon occurred on the long survey. Qualification started out at 64% and finished at 47%.

By 2009, we did not expect the incidence of short break holiday takers to change substantially. The question did not ask "Did you take a short break holiday this year?" but rather "When was the last time you took a short holiday break?" It was the equivalent of ever having taken such a holiday.

Within the long survey, the starting qualification level was 6% lower than it had been 5 years previously—not a significant finding. In the 2009 long survey, the pattern observed in the 2004 data was repeated; there was a decline in qualification rates as the question block moved

back in the survey. The difference between qualification at position 1 and qualification at position 4 is statistically significant at the 95% confidence level in both surveys. It is hard to come to any conclusion other than that a small minority will look for ways to shorten their survey taking experience.

On the 2009 short survey, the pattern was less clear. The qualification rate was statistically significant at 17% below that of the 2004 short survey for the question block when placed in first position. However, the qualification rate increased and decreased as the block changed position. In fact, none of these differences in qualification found in position of the question prove to be statistically significant at the 95% confidence level.

***Table 6**–Qualification for Short Break Holiday Section*

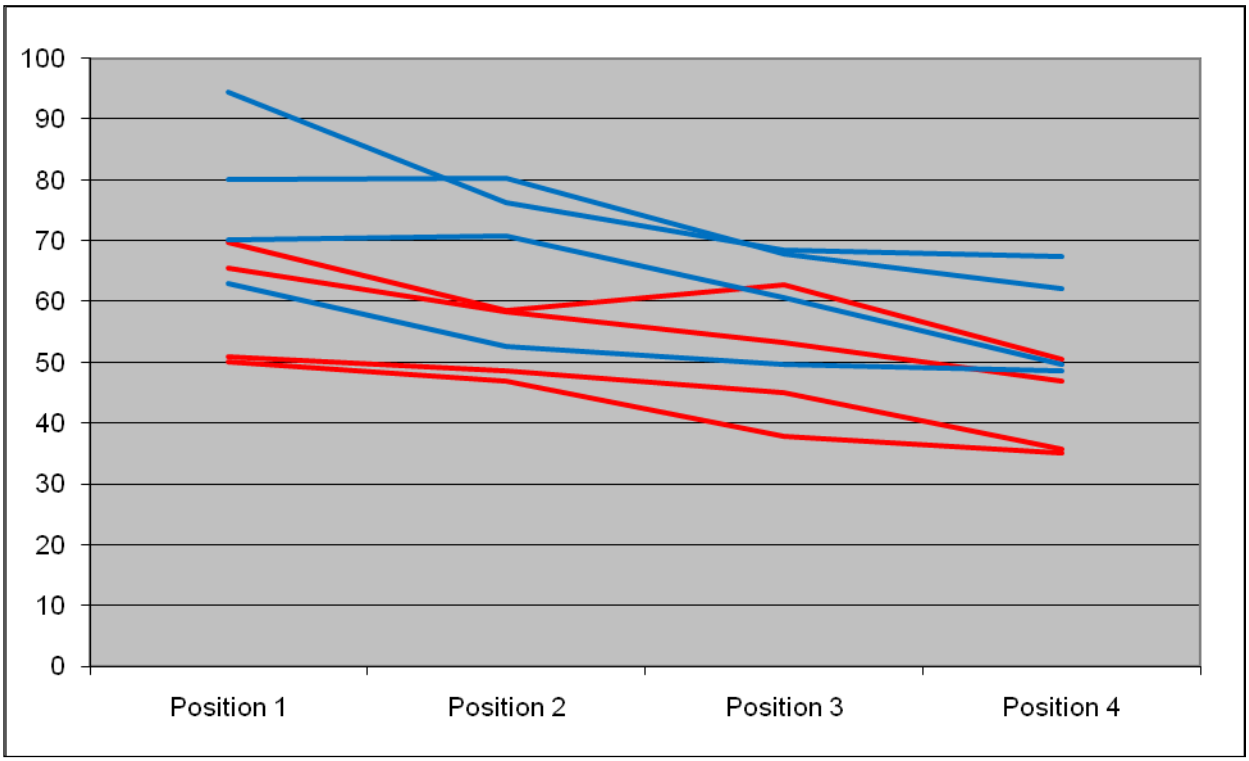
	Position 1	Position 2	Position 3	Position 4
2004 Long Survey	64%	60%	56%	47%
2009 Long Survey	58%	52%	46%	47%
2004 Short Survey	68%	55%	58%	50%
2009 Short Survey	51%	60%	53%	55%

The 2009 short survey is 3 minutes or so shorter than the 2004 short survey. Perhaps this brings it down to an acceptable interview length for respondents so they do not feel the need to cheat their way through a faster route.

Data quality can also be measured by engagement in open questions. There should be no real difference between the number of words, or number of characters, used in the same open question when asked first or last—it is the same question.

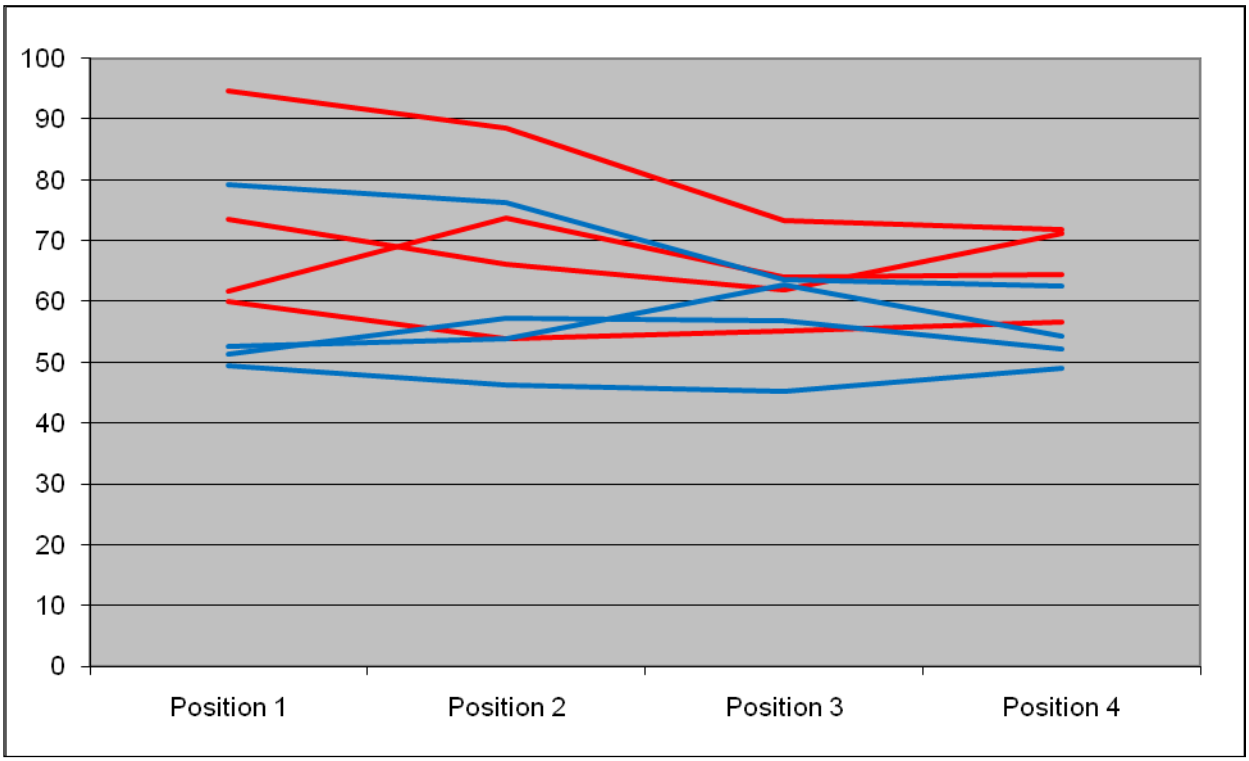
From the results of the long survey, two things are very clear. First, the number of characters typed decreases as the open question moves back in the survey. Second, the number of characters per open question has decreased in the past 5 years.

***Chart 5**–Number of Characters per Open Question: Long Survey (Blue=2004, Red=2009)*



In the short survey the pattern is less clear, although certainly flatter.

***Chart 6**–Number of Characters per Open Question: Short Survey (Blue=2004, Red=2009)*



Again, the short survey seems to be an acceptable length compared to the long survey.

Conclusions

In both 2004 and 2009, the long survey proved itself too long. It fatigued the respondent and led to satisficing behavior. Where questions could legitimately be skipped, they were. The answers to open questions were less full, and less time and effort generally was put into the end of the survey compared to the start.

Perhaps the most unsettling finding was that the instances of cheating, deliberately telling a falsehood in order to skip an entire section, also increased as the survey progressed.

Following additional analysis of the data after 2004, Rathod concluded:

"This research indicates that there is a critical point in online survey response when the fatigue effects become significantly more pronounced. That critical time is around the 20 minute mark. If researchers work to keep surveys shorter, it will not only help ensure response quality, but it will also make for more motivated and responsive respondents."

Survey lengths of 25 minutes (and longer) are not uncommon. In many quarters, they would be considered normal or even fairly short. Perhaps our inability, or unwillingness, to act on Rathod's warning at the ESOMAR Panel Conference in Budapest has contributed to the general decline in response rates and data quality in the intervening period. We are the ones who have continued to foist over-long surveys onto already overburdened respondents. When we seem to care so little for their well-being, is it any wonder that they care less about the quality of the work they do for us?

It is worth remembering that the only reason we can see the data quality effects is that we have a block rotation design in this experiment. If we did not, then how would we judge the quality of the responses we got later in the survey? Respondents (especially online panelists) tend to soldier on to the end. Is it their fault that they are tired and cannot think as clearly as we would like them to? In the "old days" of telephone and face-to-face interviewing, the interviewer would hear when respondents became tired and started to satisfice. They would take pity on them, perhaps dropping out of interview mode for a moment or two, providing a mental tea-break if you like, before going back to the task in hand. Where in our online surveys do we offer our respondents the same courtesy?

It appears from the 2009 research that our current crop of respondents is not as diligent as respondents were in 2004. It is hard to believe that they are hugely different people; that it is actually we who have caused this situation. The survey we presented them was certainly not "best-in-class" given everything we know now about survey design. This was deliberately so to maintain comparability. We'd like to think that the respondents treated it in part with the disdain it deserved rather than being inherently less involved and engaged than their forebears.

Interview lengths of 20 minutes or less can produce wonderful and engaged responses if well designed. The fact that we saw very much less satisficing and cheating in the short survey attests to this. We make no apology in repeating the conclusion of 2004:

"If researchers work to keep surveys shorter, it will not only help ensure response quality, but it will also make for more motivated and responsive respondents."

Originally presented at ARF Re:think 2010