DISQO
DISCOVER WHY

Evaluating Open-Ended Responses: One More Tool in the Quality Toolbox

August 2018

www.disqo.com

In the marketing research industry, there is a lot of debate regarding how to best measure the quality of online survey data. Multiple checks are typically employed to identify potentially suspicious behavior or poor-quality data. Verbatim responses to open-ended questions are often included among the checks to remove respondents from data sets. Yet, some respondents are less inclined to provide robust answers to open-ended questions. Does that mean they are unengaged or "bad" respondents and that their data is not valid?

Given the importance of data quality, DISQO conducted research to better understand the implications of using open-ended questions in conjunction with other quality checks. The research helps determine which respondents are of concern and provides guidelines to identify them.

Key Findings

Responses to open-ended questions are an important tool in the toolbox to assess quality; however, they should not be used as the sole criterion for identifying respondents with suspicious data quality. Rather, they should be used in combination with other quality checks such as red herrings, straight-lining, speeding, etc. Eliminating respondents based solely on their responses to open-ended questions can mean throwing away valid data.

There are several things to keep in mind when using open-ended responses for quality checks:

- The focus should be on respondents who provide answers of no substance and/or nonsensical answers when flagging comments for poor quality. Even in these cases, respondents should only be eliminated if they have failed other quality checks.

- "No opinion" answers such as nothing, n/a, and don't know are, more often than not, valid answers and therefore, should not be flagged for poor data quality.

- Word count should not be used to judge the quality of open-ended responses. Short answers, even if consisting of only one word, are often valid.

- Short answers are even more prevalent on smartphones and tablets. This makes sense since respondents using their mobile devices are conditioned to provide quick, abbreviated communication as in text messages. Short open-ended responses are likely to become more common as the proportion of respondents taking surveys on their mobile devices continues to grow.

- Survey length can impact the quality of open-ended responses. Later placement in surveys with longer LOIs can lead to a degradation in responses due to fatigue. With shorter survey lengths, placement of the open-ended questions does not necessarily impact the quality of the responses.

This research highlights the quality of the DISQO Panel. Only 3% of respondents fail the quality check for open-ended responses. The quality of DISQO panelists is also evident among those taking the survey via smartphone with only 1% failing the open-ended quality check.

Research Design

DISQO fielded a survey among 1,594 of its consumer panelists in April 2018. The sample was comprised of US adults age 18+. Controls were in place to ensure the sample was representative of the general population based on age, gender, region and income. Respondents could complete the survey on any device (desktop/laptop, smart-phone, or tablet).

Panelists participated in one of two versions of the survey:

- A short version with a median length of 9 minutes
- A long version with a median length of 16 minutes

The short version was a subset of the questions of the long version. The topic focused on device usage and habits. Both versions included questions regarding devices used, technology savviness, television viewing habits, and "internet of things" usage. The long survey version also included questions about social media, communication methods, mobile phone usage, tablet ownership, and online shopping habits.

Both survey versions included the same open-ended questions. Respondents were required to enter at least one character to progress to the next question. The first open-ended question occurred in the first two minutes of the survey at the same location in both versions.

> What do you ***like and dislike*** about your *<insert device used most often from previous question>* that you use most often to access the internet?

The second question asked opinions about "the internet of things" and was shown much later in the long survey version than the short survey version. Respondents received one of two questions depending on whether or not they already control items in their home via an app. Since each respondent only saw one of these questions and both fall in the same location in the survey, the data has been combined for analysis.

> *What do you like and dislike about controlling items in your home via an app?*
>
> *Why are you <insert interest level from previous question> in controlling items in your home via an app?*

Categorizing the Quality of Open-Ended Responses

Each open-ended response has been examined and coded into four possible categories:

1) Quality answer – Responses that answer the question and are relevant.
2) No opinion answer – Responses such as nothing, n/a, not sure, and don't know. In many cases, these are likely valid answers, but are coded separately to understand the magnitude.
3) Non-substantive answer – Generally 1 to 2-word responses that don't answer the question such as good, cool, nice, yes, OK, or even foul language.
4) Nonsensical answer – Consist of random letters and/or numbers such as abc, ghfhgh, and 5745. These answers are often a quick way for respondents to proceed to the next question.

To provide further clarification, examples of open-ended responses deemed as "Quality" from the DISQO panelists are shown below. The responses pertain to the question regarding reasons for interest level in controlling items in your home via an app (i.e., major appliances, thermostat, lighting, locks, irrigation system, security system speakers and television sets).

	EXAMPLES OF QUALITY RESPONSES
"Extremely interested"	• "It would save energy and time when I am away to adjust the air. Or turn on and off lights". • "It would be convenient to control items away from home. No more wondering if I left something on."
"Very interested"	• "It would make things easier but it's a bit pricey to set up." • "I would really like to be able to voice control the lights and television since I very nearly always have a cat on my lap or laying across my chest in bed."
"Somewhat interested"	• "It would be nice to be able to turn off an appliance that I forgot about or to check on the security of my home while I am away." • "I think it's a cool idea, but you also have to invest in a lot of technology to do it."
"Not very interested"	• "What if my phone broke or lost and then I can't control it." • "Small apartment, I can walk to what I want to control."
"Not at all interested"	• "Using apps for nearly everything is a sign of laziness and allowing something else to have control over your life. I want to be in control." • "Concerns about security, privacy. Concerned about the potential for the device to be monitored or hacked into."

Quality of DISQO's Panelists

It is important to understand the distribution of the categories of the open-ended responses. The majority of open-ended responses provided by DISQO panelists are deemed quality answers. Only 1-2% of respondents give non-substantive or nonsensical answers for each of the open-ended questions. Those giving non-substantive or nonsensical answers are more likely to be male. Those giving non-substantive answers are also more likely to be age 18-39.

Table 1: Percent of Respondents Failing by Each Open-Ended Response Category

(n=1594)	First Open-End	Second Open-End
Quality answer	92%	91%
No opinion answer	5%	5%
Non-substantive answer	2%	2%
Nonsensical answer	1%	2%

Caution should be used in removing respondents if they aren't showing a pattern of behavior that brings into question the quality of their data.

- Respondents who raise the most concern with open-ends are those giving non-substantive or nonsensical answers to both open-ended questions (Pink highlighted cells). This is 42 respondents (3%).
- Those who give a quality answer to at least one open-ended question should be given the benefit of the doubt (blue highlighted cells).
- It is a little more difficult to decide how to handle respondents who give no opinion answers to both questions, or a no opinion and a non-substantive (or nonsensical) answer (light blue highlighted cells) because no opinion answers can be valid answers. This is an additional 28 respondents (2%).

Table 2: Comparison of Response Categories by First versus Second Open-Ended Question

(n=1594)		First Open-End			
		Quality answer	No opinion answer	Non-substantive answer	Nonsensical answer
Second Open-End	Quality answer	1402	48	5	0
	No opinion answer	49	22	4	0
	Non-substantive answer	17	2	16	1
	Nonsensical answer	3	0	5	20

Before final decisions are made on how to handle each of these, it is important to determine if these respondents are flagged for any other suspicious behavior checks.

Suspicious Behavior Beyond Open-Ended Quality Checks

In order to understand the relationship between open-ended quality checks and other quality checks, seven additional suspicious respondent checks have been included in this survey. (See the Appendix for more details on these seven checks.) For this analysis, the focus is on three groups identified through their open-ended responses:

1) Respondents giving non-substantive or nonsensical answers to one or both open-ended questions.
2) Respondents giving no opinion answers to one or both open-ended questions.
3) Respondents giving quality answers to both open-ended questions.

Respondents giving non-substantive or nonsensical answers to the open-ended questions are most likely to exhibit other suspicious behavior.

Table 3: Quality Check Failures by Each Open-Ended Response Category

	Non-substantive or nonsensical	No opinion answer	Quality answer	Total
Total	73	119	1402	1594
Low incidence	15%	6%	5%	5%
High incidence	18%	11%	1%	3%
Red Herring	6%	3%	1%	1%
Straightlining	18%	9%	2%	3%
Number of items	21%	2%	1%	2%
Convergent validity	7%	1%	5%	5%
Speeding	15%	12%	2%	3%

Many respondents are only flagged on one suspicious behavior check (38% for respondents giving non-substantive answers/nonsensical and 20% for those giving no opinion answers). In fact, 40% of those giving non-substantive or nonsensical answers aren't flagged for any other suspicious behavior checks. This number increases to 69% for those giving no opinion answers which substantiates that these can be valid answers.

Table 4: Number of Quality Checks Failed by Each Open-Ended Response Category

	Non-substantive or nonsensical	No opinion answer	Quality answer	Total
Total	73	119	1402	1594
Flagged on 0 checks	40%	69%	86%	83%
Flagged on 1 check	38%	20%	12%	13%
Flagged on 2+ checks	22%	11%	2%	4%

In this research study, we include the open-ended quality check plus the seven checks above to identify respondents whose data is deemed suspicious. Since very few of the respondents who give a no opinion answer fail other quality checks, respondents are only flagged if they give non-substantive or nonsensical answers to both open-ended questions.

When looking at all eight quality checks, only 74 respondents are flagged on two or more checks representing 5% of the sample. Due to the small number of respondents identified as providing poor quality data, it can be surmised that cleaning these respondents from the data set would not have a dramatic impact on the data. In practice, it would be up to the researcher to make the final determination of how to handle these respondents.

Using Word Count

The word count for each open-ended response has been calculated. Some respondents provide very short open-ended responses; however, the responses answer the question and provide important insight. For example, when asked what they like and dislike about the device they use most often to access the Internet, multiple respondents responded with single words such as convenience, size, and slow.

Females are more likely than males to give longer answers to open-ended questions. This is true for the first and second open-ended questions in both questionnaire versions.

Table 5: Word Count for Each Open-End by Survey Length

	First Open-End		Second Open-End	
	Long	Short	Long	Short
Total Base	794	800	794	800
1 word	12%	13%	16%	14%
2-3 words	15%	19%	21%	24%
4-5 words	17%	13%	18%	17%
6-10 words	25%	24%	23%	25%
11-15 words	15%	14%	11%	10%
16-20 words	8%	8%	7%	6%
21+ words	9%	9%	5%	6%
Average Words	8.96	8.67	7.15	7.21

Respondents who give short answers (1 to 3 words) to the early open-ended question don't necessarily give short answers to the later open-ended question. In fact, 36% give a longer answer to the later open-ended question.

Table 6: Word Count for Second Open-End Among Those Giving 1-3 Word Answers to First Open-End

		1-3 Words First Open-End (n=466)
Number of Words Second Open-End	1 word	30%
	2-3 words	35%
	4-5 words	16%
	6-10 words	14%
	11-15 words	4%
	16-20 words	1%
	21+ words	1%

Since many of the short word count answers are valid, DISQO does not recommend using word count as a means of judging the quality of open-ended responses and suspicious behavior.

Impact of Length of Interview

There is some additional risk to asking open-ended questions later in a survey. The open-ended responses in the longer version exhibit slightly higher levels of poor quality in the later question. For both the short (median LOI of 9 minutes) and long (median LOI of 16 minutes) versions of the questionnaire used in this research, the question asked later contains fewer average words than the question asked earlier.

Table 7: Open-Ended Response Category for Each Open-End by Survey Length

	First Open-End		Second Open-End	
	Long	Short	Long	Short
Total Base	794	800	794	800
Quality answer	92%	93%	90%	93%
No opinion answer	4%	5%	5%	4%
Non-substantive answer	3%	1%	3%	2%
Nonsensical answer	2%	1%	2%	1%

Word count may not be indicative of the quality of a respondent, but it can be indicative of respondent fatigue. Thus, asking open-ended questions late in a longer survey will likely lead to more poor quality open-ended responses.

Impact of Device

The device used to answer the survey impacts the open-ended word count. With both open-ended questions, the average word count is higher for those answering via desktop than via smartphone or tablet.

Table 8: Average Word Count by Device

	Average Word Count	
	First Open-End	Second Open-End
Desktop	9.37	7.95
Smartphone	8.27	6.31
Tablet	7.69	6.02

The proportion of respondents giving non-substantive or nonsensical answers to both open-ended questions is slightly higher for desktop than for smartphone (4% versus 1%). There is also a relationship between survey length and device, with fewer respondents completing the long survey via smartphone.

Table 9: Device Used to Complete the Survey by Survey Length

	Long	Short
Total Base	794	800
Desktop	58%	52%
Smartphone	31%	40%
Tablet	11%	8%

About DISQO

DISQO provides Opinion and Behavior data to deliver the most complete view of the consumer. DISQO empowers its clients with accurate and reliable first-party data to improve business decisions, drive insight, strategy and overall value. It has developed the highest quality single-source consumer research panel which utilizes a human-centric approach that engages people to share data. The company provides data to the world's largest market research and analytics companies to help them discover the "Why" behind consumers' opinions and behaviors.

Visit www.disqo.com

Appendix

The suspicious behavior checks used in this analysis are summarized below. While the use of multiple checks is recommended, all of these checks are not necessary for every survey. The appropriate checks can be selected based on the questionnaire design.

Low incidence

Respondents are asked which product categories have been used in their household in the past 12 months. The list includes three low incidence categories and four high incidence categories. Fraudulent respondents are more likely to claim they have used multiple low incidence categories to try to qualify for the study. Respondents claiming to use two or three of the low incidence categories have been flagged.

High incidence

This quality check uses the same question as the low incidence check. Inattentive respondents are more likely to not check something that most people use because they are rushing through the survey. Respondents claiming to use none of the high incidence categories have been flagged.

Red herring

In a brand usage question, two fake brands (red herrings) have been included. Fraudulent respondents are more likely to claim they have used the red herring brand to try to qualify for the study. Respondents claiming to have used either of the red herring brands have been flagged.

Straight-lining

A respondent straight-lines when they give the same rating to all items in a grid question. The use of straight-lining as a quality check needs to be done with caution. Depending on the items and the scale, straight-lining is often a legitimate response especially down a neutral middle point. The grid in this survey has a mix of both positively and negatively worded attributes. Respondents straight-lining across all of the attributes have been flagged.

Number of items chosen

When asked to "check all that apply," some respondents will check everything or nearly everything. This is often done in order to try to qualify for a study. In some cases, this can be valid, but doing this across several different questions is less likely to be valid. The number of answers given to seven different "check all that apply" questions has been calculated. For each question, respondents have been grouped into "gave an extreme number of answers" or "not". Respondents who give extreme answers on three or more of the seven questions have been flagged.

Convergent validity

With convergent validity, similar questions should be answered in a similar fashion (e.g., if someone agrees to one item then they should also agree to another item). Four pairs of questions have been compared to make sure the answers match. Respondents have been flagged when their answers don't agree on two or more of the sets of questions.

Speeding

Respondents who rush through a survey may be answering inattentively. There are no hard and fast rules on what constitutes too fast. Respondents who completed the survey in less than half the median time for each survey length have been flagged.