



Who Are You Talking To?

Three Ways to Maximize Data Quality When Using Online Panels

July 2019

Zeldis Learning Center

Zeldis Research Associates
1230 Parkway Avenue, Suite 311
Ewing, NJ 08628
Phone: 609-737-7223
www.ZeldisResearch.com

A ZELDIS WHITEPAPER

© 2019 Zeldis Research Associates All rights reserved.

We all know that using panel samples isn't ideal; panels have limits in terms of respondent validation, quality, and representation.

But we live in the real world, where panel is still "king" and often the most realistic and cost-effective way to field an online survey. So if you need to use a panel to field your sample, here are some steps you can take to optimize the quality and validity of your survey results.

01

Panel Selection

All panels aren't created equal. Make sure to vet the panels you are using. Some of the most important questions to ask center on how panelists are recruited and cleaned. It is important that panels constantly clean out bad respondents and refresh with new ones.

Panel quality is even more important in B2B work. Ideally, you want to find panels that recruit in creative ways (e.g., partnerships with professional organizations or reward programs) and don't just screen consumers for their occupation. In addition, there are some panels that actively validate B2B samples, either using publically available data or by verifying through LinkedIn profiles. We've found the data quality from validated samples to be much higher than from others.

Example: We often conduct surveys with HR/employee benefits managers. One panel company we work with partners with the major association of HR professionals in recruiting their panel, resulting in much higher data quality.

02

Programming and Sampling

When creating your sampling plan, know that panels are not naturally representative of the general population. Make sure to work with panel providers to generate representative sample sends, and if needed, create quotas to make sure that your data reflects the larger population. Use Census data to ensure your results reflect the population on factors such as gender, income, age, etc. If you need to have quotas that skew your data to ensure a readable base in a particular subgroup, you can use weighting to make sure your total reflects the population as much as possible.

[more...](#)

Programming and Sampling, continued...

There are several ways you can program in better data quality in your survey:

- One of the most commonly used data quality checks is to terminate “speeders” (we typically eliminate those who complete in less than one-third of the median time).
- Use programmed “honeypot” traps that will trick and identify bots that may be going through the survey.
- To ensure that respondents are truly paying attention, create one or two questions where respondents need to select a specific answer (e.g., “please select response 2 for this item.”)
- Often respondents will select all items in a list in the screening portion of the survey, perhaps because they know this will help them qualify. Ways to identify those respondents include:
 - Create a “red herring” (fake) item in your list to identify those who are blindly selecting everything.
Example: On consumer studies about financial products, when we ask consumers if they own certain products (CDs, annuities, 401(k)s, etc.), we add in “Bronze accounts” (this not a real investment product) to identify those who are simply selecting all.
 - If you know it is impossible or extraordinarily unlikely that a respondent will own/experience all items in a list, simply terminate those who select all. It is possible you may lose a small proportion who can legitimately say yes to all, but it is worth that loss to maximize the overall data quality.
- Make sure to include a mandatory open-ended question, whether or not you plan to code or analyze the results, just as a cleaning tool (see #3 below).

[more...](#)

03

Back-end Cleaning

We find the most important part of maximizing data quality is in the back-end data cleaning.

- While some of the “traps” in #2 above should terminate poor-quality respondents as they are taking the survey, others may be flagged for analysis during data processing. Ask the programmer to create an indicator in the data file for any traps and red herrings you programmed, as well as for every set of attributes, to look for respondents who may have straight-lined that set. Use the indicators to decide who to remove from your data – create a rule of one, two, or three “strikes” and they are out.
- Look at incongruous data. Respondents often answer questions in contradictory ways, and this can be an indication of bad data. If you have two questions that relate to one another, make sure respondents don’t directly contradict themselves.

Example: On a survey of new moms, one question asked whether respondents receive WIC assistance (financial assistance for low income women and children). In comparing answers to this question with answers to an income question, we identified several respondents who claimed both WIC assistance and six-figure incomes.

- Read every open-ended response. We find this to be the most important check of data quality and a good investment of your time. Make sure the respondent answered the question you asked in a logical way (e.g., remove those who answer “good” to “Why did you decide to become a teacher?”) Some researchers will keep respondents who answer with random text to avoid an open-ended question (e.g., “gfhfghfghfghfg”), but we recommend removing these. Panel respondents are being compensated for participating in the survey, and should be expected to answer required open-ends.
- Be strict. The more stringent your cleaning standards, the better your overall data quality will be. Reputable panel companies should be willing to replace bad respondents for no additional cost.