Defining Quality in Sample

innovate

Introduction

One of the most significant concerns in the market research industry today is sample quality. Since the early 2000s the industry has been rapidly growing, with an increased focus on speed and cost savings as market research moved online. But that speed at times came at a cost.

Data quality issues have proliferated throughout the industry, with increased incidences of fraud, straight lining, the initiation of a portable experience, and the increased need of clients to see results in just a matter of days or even hours.

The focus on speed makes it more challenging than ever to build systems that protect sample quality, ensuring a good cross-section of the right respondents. At the same time, as fraud continues to morph and evolve with the advance of new technologies, how do we define quality? What steps can and should a sample provider take to ensure the integrity of their sample, while combatting potential fraud and dips in quality?

In this guide, we're going to look at some of the most important factors in defining sample quality, how Innovate approaches these industry-wide problems, and what you should look for when preparing your next survey. From technical layers of protection to sample and survey management, there are several factors that directly impact quality in sample, each of them important to ensure good results for clients.

Changes in the last several years

The market research industry is currently facing the newest in a long line of changes and challenges. Since the move of surveys online, the industry has been enamored with speed and cost savings.

That fixation led to a rapid decline in data quality in the early 2000s, so much so that major clients like P&G spoke out as early as 2006 about the issue. While the industry responded in turn, sparking a very real discourse on the definition of sample quality and the steps that need to be taken by both market research firms and sampling companies, the issues haven't fully gone away – they've merely evolved.

Major initiatives have been ongoing to restore and maintain online research quality, but technology's breakneck pace of innovation hasn't slowed. The rapid evolution and diversification of devices, in particular the shift from proximity fixed experience to portable, has had a major impact not only how surveys are given, but how they and respondents in general are managed.

Today, the emphasis remains on speed, and fraud is morphing and evolving to take advantage of this fixation. This means dealing with very real issues like:

A.I. Data – Bots are becoming increasingly sophisticated, to the point that the data from AI is nearly indistinguishable from real data. It's a mess as people increasingly use bots to mimic people and fill out surveys, often in just a few seconds.

Tor Browsers – Internet privacy tools have evolved as well. Tor browsers that are designed to protect users against tracking provide an intangible barrier between many current tracking technologies and the users who are attempting to defraud the system.

Survey Click Shops – Mass efforts to systematize and complete surveys at scale for the monetary reward create massive data issues and combined with the above tactics can be hard to track.

Device Emulators – The ability of technology to emulate different types of devices to fit survey requirements is growing as these tools become easier to access and use. These are just some of the threats faced as technology advances.

Datacenters, VPNs, anonymous proxies – the more advanced the tools available to those seeking to defraud the system, the more vigilant the industry must be in tracking and maintaining quality at all times.

The State of Current Panels

Today's panels are built on an understanding that the threats listed above are real and there are people actively attempting to game the system.

Innovate has structured its panel building process around these realities, using a thorough vetting process when recruiting. This includes careful vetting of the partners' recruitment methodologies, sampling protocols, deduplication technology and incentive management.

More than recruitment, we utilize a series of evaluation tools to test sample quality based on:

○ Demographics

○ Attitudinal benchmarks

○ Behavioral benchmarks

The use of carefully curated benchmarks, constant evaluation of all partners, and a keen understanding of what the most common pitfalls might be are important to avoid common problems related to sample quality.

This also means a careful control of the percentage of sample and distribution across several demographic variables, tenure, and activity levels with our sampling tools. Designed to combat learning behaviors or conditioning by respondents who are on a panel or in a survey research process for a long period, it also helps to detect some of the most common forms of fraud and potential quality issues.

Technology is vital to this approach, allowing us to approach quality respondent management with cutting edge solutions targeting advanced online and mobile fraud techniques. The result is more than two dozen quality check-points in the registration process, with points assigned to respondents based on certain behavior. A single point isn't always indicative of a problem, but taken as part of the whole, multiple points at different stages of the registration and participation process can indicate someone we don't want on the panel is attempting to join.

Because these tools are built into the technology we use, it allows us to silently deactivate the account. It also allows us to further implement even more advanced protections against fraud for targets like our B2B sample.

Let's take a closer look at some of the technologies and tactics implemented to ensure quality at this level, and what it means for the state of panel health overall.

Los Angeles   ●   New York   ●   New Delhi

888-229-6664   ●   sales@InnovateMR.com   ●   InnovateMR.com

innovate

Technical Approaches to Fraud Detection

The fraud detection process requires a careful understanding of the different tactics that might be used by those with nefarious intent. Many of these quality checks should be built into the process from the moment of registration, but also as part of panel quality control over the long term.

What does that mean? It means carefully building a system that protects the integrity of the entire panel against fraud. Here are some of the most important safeguards that should be implemented:

Device Fingerprinting – This information, collected from the device being used by the respondent, serves as a specific identification that can be used in validation. Ideally, every device has its own fingerprint and can be identified in future engagement. Per a 2016 Fraud Report, this technology has been implemented by 32% of companies with 17% more indicating a plan to implement in the future. TrueSample verification places a digital fingerprint on each user when they enter a survey, allowing us to assess for potential bot activity, black-listed IP addresses and other fraud markers.

IP Geo Location Information – IP GEO location information has been implemented by more than half of all companies with another 13% indicating plans to implement. Pegging the location of users based on IP address can be done for most devices, though other factors here can become an issue if IP is overridden by anonymous browsers, VPN, or other privacy tools.

Identity Address Validation – By validating real world identity via name, postal address, email address, and other personal information against existing consumer databases, a real name can be assigned to each respondent. This allows for effective removal of duplicate entries as well.

GEO Location Distance Check – By combining the above two factors, we can evaluate where someone says they are versus the IP address logged when they are active to determine if they are where they say they are (or are using a secondary device).
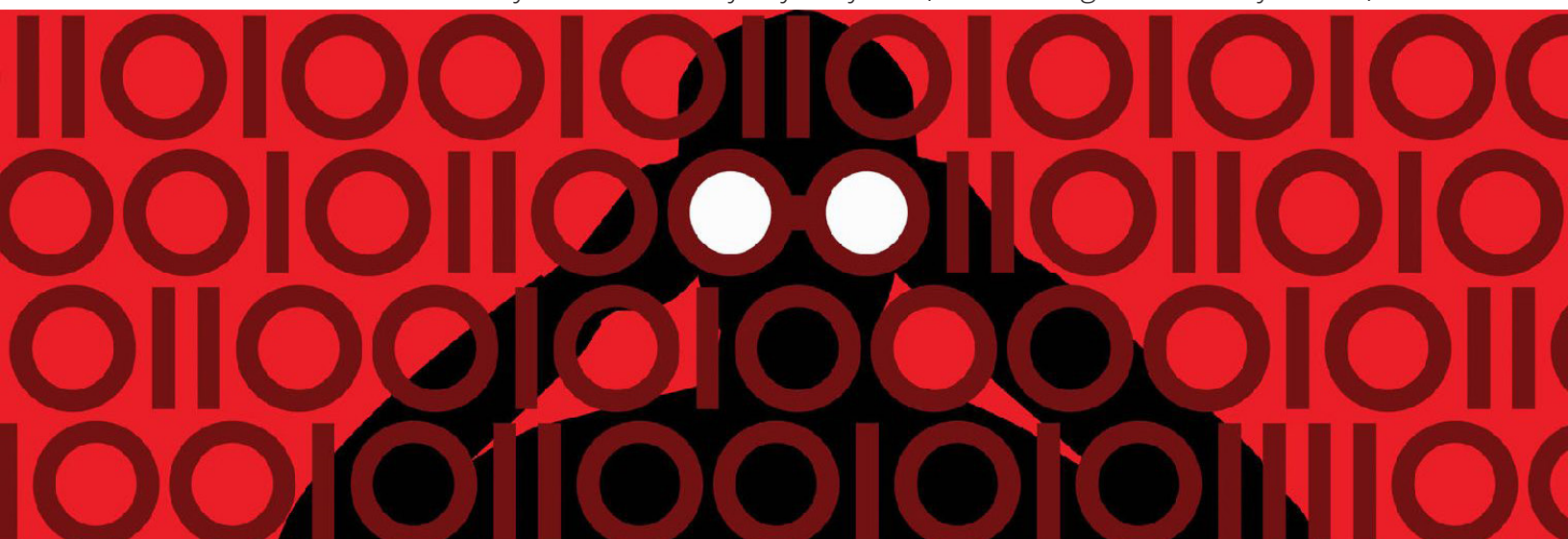
Additional resources can be drawn on at the device level to check for potential fraud indicators with first time users. These include

- Language Check
- GEO Browser Language Check
- GEO Time Zone Check
- GEO Country Check
- Multi-Device Check
- Bot Check
- Anonymous Check
- Blacklist Check
- Browser Status Check

While fraud on a larger scale may be able to fool several of these factors, the multi-step process can help identify those who are using device emulation, spoofing location, farming survey responses, or who are otherwise bypassing frontend quality checks during registration and survey activities.

Response Validation

A good response validation tool leverages real-time Bayesian statistical models and analysis to determine the engagement of users. In short, are they engaging in a way that matches typical human behavior?

Using these statistical models, a respondent can be flagged as unengaged in a survey if they speed on a certain percentage of pages they saw during the survey. There are established norms and standard deviations for each page that can be calculated and updated in real time as the page submissions from this and other respondents are received by the platform. In short, outliers are flagged almost immediately and can be evaluated for engagement levels.

Another important factor that this technology can measure is response pattern. Undesirable response patterns on a certain percentage of pages can also be considered unengaged with the survey and flagged as such.

Behavioral Approaches to Data Quality

Even with extensive fraud detection and resources in place to identify potential high-risk respondents during the registration process, it's important to have resources in place to ensure high data quality in studies. This can be done through sample design and management, survey design, and ongoing member management. Let's take a closer look at each of these three factors:

Ⓘ Sample Design and Management

Before a single person reaches your survey, how is your sample designed and who are you sourcing through? These questions can have a profound impact on the overall quality of your responses when a survey is completed.

Vendor selection will have a huge impact here. The method by which sample is sourced by your vendor, along with the management and incentive system used will all impact the quality you receive. You should ask important questions that will directly impact how your sample is built, including:

○ How sample outgo is balanced
○ The measures implemented to ensure high quality
○ Demographic balance
○ Survey field time
○ Invitation and introductory language used
○ Competing survey inventory
○ Survey frequency and variation
○ Routing and project prioritization methods

Know how your vendor sources their sample, what sampling methods they use to match your client's needs, and the steps they have taken to ensure quality in their panel. Consider how they source their sample as well. Uniqueness is a major challenge for online panel development, leading to high levels of overlap between panels as sample companies all fish from the same pond. Offline   recruitment methods such as television are an effective way to combat this.
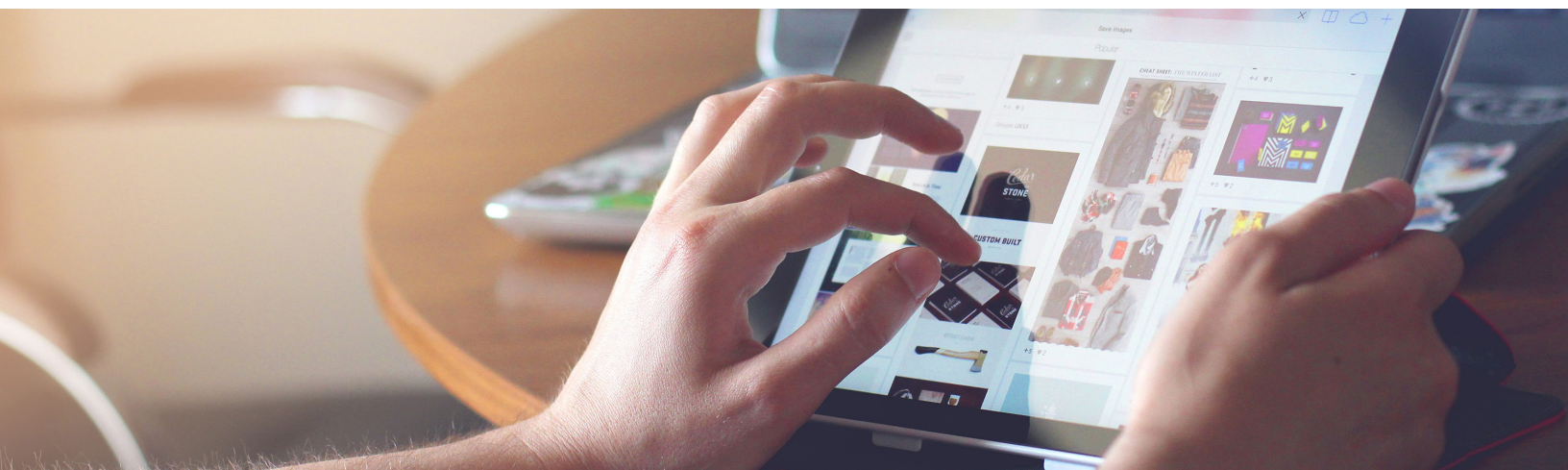
(II) Survey Design

Not all quality issues are related to intentionally misleading answers or fraudulent responses in a panel. Survey quality can be equally impacted by poor design, ineffective questions, leading language, and formats that don't encourage the kinds of answers you're looking for.

Question design is an area where we have more work to do, especially as attention spans shorten and mobile devices provide a faster, more accessible method of interaction for users. What should you focus on in your survey design? There are several factors to keep in mind, including:

- **Non-leading Wording –** Leading respondents towards certain potential responses can reduce data quality and push results in a certain direction.
- **Outs for Respondents –** Give necessary outs for your respondents to take if needed.
- **Sparing Use of Open-Ended Questions –** Open ended questions can be burdensome and are not always clearly labeled in terms of what is needed. Users are less patient than ever before and are frequently on mobile devices, making these harder to complete.
- **Avoiding yes/no Format –** The quality of data you get from yes/no is minimal. It can be easily straight-lined and allows for little or no ambiguity.
- **Avoid Burdensome Question Formats –** Grids, lists with more than attributes, anything that requires scrolling, or those that don't display well on mobile devices should be avoided as much as possible.
- **Concise Wording –** Keep language short and concise so users can get through it as quickly as possible. Aim for a 5th grade reading level.
- **Reduce Visual Clutter –** Keep screen clutter to a minimum. This means minimization of the number of questions on screen as well.

innovate

Mobile Friendly Design

Upwards of 40% of all web activities happen on mobile devices, and the number keeps growing. Build your surveys with mobile users in mind.

Research increasingly shows that users are less engaged with non-friendly mobile pages, whether it's a poorly optimized survey or one that doesn't account for mobile use at all, despite the large number of people who use mobile devices for all of their web activities.

Good survey design will not only ensure it's easier to track and identify potential fraud – it enables good respondents to provide the best possible answers in the format that is most accessible to them.

(III) Member Management

Finally, there's member management. How do you ensure survey respondents are legitimate once they are presented with a survey, and how do you maintain the integrity of your results over time?

From a technology standpoint, there are several tools you can use to identify potential fraud in a survey:

- **Honey Pots –** Using a programmatic computation behind the scenes you can add a hidden question to your survey – one that humans cannot see but that a bot can. If you see data for this question in a response, you know it was completed by a machine, and not a person.
- **Algorithmic Solutions –** Algorithms that track activity over time and identify LOI completions and repeat issues or invalids are highly effective when properly implemented.
  In addition to technology, there are several common sense hands-on things you can do to validate user identity over time:
- **Profiling and Third Party Data Validation –** There are services that will perform these validations and ensure the data in your member list is accurate and remains that way.
- **Demographic Consistency Checks –** If you ask for basic demographic information during registration, this information can be rechecked later with validation questions.

Finally, there is the use of trap questions, which can be highly effective if used properly. These can be tricky, however, skewing either too complicated or too simple. On the one end of the spectrum, they can be frustrating and burdensome for users. On the other, machines can learn and overcome them, so it's important to build them effectively

- **Instructional Trap Questions –** These ask users to do a certain task, such as choose all images with a house in them or identify all the numbers or a specific image with an object in it.
- **Skill Based Questions –** These are commonly used on front-end forms and ask users to make basic computations, like 5+2.
- **Honesty Based –** These will ask users to describe things like the brands they have interacted with or activities they have participated in over a certain time frame.

The key to effective use of trap questions is to implement multiple measures and not rely on a single question to measure quality. At the same time, don't place these questions at the end of a long survey as false positives can invalidate good surveys. Use them in places where users are more likely to respond accurately if they are legitimate respondents.

innovate

Maintaining Data Quality in Your Survey Responses

Maintaining quality in sample will always be an issue faced by both market research and sample companies. It requires vigilance and a targeted approach to screening, registration, survey design and member management.

It's why the Innovate team has invested heavily in technology to remain at the forefront of fraud detection and sampling accuracy for our panel. This isn't a one-time issue that can be addressed with a few small changes. It requires constant attention.

For market research firms, you need to know your sample source, use multiple trap-based questions based on the specific knowledge of your target audience, implement several tiers of quality measurements, and leverage 3rd party data sources to validate member data at recurring intervals.

By building a system that actively monitors for fraud, detects potential red flags, and provides well-structured surveys that encourage accurate and effective responses, you can combat potential quality issues in your data while taking advantage of the vast pool of insights available from a quality panel.

Learn More About Innovate

Learn more about Innovate and our first-class panel on our website. You can download a full description of our panel in our 2017 Panel Book or a recent webinar recording in which we discussed sample quality in greater depth. Click the link below to learn more:

Learn More About Innovate